EU RIA H2020 Proposal Template

ROBHOOT

Abstract

Eco-evolutionary biology reveals how interactions and traits diversify across multiple scales of organization, from neurons to populations and ecosystems. Evolving networks in nature with ever changing traits and connectivity patterns can inspire a new discovery computation for a global-sustainable knowledge-inspired society. Many studies have shown sustainability could be achieved by strengthening transparency, communication, and rapid access to discovery technologies. Sustainability goals, however, depend on global access to discovery-based knowledge. Yet, science-enabled technologies targeting knowledge discovery to reach sustainability goals are not in place. We propose an eco-evolutionary diversification-inspired discovery computation technology for a knowledge-inspired society. We introduce evolutionary diversification-inspired and artificial intelligence solutions for sustainability in natural ecosystems. We validate our approach with a sustainability of the Seas case study in federated networks, where many distinct groups of species, humans and technologies coexist exploiting resources in complex ecosystems. Knowledge discovery running on a federated network encompasses a hybrid-technology to lay out the foundation of an open- and cooperative-science ecosystem for computation discovery in the face of global sustainability challenges. The project summarized here is not only set out to deliver knowledge discovery computation in federated networks, but also to provide fully reproducible open-source software solutions of a science-enabled technology to connect knowledge-inspired societies to global sustainability challenges.

Knowledge discovery in eco-evolutionary diversification-inspired federated networks

ROBHOOT

1 Excellence

1.1 Radical vision of a science-enabled technology

Ecosystems collapse around the globe in the absence of technologies to discover novel ways of sustainable exploitation of complex ecosystems. In this regard, rapid, real time, heterogeneous- and cooperation-based, discovery computation is currently a major issue revolving around data-driven intelligent machines and knowledge inspired societies facing global sustainability challenges. However, diversifying networks are not used for discovery computation yet, despite rapid changes in trait and interaction have been observed in experimental and theoretical systems [15, 17]. Evolving networks are characterized by feedbacks between the ecology and evolution of interacting traits, the eco-evolutionary feedbacks, to produce novel trait changes with new functional properties in ecosystems. This results in new computational properties, like interactions (i.e., cooperation, competition, antagonism, etc), and information processing and learning capabilities. Conventional Artificial Intelligence (AI) computation is rapidly moving towards explainable and discovery pattern inference [21] but often avoids evolutionary diversification for exploring new computing capabilities [26]. The same situation occurs for artificial neural networks that also make limited use of novel computing capabilities as a consequence of the emergence of new interactions and traits [28]. **The goal of this project is to implement eco-evolutionary-diversification inspired solutions to make discovery computation based on rapidly evolving traits and interactions**. The exploitation of evolving connections and traits will allow us to create novel discovery computation solutions for natural ecosystems facing sustainability challenges like overexploitation of the Seas, where harvesting renewable resources are in the point of diminishing returns for many species, communities and ecosystems (refs +++). Why should we go deeper into diversifying information processing for discovery computation? With connections and traits represented in a spatially distributed network, as found in natural ecosystems, it is possible to untangle mapping of many spatiotemporal inputs onto many output functions. This allows considering not only evolutionary processes changing traits and agents but the diversification of new entities to decipher new solutions for harvesting renewable resources. This also allows representing real-time solutions for ever changing renewable resources, which is a key problem in many digital and natural ecosystems.

To show the capabilities of the ROBHOOT approach, we will complement novel implementations of evolutionary diversification-AI discovery computation with full reproducibility, automation, visualization and reporting to trigger its citizen science and scalability properties at large-scale (Figure 1). **The main impact of ROBHOOT is to provide novel open-source software for reproducible discovery computation solutions to substantially improve ecosystem sustainability relevant for community-rich digital and natural ecosystems**. To support this notion, we will perform eco-evolutionary diversification and AI inspired simulations accounting for heterogeneity in data-sources. The central goals of ROBHOOT are:

1. To extend existing theories of eco-evolutionary diversification and AI inspired solutions to decipher the factors driving discovery computation in federated networks. This will allow us to identify novel paths of reliable solutions for ecosystem sustainability.
2. To investigate how spatiotemporal evolutionary diversification and AI inspired networks can mimic the empirical patterns of natural and socio-technological ecosystems when large and heterogeneous exploiting human groups and species coexist.
3. To develop fast, reproducible and automated eco-evolutionary biology-inspired discovery computation prototypes for real-time information processing tasks.
4. To arrive at powerful discovery computation principles for forecasting in federated networks when diversification in interactions and traits occur in a large and heterogeneous pool of species, technologies and human groups.

1.2 Science-to-technology breakthrough that addresses this vision

Data knowledge discovery (WP1)
Evolutionary biology-inspired semantic algorithms: Most studies of data discovery focus on advanced analytics functions to reveal insights, ignoring the discovery of data-source heterogeneity almost completely. Currently, only a few databases are semantically annotated from many data-sources (e.g., gene ontology database, COVID-19). Ontology development is time-consuming and requires expert knowledge. It is also ideally paired with data-driven research that iteratively checks the soundness of the ontology as it simultaneously seeks discovery. Thus, software tools for mapping and linking the terms between different ontologies accounting for many data-sources are still not in place [6, 4].
Going beyond ROBHOOT will go beyond state-of-the-art to implement evolutionary-biology inspired semantic algorithms. We will explore insertions and deletions, different types of recombination and crossover, and other evolutionary-based functions to find datatype properties from ontologies and raw-data from non-semantic databases. ROBHOOT will explore algorithms to gain understanding of the replicability of data heterogeneity contrasting different evolutionary algorithms.

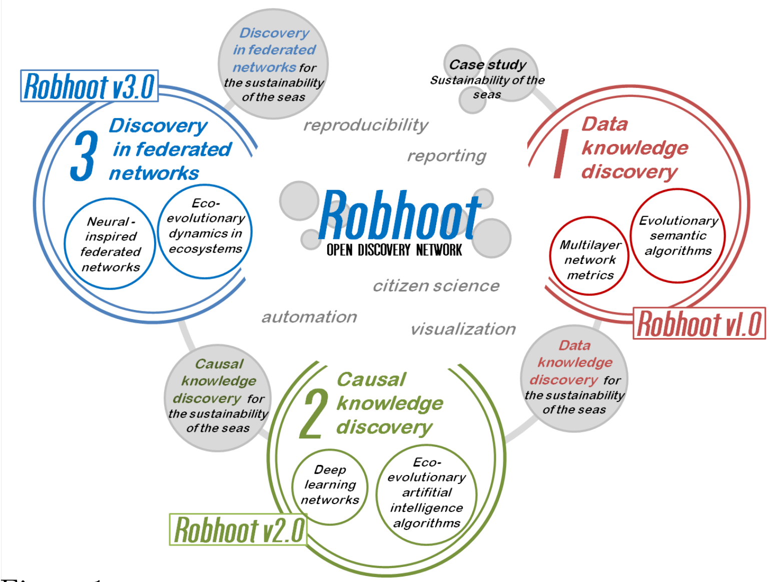

Figure 1: **Discovery in evolutionary diversification-inspired federated networks.** ROBHOOT target knowledge discovery when heterogeneous groups of species, humans and technologies share ecosystem resources for a sustainable knowledge-inspired society: It introduces three science-enabled technologies: Evolutionary biology-inspired semantic algorithms for ROBHOOT v1.0 (data knowledge discovery, red), eco-evolutionary diversification-inspired AI models for ROBHOOT v2.0 (causal knowledge discovery, green), and evolutionary neural diversification-inspired federated networks for ROBHOOT v3.0 (discovery in federated networks, blue). ROBHOOT uses the sustainability of the Seas case study in federated networks to offer a compact open-source technology with full reproducibility, automation, visualization and reporting for an open citizen science.

Causal knowledge dicovery (WP2)
Eco-evolutionary diversification-inspired AI algorithms: Causal, explainable or interpretable discovery from observable data has been extensively studied (refs ++). Many of these studies have used symbolic reconstruction of equations by symbolic regressions or evolutionary methods (refs +++). However, a common gap throughout much of the literature is that of reconstruction from partially known models where the parameters represent eco-evolutionary and diversification processes from where explainability of the data can be done efficiently. The classical view on biology-inspired information processing technologies is to consider plasticity without structural changes, or without diversification among many interacting components (refs +++). Recent experimental evolution studies shows that rapid trait changes with new information processing capabilities is far more complex (refs +++). For example, eco-evolutionary dynamics strongly affect feedbacks between ecological and evolutionary processes, which in turn influences trait changes to open new properties with new information capabilities (refs +++). Furthermore, recent studies suggest that the interplay between trait dimensionality and adaptation is key to understand the emergence of new traits and information processing abilities to elaborate new discovery computation strategies in ecosystems (refs+++).
Going beyond ROBHOOT will, for the first time, employ eco-evolutionary diversification-inspired solutions to implement AI process-based methods to represent spatiotemporal causal inference in systems containing large heterogeneity and dimensionality (Figure 2). Eco-evolutionary diversification-inspired models will be extended to deep process-based learning networks including trait and interactions as evolutionary changes to understand patterns in these systems. The search for causal knowledge discovery will be applied to the data knowledge discovery generated in WP1 for the sustainability of the Seas. This database started in 1965 and currently contains 9 mill. entries, 1612 species (i.e., 50 variables and

traits per species), around 20 countries and 11 sampling methods (Figure 2). Our approach will explore broad classes of evolving functions from eco-evolutionary diversification-inspired AI algorithms combining them to automated Bayesian machines ensuring the search, the evaluation of models, trading-off complexity, fitting to the data and quantify resource usage [20, 29].

Discovery in federated networks (WP3)

Evolutionary neural diversification-inspired federated networks: Technologies in digital ecosystems around federated networks are rapidly increasing and mostly focus on decentralization, scalability and security fronts [19, 13, 14, 8, 24, 9]. Yet, the discovery of novel algorithms in diversification-inspired federated networks for forecasting of global sustainability problems when heterogeneous groups change and learn from each other is currently not in place. Recent studies have shown the importance of evolutionary search of mathematical and symbolic operations as building blocks to discover ML algorithms ([26, 20]). In this regard, evolutionary diversification-inspired search for algorithmic discovery can help to decipher how interactions among heterogeneous groups evolve and learn to solve complex sustainability problems. Evolutionary dynamics can explore open-ended language of models with varying trait evolution functions to discover biologically inspired solutions in multidimensional systems ([26],+++). ROBHOOT v.3.0 deploys biology-inspired federated networks accounting for heterogeneous agents to discover novel biology-inspired solutions for the sustainability of the Seas federated network.

Going beyond: Our understanding of the outcomes from diversified information processing systems formed by highly heterogeneous groups, a kind of large-scale meta-learning in the federated setting [13], is currently quite limited. Therefore, new science-enabled approaches accounting for diversifying information processing in heterogeneous and highly dimensional systems are required. This allows to develop science-enabled technologies where heterogeneous agents with different interests find (non optimal) solutions for the sustainable exploitation of ecosystems. Federated objects can be seen as "neural networks" containing many types of heterogeneous nodes with varying degrees of learning, connectivity and firing probabilities [22, 23]. ROBHOOT v.3.0 connects knowledge discovery to diversification-inspired federated netwoks to study the properties of cooperative

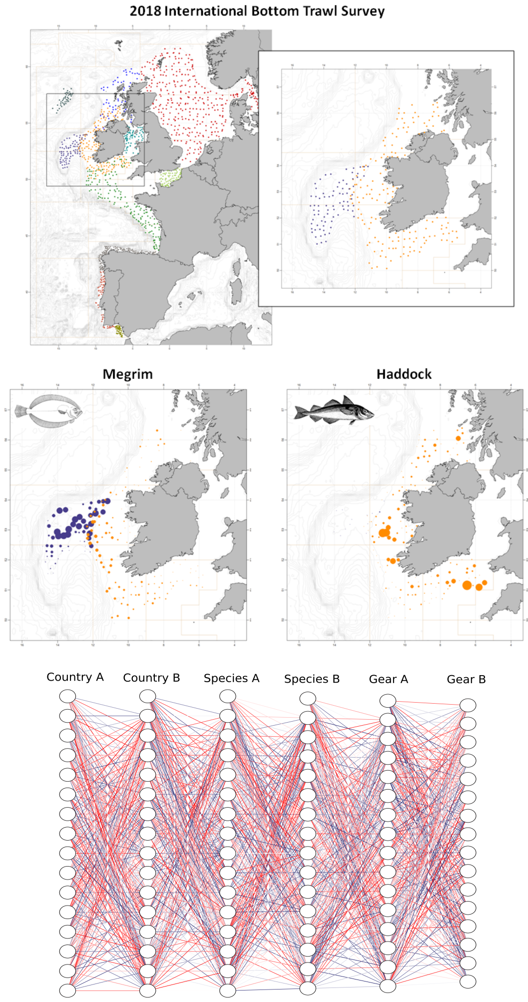

Figure 2: **Discovery for Sustainable Ecosystems**. ROBHOOT target discovery computation for new sustainability paths in complex ecosystems. The sustainability of the Seas case study [7] will be enriched to validate the technology when many species, human groups and technologies exploit resources (**Top left**, data-color points represent sampling from different countries. Zoomed in is the Irish ground Fish (IE-IGFS, Orange) and the Spanish survey on the Porcupine Bank (SP-PORC, Blue). Countries produce strong bias in the distributions maps because they use different Gears according to their commercial interest (Megrim, *Lepidorhombus whiffiagonis*, consumed largely in Spain and France, **Center left**) and Haddock, *Melanogrammus aeglefinus*, highly priced in northern Europe, **Center right**). This generates strong bias for sustainability in natural resources. ROBHOOT integrates evolutionary biology-AI-inspired solutions represented as networks with many layers to discover sustainability paths with many coexisting species, human groups and technologies **Bottom**. Each country, species and gear is composed by many nodes: country contains fishery, environmental agency, stakeholders, etc. Species contains size-classes, habitat preference, species interactions, etc. Red and blue links mean competition and cooperation links connecting each pair of nodes.

forecasting in the face of global sustainability challenges.

1.3 Interdisciplinarity and non-incrementality of the research proposed

To succeed with ROBHOOT, it is essential to build an interdisciplinary team that includes scientists from different disciplines, i.e., evolutionary biology, ecology, computational neuroscience, computer science, data science, complex systems and experts in communication and biodiversity sampling methods. Data knowledge discovery gained by the analysis of the computation discovery capabilities of evolutionary-inspired semantic algorithms will be developed by the evolutionary biology, computer science and complex system members of the consortium (EBD-CSIC, IFISC-CSIC, SDSC). Data discovery can be transferred to the causal domain addressed by the other part of the consortium with expertise in ecology, evolutionary biology, data science and causal inference (EAWAG and TARTUR). The whole process will be enriched with full automation, reproducibility and visualization supported by ICREA, SDSC, and our company-partner SCITE, respectively. Conversely, those scientists working on neurobiology and eco-evolutionary dynamics in ecosystems will feed information back on fundamental discovery computational challenges in federated networks (i.e., role of heterogeneity, evolving and diversifying traits and interactions, cooperation, and dimensionality). This allows feedbacks in their implementations to explore to what degree this is reflected also in eco-evolutionary biology-inspired and neurobiology inspired discovery computation models to augmented their models. This cross-fertilizing back-and-forth interaction will allow the project to keep high modularity within the WPs while keeping cross-interactions among the groups to run efficiently the stages of the project. To bring together evolutionary biology-inspired semantic algorithms for data discovery and evolutionary neurobiology-inspired discovery in federated networks require a long stride and this has not been attempted so far. This way, we expect to realize a truly novel, sustainability-driven knowledge-inspired society technology for which there are no predecessors. Thus, ROBHOOT will not be incremental, but a leap opening a new direction for eco-evolutionary diversification-inspired discovery computation.

1.4 High risk, plausibility and flexibility of the research approach

ROBHOOT represents a novel approach for complex, adaptive and multidimensional discovery computation in ecosystems. The transfer of eco-evolutionary diversification-inspired principles onto fully reproducible and automated software, progressing towards a process-based discovery technology, will be a major qualitative step, defining ROBHOOT as a high-risk project, fitting into FET-Open. To achieve the ambitious goals, we will combine expertise from all involved areas, mitigating risk in a gradual way, following a strict line and increasing step-by-step in complexity of the problems addressed. We will start with evolutionary biology-inspired semantic algorithms for data discovery applying them to the sustainability of the Seas case study. This is followed by the implementation of more complex eco-evolutionary diversification-inspired AI modeling to infer causality in our case study. Then, we will advance to more complex situations, where the evolutionary neurobiology-inspired modeling will expand the search along many diversifying and cooperative forecasting schemes to find paths of sustainability in our case study. To keep the project technically feasible, and to be able to identify the mechanisms and their properties from data and causal discovery computation to discovery in federated networks, we will limit methods to three main approaches. All of the above will be done by combining theoretical work and numerical simulations with a real empirical case for the sustainability of the seas. The knowledge gained along these three lines will allow us to compactly represent all the steps into a compact science-enabled open-source technology. We will develop the modeling in fast computing languages implementing low-level Agent Based Models (ABM) along all the theoretical development of the proposal (i.e., Julia, C++). We will contrast the ABM with differential/difference equations methods when a large number of agents, traits, and interactions diversify in time and space. This feature represents a very desirable fallback in case of speed and convergence problems for multidimensional and nonlinear systems (Table 1.4a, Critical risks for the research approach). Our implementation activities are all complemented by numerical investigations contrasted for speed, replicability, and robustness with the sustainability of the Seas case study (Figure 2). The success of ROBHOOT would represent a breakthrough in the current discovery computation with direct application to sustainability of ecosystems and beyond. The combination of rapid, data heterogeneity and cooperation for discovery computation based on open-source code will lead to fast implementations of the demonstrators with high flexibility that will permit a rapid transit to the public.

Table 1.4a: Critical risks for the research approach

Description of risk	**Objective**	**WP**	**Proposed risk- mitigation measures**
Evolutionary semantic algorithms insufficiently developed: Medium	2	WP1	Use traditional non-semantic genetic algorithms to infer data connections.
Multilayer metrics accounting for spatiotemporal patterns along many datasets insufficiently developed: Low	2	WP1	Implementation of more standard complex networks metrics to characterize data knowledge discovery.
Low number of training data available: Medium	2,3	WP2	Alternative methods focusing on matrix decomposition methods.
Automated evolutionary-inspired expressions for causal knowledge discovery insufficiently developed: Medium	2,3	WP2	Symbolic regression methods to full automation for causal discovery accounting for evolutionary rules.
Eco-evolutionary dynamics of multiple traits in species-rich ecosystems insufficiently developed: Medium	1-4	WP3	Mean-field approximations using classical ODE systems and novel universal differential equations for scientific machine learning.
Evolutionary neurobiology-inspired federated networks insufficiently developed: Medium	1-4	WP3	Spiking neural network models as alternatives to evolutionary neural biology-inspired algorithms in federated networks.
Cooperative forecasting mixing eco-evolutionary dynamics and neural nets in large scale federated networks insufficiently developed: Medium	1-4	WP3	Mix eco-evolutionary dynamics models with less alternative neural nets models working a smaller spatiotemporal scales.

2 Impact

2.1 Expected impact

- **Scientific and technological contribution to the foundation of a new future technology**: ROBHOOT target discovery of novel evolutionary diversification-inspired algorithms to substantially improve solutions for sustainability in ecosystems. **Discovery of novel evolutionary-inspired algorithms in the context of diversifying traits, interactions, technologies and human groups for biodiversity maintenance have been hardly been investigated in this context so far**. Therefore, several predictors related to biodiversity, technological and social times series analysis will be tested and further developed to enable robust prediction of sustainability. Altogether, this project will lay the foundation for future sustainability studies.
- **Potential for future social or economic impact or market creation**: Our approach accounts for heterogeneous sources of data, the evolving mechanisms driving technological, environmental and social changes required to make ecosystems sustainable. This will allow to use the technology in public and private industry, for example, to generate rapid and robust scenarios when facing complex problems including global sustainability challenges (e.g., global health, food and feed production, ecosystems degradation).
- **Impact on transparency and reproducibility**: Decision making and governance at local, regional and global scales require access to transparent and reproducible information containing

the factors and their plausibility explaining the empirical patterns. ROBHOOT consortium brings together different partners in the fields of computer science, neurobiology, complex system, biology, social sciences, evolutionary ecology and one SME all focusing on reproducibility, automation, visualization and reporting scientific data to different audiences. A reproducible, open-access, and automated tool will be developed accounting for global data-arquitecture and addressing scenarios of future strategies for sustainability.

- **Ecosystem health impact**: ROBHOOT focus on novel discovery solutions for ecosystems that can be under a varying degree of disturbances (e.g. fires, floods, droughts, overexploitation of natural resources). It uses a case study for overexploited ocean ecosystems with highly heterogeneous social groups and different interests in the exploitation of limited and shared resources. This is a technology designed to provide novel solutions for ecosystem sustainability, improving the underlying discovery paths, thereby connecting ecosystem sustainability and ecosystem health. This feature aligns to the EU Reflection paper towards a Sustainable Europe by 2030 and the UN's Sustainable Development Goals. ROBHOOT can be seen as an horizontal enabler for a scientific-based transition to sustainability based on large amounts of heterogeneous data, artificial intelligence and evolutionary-biology inspired solutions.
- **Building leading research and innovation capacity across Europe**: This consortium brings together excellent partners from the fields of computer science, machine learning, deep learning networks, neurobiology, complex systems, experimental biology, biology and evolutionary ecology, physics, theory and applications of complex systems in social networks, delivering a highly innovative science-based reproducibility, automation, reporting and communication tool focusing on sustainability solutions. All consortium partners exhibit a long-standing experience in interdisciplinary research across the boundaries of the individual disciplines. A web-based sustainability discovery portal will be produced (WP3), which will allow researchers, NGO, managers and the public to train students in the discovery process to manage over-exploited ecosystems. This will also allow to scale up the number of people participating in the sustainability process thus mobilising forward thinking researchers and excellent young researchers to work together and explore what may become a new discovery technology paradigm in sustainability research.

2.2 Measures to maximize impact

Dissemination and exploitation

A plan for dissemination and exploitation (PDE) will be developed and managed under WP4. It will address the project strategy and concrete actions related to: i) Dissemination: Open Access format; ii) Data Management: how data will be handled; iii) Protection: IPR strategy; iv) Exploitation, namely "business models", and v) Communication, particularly the different action to communicate the project's results and prototypes to key groups of end-users. The PDE will also have a Dissemination and Exploitation Board (DEB).

- **Open Access**: Project reports and ISI journals publications will be under the Open Access format. Following the Open Science principles, software and scientific publications will be deposited in the online institutional repositories and on the EC Participant Portal. ZENODO (http://zenodo.org), a public repository recommended by the European Research Council and the EC, and supported by EUs OpenAire platform (https://www.openaire.eu/) will be also used for dissemination and communication purposes (publications, presentations, datasets, images, videos/audio and interactive materials such as lectures).
- **Open access to research data**: recommended data repositories (e.g. PANGAEA, NASA Goddard Earth Sciences Data and Information Services Center) will used to share the generated data and software. Open-source codes and analysis of standardized inputs/outputs and software will be made public through an online platform with the aim of converting it in *the Reference Point* for any future research in knowledge discovery.
- **Data management**: Good research data management practice will ensure that data produced or used during ROBHOOT is registered, stored, made accessible for use, managed over time and/or disposed of, according to legal, ethical, funder requirements and good practice. This management

will provide benefits such as reducing the risk of data loss, improving data workflows and data availability and discovery, visibility of research outputs, attracting new collaborators and research partners, strengthening of the research environment and infrastructure. A data management plan (DMP) will be created by Project Coordinator in close cooperation with the partners and approved by the Steering Board at the start of Project. The DMP will follow the FAIR principles. The document will describe how to collect, organize, manage, store, secure, back-up, preserve, and where applicable, share data.

- **Innovation and IPR**: The Consortium will benefit from the innovation and technology transfer environment in place at EAWAG, which will examine for the individual case if an invention is patentable and if a patent application would be economical reasonable. Support is also available to assist the realization of innovative ideas into efficient business concepts. The necessary precautions will also be taken to protect the IPR of individual institutions. A Consortium Agreement will be signed prior to the beginning of the project to take into account the different interests of the partners, in particular how to treat pre-existing knowhow, the ownership of the results and the intellectual property rights in order to prevent conflicts during the project. The Steering Board will ensure that all innovations and generated data are exploited to the benefit of the involved partners.
- **Exploitation, including business models**: The project's results will be showcased in trade shows (e.g. WebSummit), by communicating through specialized trade press media, and also to a targeted audience (policy makers, funding agencies, industry and SMEs). A detailed business plan will prepared during the project work in collaboration with the SME and academic partners involved, with the ultimate goal of creating a Start-Up at the end of ROBHOOT. The value proposition of the project is develop computation discovery solutions for rapidly diversifying traits and complex interactions that improve the sustainability of exploited natural ecosystems.

Communication activities

ROBHOOT has very general communication targets, from scientists and decision-makers, to the business community and the public. ROBHOOT's general dissemination measures will focus on project results and stakeholder engagement through the following activities.

- **Scientific manuscripts and conference presentation**: High-impact scientific manuscripts are expected, together with the presentation of results in scientific conference, as well as organization of special sessions in international scientific and technological meetings.
- **Website**: A dedicated website and a public git ROBHOOOT repository, which is already available (ROBHOOT git repository), will be used for communicating results and sharing updated versions with all target audiences.
- **Hackatons and robhacks**: activities to attract multipliers and developers from the open-source community to the community who engage in data analytics and build hybrid evolutionary biology-inspired AI algorithms. Workshop: At the end of the project we will organize a workshop specifically on *Next generation evolutionary-biology AI inspired solutions for global sustainability challenges* for disseminating our results to a broad set of groups and experts in fields related to global sustainability for assessing future exploitation potential, inviting partners from academia as well as industry.
- **Testnet**: ROBHOOT will launch a testnet to help disseminate the main results of discovery in federated networks (Section 3.1.3). The launch will have invited NGO's and GO across disciplines and social, economical and technological sectors. The ROBHOOT Open Discovery Network will be launched as a Biodiversity and sustainability open discovery network to offer the solutions for the exploration of the Seas case study and to integrate additional public databases and data collections into the open discovery network to facilitate NGOs, GOs and other organizations transparency, reproducibility, and governance in ecosystem management.