EU RIA H2020 Proposal Template

ROBHOOT

Abstract

Eco-evolutionary biology teaches us how interactions and traits evolve and diversify across levels of biological organization, from neurons to populations. Evolving networks in nature with ever changing traits and connectivity patterns can inspire a new computing discovery for a global-sustainable knowledge-inspired society. Many studies have shown global sustainability could be achieved by strengthening transparency, communication, and rapid access to discovery technologies. Sustainability goals, however, strongly 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 biology-inspired computing discovery technology for a knowledge-inspired society. We introduce evolutionary biology-inspired and artificial intelligence solutions to explore sustainability of the Seas in federated networks, networks composed by many distinct groups of individuals within species, humans and technologies exploiting resources in complex ecosystems. Knowledge discovery running on a federated network encompass a hybrid-technology to lay out the foundation of an open- and cooperative-science ecosystem for computing 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 the architecture of a science-enabled technology, as a proof-of-principle, to connect knowledge-inspired societies to global sustainability challenges.

Knowledge discovery in eco-evolutionary biology-inspired federated networks
ROBHOOT

1 Excellence

1.1 Radical vision of a science-enabled technology

Rapid, real time, data heterogeneity- and cooperation-based, discovery computation is currently a major issue revolving around data-driven intelligent machines and knowledge inspired societies. Several of these properties are found in evolving networks being these changes occurring in dynamic connectivity patterns and/or traits in neurons, and populations in natural ecosystems. However, evolving networks are not used for discovery computing yet, despite rapid trait evolution has been observed in experimental and theoretical systems [15, 17]. For example, 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 novel interaction types (i.e., cooperation, competition, antagonism, etc), morphologies and/or evolving learning capabilities among agents to add discovery computing properties to the network. Conventional Artificial Intelligence (AI) computation is rapidly evolving towards explainable and discovery pattern inference [22] but often avoids evolutionary changes for exploring new computing capabilities [27]. The same situation occurs for artificial neural networks that also make only limited used of novel computing capabilities as a consequence of evolutionary changes in interactions and traits [29]. However, the rapid novel properties of evolving connections and traits, the diversifying power of biological systems, that make evolutionary-biology inspired networks highly plastic and resilient, have not yet been exploited in discovery computation. **The goal of this project is to implement eco-evolutionary-biology inspired solutions to make discovery computation a cooperative game of rapidly evolving traits and interactions**. The exploitation of evolving connections and traits will allow us to create novel types of 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 evolving and diversifying information processing systems for discovery computation? With connections and traits (i.e., nodes and links in networks) 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 considering learning among the interacting and heterogeneous traits and agents to decipher new solutions for harvesting renewable resources. This allows representing real-time solutions for spatiotemporal ecosystems with renewable resources, which is a key problem in many digital and natural ecosystems.

To show the capabilities of the ROBHOOT approach, we will complement the novel implementations of evolutionary biology-AI discovery computation with full cycle reproducibility, automation and visualization to trigger its properties at large-scale (Figure 1). **The main impact of ROBHOOT is that we provide novel discovery computation solutions to substantially improve ecosystem sustainability especially relevant for community-rich digital and natural ecosystems**. To support this notion, we will perform eco-evolutionary biology-AI network inspired simulations of multiple data-heterogeneity based networks. The central goals of $\mathcal{ROBHOOT}$ are:

1. To extend existing theories of eco-evolutionary biology-AI inspired in networks to obtain understanding of the factors and their interactions underlying discovery computation in cooperative federated networks. This will allow us to identify novel paths of reliable solutions for ecosystem sustainability.
2. To investigate how spatiotemporal evolutionary biology-AI-inspired networks can mimic the empirical patterns of natural and socio-technological ecosystems when large and heterogeneous exploiting groups and species coexist.
3. To develop fast, reproducible and automated discovery eco-evolutionary biology-inspired computation prototypes for real-time information processing tasks.

4. To arrive at powerful discovery computing principles for cooperative forecasting in federated networks, models of evolutionary neural biology-inspired to investigate cooperative forecasting for ecosystem sustainability when changes in learning, interactions and traits occur in a large and diverse pool of species, technologies and human groups.

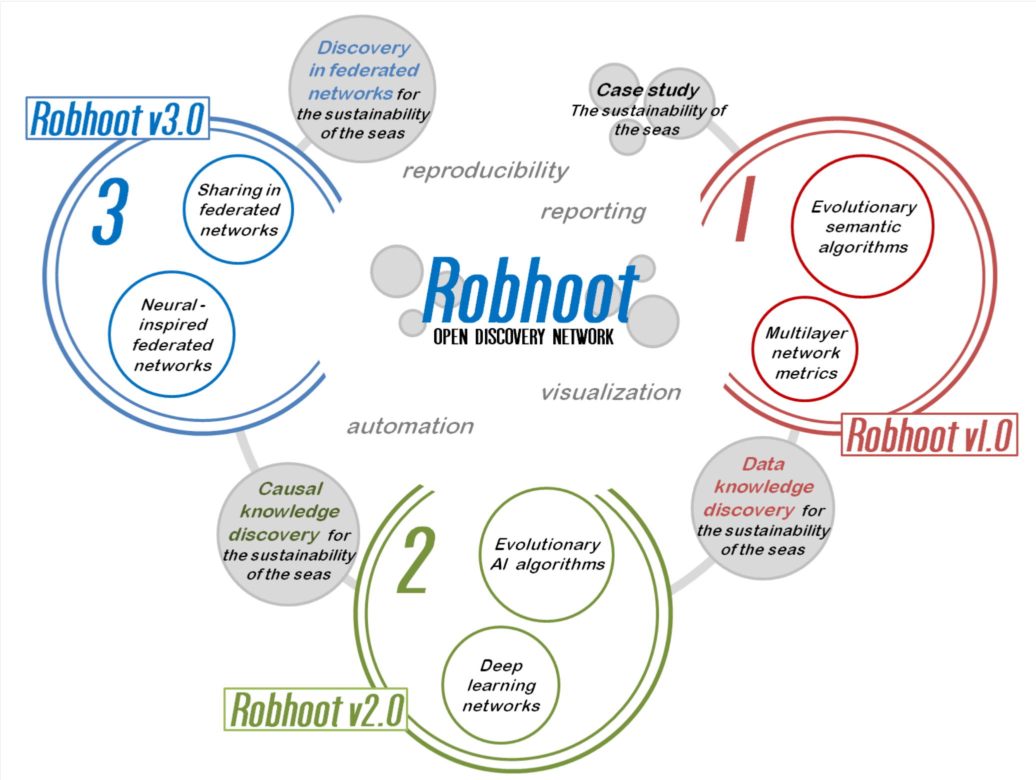

Figure 1: **Discovery in Evolutionary Biology-Inspired Federated Networks**. ROBHOOT targets knowledge discovery in federated networks. Federations are composed by highly heterogeneous groups sharing ecosystem resources for a sustainable knowledge-inspired society: ROBHOOT introduces three science-enabled technologies: Evolutionary biology-inspired semantic algorithms for ROBHOOT v1.0: data knowledge discovery (red), evolutionary-biology inspired AI-deep neural networks for ROBHOOT v2.0: causal knowledge discovery (green), and evolutionary neural biology-inspired ROBHOOT v3.0: discovery in federated networks (blue).

1.2 Science-to-technology breakthrough that addresses this vision

Data knowledge discovery (WP1)

Evolutionary biology-inspired semantic algorithms (WP1): The majority of studies of data discovery focus on add-hoc algorithms, ignoring ecological and/or evolutionary biology-inspired solutions. Currently, only a few databases are semantically annotated (e.g. , gene ontology database, COVID-19). This is because ontology development is time-consuming, requires expert knowledge and community commitment, and is 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 are still to be developed, although Semantic Web technologies are included in programs such as the U.S. National Science Foundation's proposed CyberInfrastructure (refs +++).
Going beyond ROBHOOT will go beyond state-of-the-art to implement evolutionary computation concepts (i.e., genetic algorithms with different rules and selection modes) to investigate new data properties along many data-sources. ROBHOOT will provide a detailed understanding of the replicability of accounting for many data-sources on the global data architecture map contrasting different evolutionary algorithms.

Causal knowledge dicovery (WP2)

Eco-evolutionary biology-AI-inspired discovery computation algorithms (WP2): Many studies of causal discovery (i.e., explainable or interpretable discovery), focus on genetic programming (refs

+++) and symbolic regression (refs +++), ignoring eco-evolutionary biology-inspired computation from where causal inference can be obtained. On the other side, experimental evolution data shows the importance of rapid structural trait changes beyond plasticity for new functional information processing capabilities of the interactions and traits (refs +++). The classical view on biology-inspired information processing is to consider plasticity without structural changes, or without co-evolution among many interacting components (refs +++). Recent studies indicate that rapid trait changes and information processing as a consequence of these changes 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 functional properties of populations with new information capabilities (i.e., new adaptations to new habitat or niche biotic or abiotic conditions, refs +++). Furthermore, recent studies suggest that the interplay between trait dimensionality (biotic, abiotic, migration traits, etc), and adaptation is key to understand the emergence of new traits and information processing abilities to elaborate new computation strategies in ecosystems (Box 1, and refs, +++).
Going beyond ROBHOOT will, for the first time, employ eco-evolutionary biology-inspired solutions to implement AI process-based methods to create spatiotemporal causal inference in systems containing large heterogeneity and dimensionality (Figure 2). Using the above models, this will be extended to deep process-based learning networks including trait and interactions as evolutionary changes and coevolution to implement diversification patterns in these systems. The search for causal knowledge discovery will be applied to the sustainability of the Seas case study containing 9 million entries, 1612 species (around 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 evolutionary biology-AI-inspired 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 (Deliverable D2.3, [20, 30]).

Discovery in federated networks (WP3) Integrating data and causal knowledge graphs provide a mechanistic understanding of how the balance of cooperation vs. competition might alter sustainability in our exploration of the Seas case study. However, causal knowledge graphs are not enough if they only represent isolated contributions and can not "learn to learn" to find novel, emergent solutions in neural biology-inspired networks composed by highly heterogeneous groups. In this regard, federated objects can be seen as "neural networks" containing many types of heterogeneous nodes with varying degrees of learning, connectivity and firing probabilities [23, 24]. Technologies in digital ecosystems around federated networks are scarce and mostly focus on decentralization, scalability and security fronts [8, 9, 13, 14, 19, 25]. In the science ecosystem, only a few applications of open decentralized technologies exist [21]. Yet, the discovery of novel algorithms in biology-inspired federated networks for cooperative forecasting of global sustainability problems when heterogeneous groups learn and share 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 ([20, 27]). Evolutionary biology-inspired search for algorithmic discovery can help to decipher how interactions among heterogeneous groups evolve and learn to solve complex sustainability problems. For example, evolutionary dynamics can explore open-ended language of models with varying trait evolution functions to discover biologically inspired solutions in multidimensional systems ([27],+++). ROBHOOT v.3.0 deploys biology-inspired federated networks accounting for heterogeneous agents to discover novel biology-inspired solutions for the exploration of the Seas federated network (Deliverables D3.1 and D3.2, Tables 3.1a-c)
Going beyond: Our understanding of the outcomes from evolved 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 information processing with diversification of heterogeneous and highly dimensional systems in federated networks are required to develop science-enabled technologies where heterogeneous agents with different interests find (non optimal) solutions that allow sustainable explotation of ecosystms. ROBHOOT v.3.0 con-

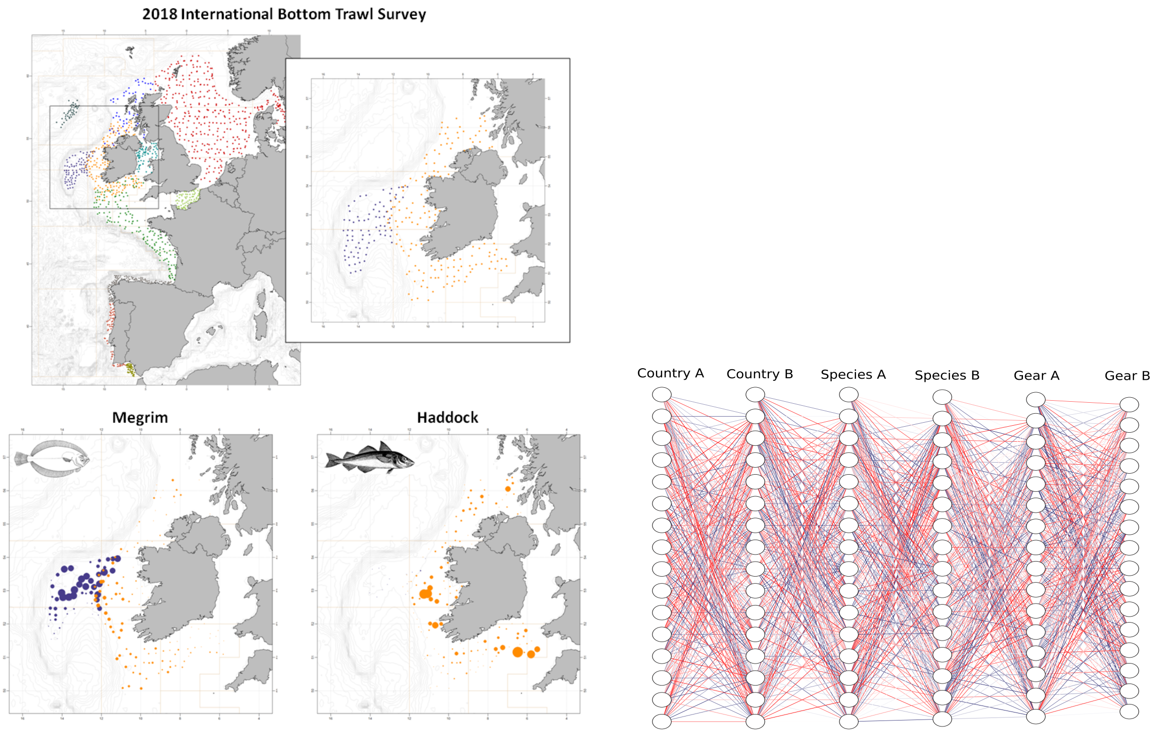

Figure 2: **Figure 2: Causal Knowledge Discovery for Sustainable Ecosystems**. **Top left**) The Irish Ground Fish Survey (IE-IGFS, Orange) and the Spanish Survey on the Porcupine Bank (SP-PORC, Blue) were part of the 2018 International Bottom Trawl Survey, coordinated by the International Council for the Exploration of the Sea [7]. Ireland and Spain use different Gears: The GOV gear has a larger vertical opening (Ireland, 3-4 m) respect to the Baka used on the Porcupine Bank (Spain, 2-3 m). This makes catchability different for fish species, such as Megrim (*Lepidorhombus whiffiagonis*, **Center left**) and Haddock (*Melanogrammus aeglefinus*, **Center right**), in which both countries have very different commercial interests. Haddock is a species of the cod family, highly prized in northern Europe, while Megrim is a species of flatfish, consumed largely in Spain and France. Spain catches Megrim better than Haddock and viceversa for Ireland. This generates a strong bias in the distribution maps (compare Megrim vs. Haddock map, Center) with potential implications for biodiversity management and sustainability in natural ecosystems. **Right** Causal knowledge discovery graph representing the 2-countries, 2-species and 2 gears for the example above. The whole data set for 2018 contains 11 countries, 461 fish species (approx. 200k individuals sampled), and 5 gears. Each country, species and gear is composed by many nodes: For example 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.

nects knowledge discovery to biology-inspired federated netwoks to study the properties of cooperative forecasting and strong inference in the face of global sustainability and biodiversity challenges (Figure 2 and Tables 3.1.a-c).

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, including evolutionary biology, ecology, computational neuroscience, computer science, data science, complex systems and experts in biodiversity sampling methods and the infrastructure related to international protocols for sampling the Seas. Data knowledge discovery gained by analysis and modelling of the computation discovery capabilities of evolutionary-inspired semantic algorithms by the evolutionary biology, computer science and complex system members of this consortium (EBD-CSIC, IFISC-CSIC, SDSC) can be transferred to the causal domain addressed by the other part of the consortium with expertise in evolutionary biology, data science and causal inference

(EAWAG and TARTUR). This will be enriched with full automation, reproducibility and visualization supported by ICREA, SDSC, and our company-partner (SME), 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 traits and interactions, cooperation, learning functions, and dimensionality) encountered 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 work packages while keeping functional interactions among the groups to run efficiently the different stages of the project. To bring together adaptive biology-inspired semantic algorithms for data discovery and evolutionary-neurobiology-inspired discovery in federated networks requires 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 biology-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. The transfer of eco-evolutionary biology-inspired principles onto fully reproducible and automated software, progressing from fragmented- and pattern-based to integrated- and 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 gradually increasing in complexity of the problems addressed. Figure 3 shows that we will start with evolutionary biology-inspired semantic algorithms for data discovery in the context of the Sustainability of the Seas case study. This is followed by investigation and implementation of more complex eco-evolutionary biology-inspired AI and deep learning network modeling to infer causality in the sustainability of the Seas case study (i.e., dimensionality, nonlinearities). Then we will advance to more complex situations, where the evolutionary neurobiology-inspired modeling will expand the search along many learning and cooperative forecasting schemes to find scenarios for the sustainability of the Seas 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 unified science-enabled technology. This leaves open the option to work with fast computing languages to develop low-level Agent Based Models along all the theoretical development of the proposal (i.e., Julia, C++), instead of differential/difference equations methods when a large number of agents and interactions change in time and space. This feature represents a very desirable fallback in case of speed and convergence problems for multidimensional and nonlinear systems (Table 1.41 Critical risks for implementation). Our implementation activities are all complemented by numerical investigations contrasted for speed and robustness with the sustainability of the Seas case study started in 1965 and containing around 9 million entries, 1612 species, 20 countries and 11 sampling methods (Figure 2). The success of ROBHOOT would represent a breakthrough in the current discovery computation with direct application to sustainability of ecosystems. It exploits eco-evolutionary biology-inspired computational capabilities of evolving traits and interactions to discovery and transfers their properties to natural ecosystems. The combination of rapid, data heterogeneity and cooperation for discovery computation based on, mostly open-source languages, 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	Consider more developed genetic programming methods to infer data interactions.
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	P2	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 targets novel approaches towards sustainable ecosystems. One of the tasks in WP3 focus on the discovery of novel evolutionary neurobiology-inspired algorithms to provide results for sustainability fisheries. Solutions around WP3 ultimately depend on merging WP3 with the rest of WP's in the proposal. For example, it is known that sustainable ecosystems strongly depend on many data sources collected by different groups using different technologies (refs +++). ROBHOOT discover data interactions combining fisheries, stakeholders, and technology data, the data knowledge discovery graph, as a first step towards the discovery process. ROBHOOT also infer the technological and environmental changes and the processes underlying the empirical patterns, the causal knowledge discovery, to provide the existing sustainability status in a human-disturbed ecosystem. Altogether, this project will lay the foundation for future sustainability