ROBHOOT
Knowledge Discovery in Evolutionary Biology-Inspired Federated Networks
v.3.0

May 20, 2020

Contents

Summary

Global sustainability is a major goal of humanity. Many studies have shown global sustainability could be achieved by strengthening transparency, communication, and rapid access to reproducible information among social, ecological, economical, technological and governance systems. Sustainability goals, however, strongly depend on global access to evidence-, and discovery-based knowledge gaps. Yet, science-enabled technologies targeting knowledge discovery to reach sustainability and biodiversity conservation goals are not in place. We propose a evolutionary biology- and AI-inspired knowledge discovery technology for a sustainable-, and knowledge-inspired society. We introduce evolutionary biology- and AI-inspired solutions to account for large and heterogeneous groups in federated networks to explore sustainable scenarios for the international exploration of the Seas. Knowledge discovery running on a federated network encompass a hybrid-technology to lay out the foundation of an open- and cooperative-science ecosystem to automate discovery in global emergency and sustainability challenges. The project summarized here is not set out to deliver automated discovery knowledge in federated networks, but to provide the architecture of a science-enabled technology, as a proof-of-principle, to connect global human sustainability challenges to knowledge-inspired societies.

1 Excellence

1.1 Radical vision of a science-enabled technology

We are in the midst of the fourth industrial revolution, a transformation revolving around data-driven intelligent machines and knowledge-inspired societies. More than half of the global population is now online using the Internet (i.e., 3.9 billion), which represents a more inclusive global information society (+++). The Internet is rapidly evolving and people are using technology in powerful ways, from adopting decentralized technologies for humanitarian efforts to improving agricultural practices and reducing waste in the global food supply chain ([1],+++). Data analytics is advancing at a pace dictated by the availability of data. A myriad of data-driven approaches are being developed to extract patterns from data ([2]). At the same time, data analytics is being challenged because the diversity of data sources keeps rising (+++). Consequntly, Artifical Intelligence (AI) approaches are rapidly evolving towards more explainable/interpretable pattern inference (+++). The digital ecosystem has to deploy science-enabled technologies that account for the increasing data heterogeneity and interpretability. However, science-enabled technologies accounting for these two features are scarce [3].

Taken together, the transformation of a digital society into a knowledge-inspired society requires solving several gaps. First, the science-enabled technological paradigm for assisting humans is biased towards a limited range of the "observable" heterogeneity in data-sources. Thus, it limits the number of interpretable patterns (+++). Second, the AI technological paradigm is rooted in single- and multiple-objective optimizations (i.e., function loss or reward, similar to fitness optima functions in evolutionary biology +++). Optimization-based technologies have produced a great deal of progress, yet, they limit a broader number of sub-optimal but plausible solutions, as usually found in evolving biological systems (+++). And third, science-enabled technologies for scientific inquiries are highly fragmented, only partly reproducible, and mostly developed in close-source software ([4, 5, 6, 7, 8, 9, 10]). To leverage the abundance and heterogeneity of data, a) a science-enabled technology should be able to obtain information from a large pool of heterogeneous data-sources, b) the analysis of the data should go beyond the identification and interpretation of patterns, and towards the discovery knowledge gap and to the end-user, c) the analysis should be performed in a federate way, such that highly heterogeneous populations can learn from each other to reach consensus about the population of plausible scenarios accounting for data heterogeneity and dimensionality, and d) the whole process should be automated, reproducible and transparent such that can be improved to benefit the public. Our project contributes towards a discovery science-enabled technology, where novel rules and interactions are exposed to federated networks inspired by evolutionary biology and neural biology (Figures 1 and 2 and Table 1). Evolutionary dynamics-inspired technologies allow studying consensus algorithms within and between groups to enrich knowledge-inspired societies facing global sustainability problems. They do so through extracting information from highly heterogeneous and multidimensional groups while minimizing the need of having optimal solutions. This is particularly relevant when discovery is obtained from heterogeneous data-sources to gain information of complex governance, social, environmental and technological problems (Figure 2) [11].

Many experimental evolution and evolutionary computation models have shown the plausibility of coexistence of multiple heterogeneous populations (+++). Many interpretable mechanisms have been proposed to explain such a coexistence, like negative-frequency dependent selection (Doebeli book and others, +++), fluctuating-selection, and many others (+++). Yet, approaches accounting for not optimal or maladaptive solutions in the context of group heterogeneity in multidimensional landscapes are rare (refs around evolution cooperation in multidimensional landscapes, +++). In ecological systems, intraspecific trait variation (i.e., a proxy for heterogeneity within a species) and trait dimensionality (i.e., biotic, reproductive, abiotic and migration traits for example) can drive functional interactions with other species (i.e., cooperative, antagonistic, competitive, or mutualistic), but most approaches have neglected the effect of trait dimensionality like competition and cooperation traits in heterogeneous populations **(On neural systems, the vast majority of neurons in the brain show highly differentitated morphological, genetic and phenotypic states? (refs, Wolfgang))**. Therefore, the understanding of functional interactions among such a highly differentiated states (groups, etc) capturing the observed coexistence patterns in ecological systems is not well understood. Taken together, these results suggest that our understanding of evolved information processing systems that are formed by highly heterogeneous groups (refs about federated networks, bacterial consortia, federated bacteria..., artificial life, problem solving artificial societies, and large-scale meta-learning in the federated setting [12]), is currently quite limited. This suggests that new science-enabled technologies accounting for diversification, dimensionality and heterogeneity of highly distinct groups are required to decipher functional information processing in federated networks following the increasing demand of reproducible discovery in knowledge-inspired societies.

Biodiversity data collected by many different countries is a good example for understanding open-problems in heterogeneous federated networks. Many international programs for exploration of the seas involve many countries collecting biodiversity data using, despite efforts of standardization, different protocols and technologies ([13]). The data is then used to understand the spatiotemporal dynamics of the ecological communities as a baseline to inform fisheries (+++). Each country collects data with different gear systems (Figure 2) because of their commercial interests in specific species. The result is that countries use different gear systems and collect heterogeneous and biased data about the same species, making it difficult to obtain accurate distribution maps of species (Figure 2). This situation can be outlined as follows: country having interest in specific gear systems vs. having shared interest using standardized gears to share more accurate species and communities maps (i.e., a problem similar to the tragedy of commons, +++). This last one strategy is built on cooperation between two countries to understand better a specific species while sacrificing their own commercial interest (Figure 2). This is a common situation when many heterogeneous nodes (i.e., countries with different interests, groups, funding and conservation strategies, etc) exploit resources (i.e., species within ecosystems compossed by a network of interacting species compossed by heterogeneous individuals within and between species, food webs, mutualistic networks, etc) using different technologies (i.e., gear systems). Many of these ecosystems are overexploited and yet science-based technologies providing forecasting scenarios accounting for heterogeneous biodiversity data (i.e., species and environment), sampling protocols (i.e., gear systems and other technologies), and groups with different interests within and between countries to mitigate risks and enhance global cooperation scenarios in such a multidimensional ecosystems are not in place ([1], +++).

The example about the exploration of the seas teaches us the need of science-enabled technologies facilitating discovery from heterogeneous data-sources, groups and technologies to overcome fragmented and partial responses to a global sustainability problem. The goal of *ROBHOOT* is to propose a compact science-enabled technology integrating heterogeneous data and mechanistic inference into evolutionary biology-inspired federated networks to lay the foundation for a novel scientific discovery technology. *ROBHOOT* also contributes towards reproducible and automated cooperative forecasting scenarios in rapidly changing global sustainability landscapes (Figures 1 to 3 and Impact section): *ROBHOOT* **v.1.0** discover data knowledge graphs obtained from semantic evolutionary algorithms using heterogeneous data-sources. *ROBHOOT* **v.2.0** discover mechanisms from evolutionary biology-inspired deep learning networks, and *ROBHOOT* **v.3.0** expands knowledge discovery along evolutionary neural biology-inspired federated networks (Tables 3.1a-c).

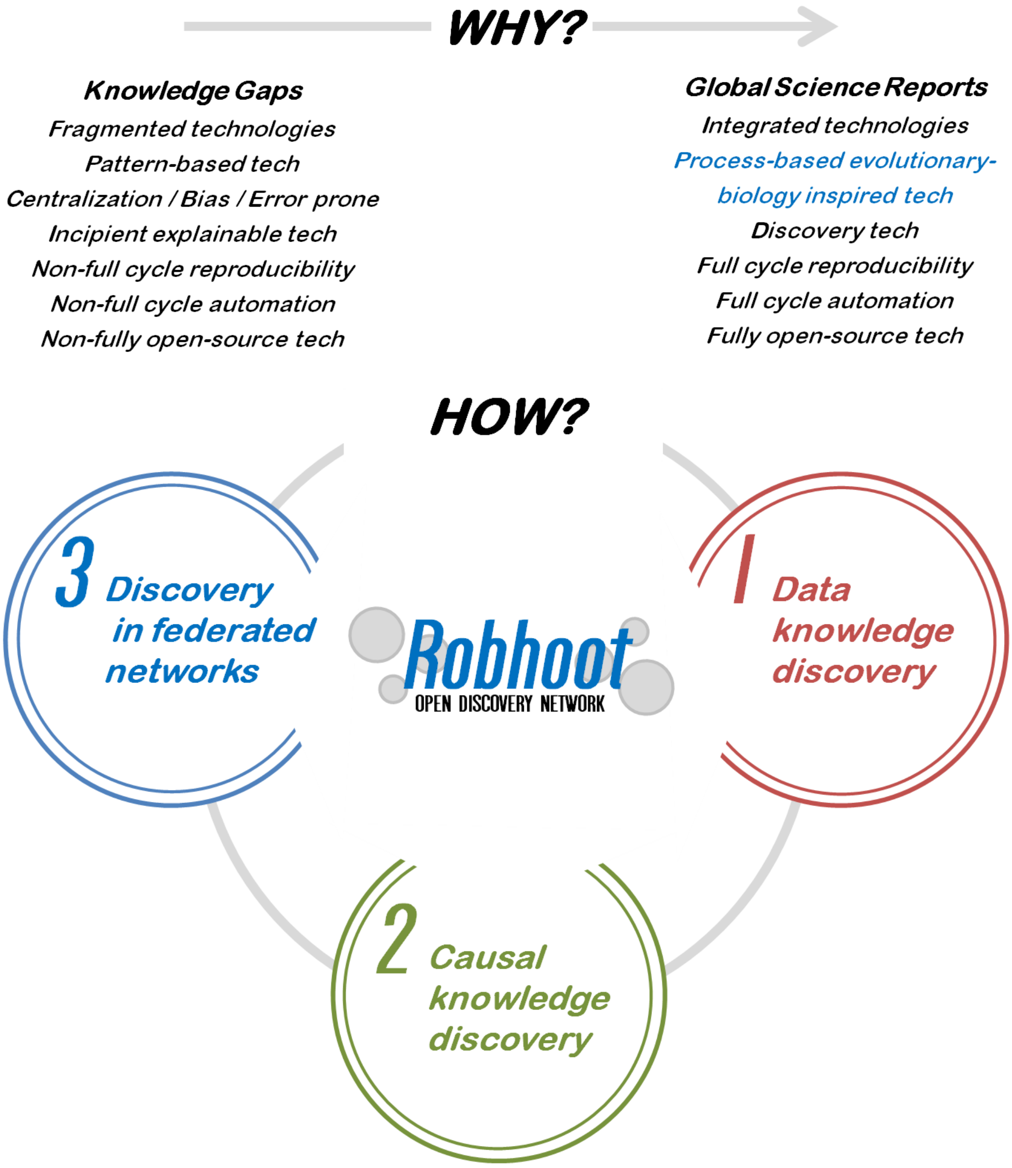

Figure 1: Discovery in Evolutionary Biology-Inspired Federated Networks. *ROBHOOT targets knowledge discovery in federated networks, networks composed by highly heterogeneous groups each with different interests in ecosystem resources, for a sustainable knowledge-inspired society:* **Top)** *ROBHOOT targets global knowledge gaps for open-access reproducible global discovery reports. ROBHOOT introduces three science-enabled technologies:* **Bottom)** *Evolutionary biology-inspired semantic and multilayer network algorithms for* **data knowledge discovery** *(red), evolutionary-biology inspired AI-deep neural networks for* **causal knowledge discovery** *(green), and evolutionary neural biology-inspired* **discovery in federated networks** *(blue).*

1.2 Science-to-technology breakthrough that addresses this vision

Interconnected global societies are constantly facing new challenges that need to be rapidly addressed. In this regard, discovery technologies providing plausible scenarios when solving complex governance, social, environmental and technological problems are lacking. Depite rapid advances of research platforms for data analytics in the last decade [14, 15, 16, 17, 18, 19, 20, 21, 22], the integration of science-enabled technologies currently lack discovery knowledge-inspired approaches impacting knowledge-inspired societies to help responding to the rapidly changing global sustainability challenges (Figures 1 to 3, and Tables 3.1a-c). Discovery technologies facilitating the understanding of complex systems still present many challenges (+++). This is particularly relevant in global sustainability landscapes, where data heterogeneity like varying sampling efforts, sampling methods, data fragmentation, and lack of transparency limit our understanding of empirical patterns to draw forecasting scenarios (Figure 2). In addition, complex ecosystems are rooted in evolving information processing systems driven by many factors and highly heterogeneous living groups. Therefore, despite many years of research and insights about complex systems (ref +++), our understanding of evolutionary biology-inspired solutions to address complex systems is quite limited

Discuss the relevant state-of-the-art and the extent of the advance the project would provide beyond this state-of-the-art. How will *ROBHOOT* go beyond state-of-the-art? *ROBHOOT* introduces evolutionary biology-inspired discovery in federated networks to make discovery an evolving feature in digital ecosytems (Figure 2). How will *ROBHOOT* explicitly deal with diversification and dimensionality when accounting for highly heterogeneous evolving groups and interactions? (refs about federated networks, bacterial consortia, federated bacteria..., artificial life, problem solving artificial societies, and large-scale meta-learning in the federated setting [12]). Describe the science-to-technology breakthrough, targeted by the project that would represent the first proof of concept of the envisioned technology. Patterns from knowledge-graphs are emerging at a fast pace in specific frontiers +++, but remains isolated from the discovery process especially in the context of cooperative forecasting in federated networks +++. *ROBHOOT* goes beyond the state-of-the-art of knowledge-graphs by fussioning data and causal knowledge graphs, and scalating these to evolutionary biology-inspired federated networks to move knowledge-inspired societies towards reaching global sustainability goals when large number heterogeneous groups share multiple resources driven by multiple factors.

Describe briefly the goals, how are we going to achieve them? *ROBHOOT* **v.1.0** deploys a data discovery technology to generate data knowledge graphs accounting for heterogeneous data-sources (Tables 3.1a-c and Figure 3). Data-architecture alone is not sufficent to outline predictive scenarios in complex sustainability problems (refs +++). Therefore, data analytics complementing data-architecture discovery is desirable to interpret scenarios in natural and digital ecosystems. In this regard, there are also many gaps in connecting data-architecture assembled from many sources into rapid and automated causal knowledge discovery. *ROBHOOT* **v.2.0** introduces automated and explainable evolutionary biology-inspired AI methods to decipher causal knowledge discovery from open-ended modeling scenarios. Still, rapidly drawing scenarios from a few labs limit the parameter phase space from where the discovery process is generated. Therefore, the scalability of fully reproducible discovery strongly depend on cooperation and learning in large scale biology-inspired federated networks. *ROBHOOT* **v.3.0** brings knowledge discovery to federated networks by introducing evolutionary biology heterogeneous-neural inspired networks to obtain cooperative forecasting scenarios from data and causal knowledge discovery (Figures 2 and 3).

1.3 Interdisciplinarity and non-incrementality of the research proposed

Explain why the proposed research is non-incremental. Describe the research disciplines necessary for achieving the targeted breakthrough of the project and the added value from the interdisciplinarity *ROBHOOT* is a science-enabled multi-feature technology for interpretable data-driven discovery in federated networks (Figures 1 to 3 and Tables 1, 3.1a-c, and 3.2a). It contains three work packages each characterized by a mixture of research disciplines. *ROBHOOT* **v.1.0** is composed by computer scientists, evolutionary biologists, and developers targeting novel evolutionary inspired semantic algorithms for data discovery. This module is complemented with scientists from complex networks taking care of quantitative methods in the data knowledge discovery to decipher the existing gaps in data discovery and data-architecture technologies (section 3.1.1). *ROBHOOT* **v.2.0** is compossed by data-scientists trained in deep learning networks and automation algorithms, theoreticians and evolutionary biologists with expertise in modeling mechanistic and Bayesian networks and biology-inspired neural networks, respectively. The combination of data-scientists, theoreticians and biologists generates a diverse team targeting synthesis between automated and explainable evolutionary biology-inspired approaches to decipher causal knowledge discovery from heterogeneous data-sources (section 3.1.2). *ROBHOOT* **v.3.0** combines computer scientists and developers targeting evolutionary neural biology-inspired models of federated networks, with social scientist, and scientists specialized in ecology and evolutionary biology (section 3.1.3). The complementarity of the teams in modules one to three makes *ROBHOOT* flexible and a science-enable functional technology in a rapidly evolving digital ecosystem [23].

ROBHOOT aims to bring global transparency in knowledge discovery generation by acting as an assistant or as an automated and reproducible discovery generator to facilitate sustainability goals in ecosystems. The multi-feature, science-enabled technology target a reduction in global knowledge gaps while transparently accounting for centralization [4, 7], bias [5], error-prone [6], and non-reproducibility [8] (Figures 1 and 2 and Table 1). Improving these features are mostly due to the rapidly evolving digital ecosystem. For example, it is increasing continuously its computing capacity, new methods integrating automated and explainable AI are rapidly advancing, and their interconnection to open-source technologies is also rapidly occurring in the digital ecosystem +++. Yet, targeting automated discovery into federated networks still require taking risky steps. It requires combining evolutionary biology-inspired solutions crossing data, causal and novel discovery in complex and heterogeneous data-sources and bring them to a global scalability...

1.4 High risk, plausibility and flexibility of the research approach

- Explain how the research approach relates to the project objectives and how it is suitable to deal with the considerable science-and-technology uncertainties and appropriate for choosing alternative directions and options. (The risks and mitigation plan should be spelled out under the Implementation section).

Knowledge-inspired societies and governance demand full research cycle transparency, reproducibility and interpretability to inform complex social, environmental and technological problems in the face of global sustainability challenges. Such needs bring many technical and functional challenges to our research proposal because obtaining robust knowledge from integrating many parts each containing its own set of methods can generate divergent, fragile and contradictory outcomes. To solve this issue, *ROBHOOT* will have a modular and flexible structure following three main work packages (Tables 3.1a-b) with a total of ten deliverables (Table 3.1c), and three main milestones (Table 3.2a)...

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-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 graph, to provide the existing sustainability status in a human-disturbed ecosystem. Altogether, this project will lay the foundation for future sustainability studies. Discovery of novel evolutionary-inspired algorithms 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. The discovery of new solutions not observed in the empirical data, but containing the plausible scenarios of maintaining high values of biodiversity and sustainability, will be the basis for estimation of the severity of overfishing and sampling bias when many groups enter in commercial conflict of interest... Such a targeted sustainability proxies would be of great interest not only for the biodiversity maintenance but also from an economic and social point of view, as it would save costs for future generations. Sustainability challenges are related to the development of future sustainable societies, which according to (Organization) **Keep elaborating**

- **Potential for future social or economic impact or market creation**:
Collapse of ecosystems can lead to serious long term economic and ecological disfunctionalities (refs +++). However, there are not well established metrics for the characterization of sustainability in complex ecosystems. Our approach accounts for heterogeneous sources of data, the (evolving) mechanisms underlying technological, environmental and social changes required to make ecosystems sustainable and novel rules that could impact positively the maintenance of biodiversity by developing cooperative forecasting strategies among the many (international) groups involved. Such a risk assessment would not only be of great interest to the groups exploiting the resources, but also from an economic and ecological point of view, as having less bias in the field data provides more accurate measures from the observed time series for planning fish stocks for a large number of speciesq. Finally, *ROBHOOT* contributes towards knowledge-inspired societies in need of radically tackling new societal and global environmental challenges: it provides reproducible and transparent methods for making sustainability goals achievable and reproducible across many sectors and economies.
In the medium-term this technology may also have interesting applications in public and private industry. For example, access to discovery with cooperative forecasting might suggest new paths and solutions that are key to generate rapid and robust scenarios when facing complex problems including global sustainability challenges (i.e., global health, ecosystems degradation, biodiversity loss, etc). First, evolutionary biology-inspired AI algorithms deciphering open-ended search of interpretable mechanisms underlying the targeted complex systems for private and public industry facing highly heterogeneous data sources. Second, cooperative forecasting challenges existing fragmented responses to emergent global sustainability problems by compactly offering reproducible forecasting emerging from many-to-many human and machine cooperative discovery, and third, open-access explainable and automated information generation account for global data-arquitecture allowing individuals and companies to address scenarios of future strategies in highly fluctuating local and global market conditions.
- **Impact on transparency and reproducibility**:
Decision making and governance at local, regional and global scales require access to transparent and reproducible information containing the interpretable factors and their plausibility to explain the empirical patterns. In this regard, the *ROBHOOT* consortium brings together excellent partners from the fields of computer science, neurobiology, complex system, biology, social sciences, evolutionary ecology and including one SME focusing on reproducibility, automation, visualization and reporting along its whole developmental life cycle (Dissemination plan below and Figure 3).
At the same time, all groups composing the consortium exhibit a long-standing experience interdisciplinary research across the boundaries of the individual disciplines (Figure 3). The subsection on related projects shows that this is a novel constellation in Europe and possibly worldwide (section 4). This consortium is also at the leading edge of developing novel evolutionary biology-AI inspired solutions to automation and reproducibility in complex systems facing sustainability challenges.
- **Ecosystem health impact**:
Ecosystem sustainability and ecosystem health are usually used as metaphors to describe the mechanisms that maintain functional and diverse systems and the condition of an ecosystem, respectively. Ecosystem sustainability and condition can vary as a result of many disturbances like fire, flooding, drought, extinctions, invasive species, climate change, mining, overexploitation in fishing, farming or logging, chemical spills, and a host of other reasons. *ROBHOOT* focus on novel discovery solutions for ecosystems under a varying degree of disturbances. *ROBHOOT* introduces a case study for overexploited ocean ecosystems when highly heterogeneous social groups with different interests exploit limited and shared resources. Thus, *ROBHOOT* is a technology designed to provide novel discovery solutions paths for ecosystem sustainability, improving the underlying discovery paths that allow draw novel connections between ecosystem sustainability and ecosystem health. This feature aligns to the EU Reflection paper towards a Sustainable Europe by 2030 focusing on the need of investing in sustainable growth and spur action by goverments, institutions and citizens, leading the way for the rest of the world using the UN's

Sustainable Development Goals (SDGs). Specifically, *ROBHOOT* can be seen as an horizontal enabler for the sustainability transition to make Europe sustainable by 2030. It introduces evolutionary biology- and artificial intelligence-inspired solutions to benefit ecosystem health and people's lives and work. By being able to process large amounts of heterogeneous data instantaneously, artificial intelligence and evolutionary-biology inspired solutions have the potential to significantly increase productivity in environmental sustainability and ultimately make informed decisions to enhance food security [24].

- **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 and in particular evolutionary biology-inspired federated networks both from a theoretical and an experimental point of view, Physics, theory and applications of complex systems in social networks and one highly innovative science-based reproducibility, automation, reporting and communication focusing on sustainability solutions. Many of the components of the consortium are first-time participants to FET under Horizon 2020 (Section 4). The use of advanced evolutionary biology-inspired and complex networks-based analyses to characterize and predict novel discovery in systems formed by heterogeneous and evolving groups and interactions combined with the implementation of intelligent learning discovery in federated networks and the development of a reproducible and automated protocol user friendly interface go much beyond the current state-of-the-art in science-based discovery technologies. All consortium partners exhibit a long-standing experience in interdisciplinary research across the boundaries of the individual disciplines (Figure 3). The subsection on related projects shows that this consortium is at the leading edge of innovation and interdisciplinarity (Tables 3.1a-c). A significant value proposition of the project is to increase the research on large-scale sustainable federated networks where many heterogeneous agents share resources embedded in complex ecosystems. This will produce valuable information and data about how federated networks work under broad set of socio-ecological scenarios, similar to natural ecosystems consoritiums where many paths produce coexistence of heterogneous poulations and high biodiversity (refs ++). It is important to consider that all ecosystems facing many human pressures are all across the world and discovery technologies facilitating the solutions in large-scale federated networks could inspire new developments improving our understanding of sustainability at global scale. For in-home, we also expect an explosion of discovery knowledge approaches and future publications, which will place Europe at the top of sustainability in federated networks.
  Moreover, in WP3, we propose the generation of a web-based sustainability discovery portal that will allow researchers, NGO, managers and the public to train students in the discovery process to manage over-exploited ecosystems, allowing to scale up the number of people participating in the sustainability process by an order of magnitude thus mobilising forward thinking researchers and excellent young researchers to work together and explore what may become a new technology paradigm in sustainability research. Members of the consortium already have experience in generating such types of training tools that are currently available online (check github repository RobhooX). This approach would provide an unprecedented capability for the access to a multitude of people interested in sustainability discovery tools that will result in facilitating consensus and a valuable source of information for science-enabled technologies in ecosystem sustainability and management.

2.2 Measures to maximize impact

2.2.1 Dissemination and exploitation of results

- **The Plan for diseminating and exploiting the project results** *ROBHOOT* allocates three research groups along its whole developmental life cycle to guarantee dissemination, transparency and easy exploitation of the technology **(when)**. **(what)** The three milestones of the project, data knowledge discovery, causal knowledge discovery and discovery in federated networks (Table 3.2a) will be fully automated and reproducible to facilitate visualization, reporting and full scalability. **(who)** Automated discovery will be implemented along Bayesian machine scientist to facilitate open-ended search during the development of the three milestones (Deliverable D3.4, *BSM*, *ICREA*, Tables 3.1a-b). Reproducible knowledge discovery graphs will be developed in the Renku open-source software (Deliverable D3.5, *DATAR* by the Swiss Data Science Center, *SDSC*). Visualization and reporting will be fully implemented in the Julia computing language for its speed and unique features (Deliverable D3.6, *VISUR*, by *DESANTANA*. Codes will be available in the public git Robhoot repository. Having the whole developmental life cycle as reproducible and automated knowledge discovery graphs facilitates the reuse and the dissemination of the technology as a whole in any platform and OS. Full reproducibility, automation, visualization and reporting provide to *ROBHOOT* legal and financial transparency and reproducibility in social governance a feature for easy replication of the discovery process by third parties, a property that can be used to facilitate reporting for governance public policy, NGO, society and thinktank in the face of local and global sustainability challenges. **why, how and which journals, conferences and with which preliminary results**.

- All the data, codes and outputs generated during *ROBHOOT* development will be open access stored in public git repositories. The project will collect data from many sources (i.e., fisheries, environmental and social data, technology data). generate data knowledge discovery graphs, causal knowledge graphs and the data and algorithms generated from the discovery in federated networks for the exploration of the Seas case study (Deliverable D1.2, *DATAX*, D2.3, *DIX*, and D3.3, *DIFEX*, respectively). **Keep elaborating**

2.2.2 Communication activities

- The full open-source developmental life cycle strategy of reproducibility, automation, and reporting generation of *ROBHOOT* targets the search of societal relevance and long-term economic impact of open and transparent science. Underlying to this strategy is to build support for future research and innovation funding, by ensuring uptake of results within the scientific community, and opening up potential business opportunities for novel products or services, and potentially contributing to better decision-making processes and valuable input for public policies formulation. *ROBHOOT* has very general dissemination 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 (stakeholder consultation processes; workshops to raise awareness, etc.) through:
  . The project website is to be set up within the first three months of the project. There is already a public git Robhoot repository.
  . Up to date information material, e.g. brochures, presentationslides, will be distributed at events to increase awareness about the *ROBHOOT* project.
  . General other publication means will be used such as newspapers, YouTube, TV and radio, social networks as well as targeted mailing lists (e.g., evodir, AI-worldwide).
  Scientific publications for the scientific community. We will target high-level journals with open access (i.e., Science, Nature Communication, etc.)
  . The consortium will visit conferences in the related scientific fields and interdisciplinary conferences in order to interactively present and discuss our results with others researchers, groups and institutions. Among other activities, the consortium will organize special sessions at several conferences in different countries. Additionally, some targeted, specific dissemination actions will be considered: We will organize hackatons and robhacks activities to attract multipliers and developers from the open-source community to the community who engage in data processing and build hybrid evolutionary biology-inspired and AI algorithms. This will be achieved by a "traveling salesman" approach using personal visits and invitations to demonstrate how *ROBHOOT* works. At the end of the project we will organize a workshop specifically on "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.
  . *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 Biodiversity management.

- *ROBHOOT* strictly adheres to the Open Access Policy of the Commission and all publishable (non-protected) results will follow the green or gold OA policy. Software as well as hardware protocols will be made openly available through standard computer science repositories. The *ROBHOOT* public git repository is already active Robhoot. Data (measured data), as such, will not be acquired by *ROBHOOT*. 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. Open access to publications will be granted under the terms and conditions laid down in the Grant Agreement, in accordance with the Rules for participation and dissemination in Horizon 2020. The beneficiaries will deposit an electronic copy of the published version or the final manuscript accepted for publication of a scientific publication relating to foreground in an institutional or subject-based repository at the moment of publication, e.g., via the OpenAIRE portal (www.OpenAIRE.eu). In addition, beneficiaries will make their best efforts to ensure that this electronic copy becomes freely and electronically available to anyone through this repository (i.e., that it becomes "open access"): immediately, if the scientific publication is published "open access", i.e., if an electronic version is also available free of charge via the publisher, or within 6 months of publication.

3 Implementation

A technology deciphering data and causal knowledge discovery to tackle global sustainability challenges is highly informative by itself, but a diverse group of scientists across Europe have decided to connect discovery and sustainability broadly. We want to advance cooperative discovery in the global digital ecosystem. To this end, the *ROBHOOT* consortium aims at integrating data and causal knowledge discovery, into evolutionary-biology and neural-inspired federated networks to provide cooperative forecasting scenarios for sustainability challenges. *ROBHOOT*'s goals are developed along three main work packages, three milestones and twelve deliverables (Tables 3.1a-c).

3.1 Research methodology and work plan , work packages, deliverables

3.1.1 WP1: *ROBHOOT* v.1.0: Data Knowledge Discovery

Fortuna and Eguíluz:: Storing and organizing much of the rapidly accumulating scientific information in rigorous, principled ways, so that finding what we want and understanding what is already known has became exhausting, frustating, stressful and increasingly costly experiences (refs +++). For information to be usable, it should be stored and organized in ways that allow us to access it, to analyze it, to annotate it and to relate it to other information and make the whole process reproducible. The first step is, therefore, to translate the scientific knowledge stored in existing databases to machine-readable format. Only a few databases are semantically anotated (e.g. , gene ontology database). 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. 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 +++).
The approach that we are adopting, therefore, begins with the raw data, ontologically interprets and transforms the data in order to extract its semantics and express such semantics in ontologies. These ontologies are then mapped to a graph-based data architecture. Neo4j is the implementation technology that we have currently

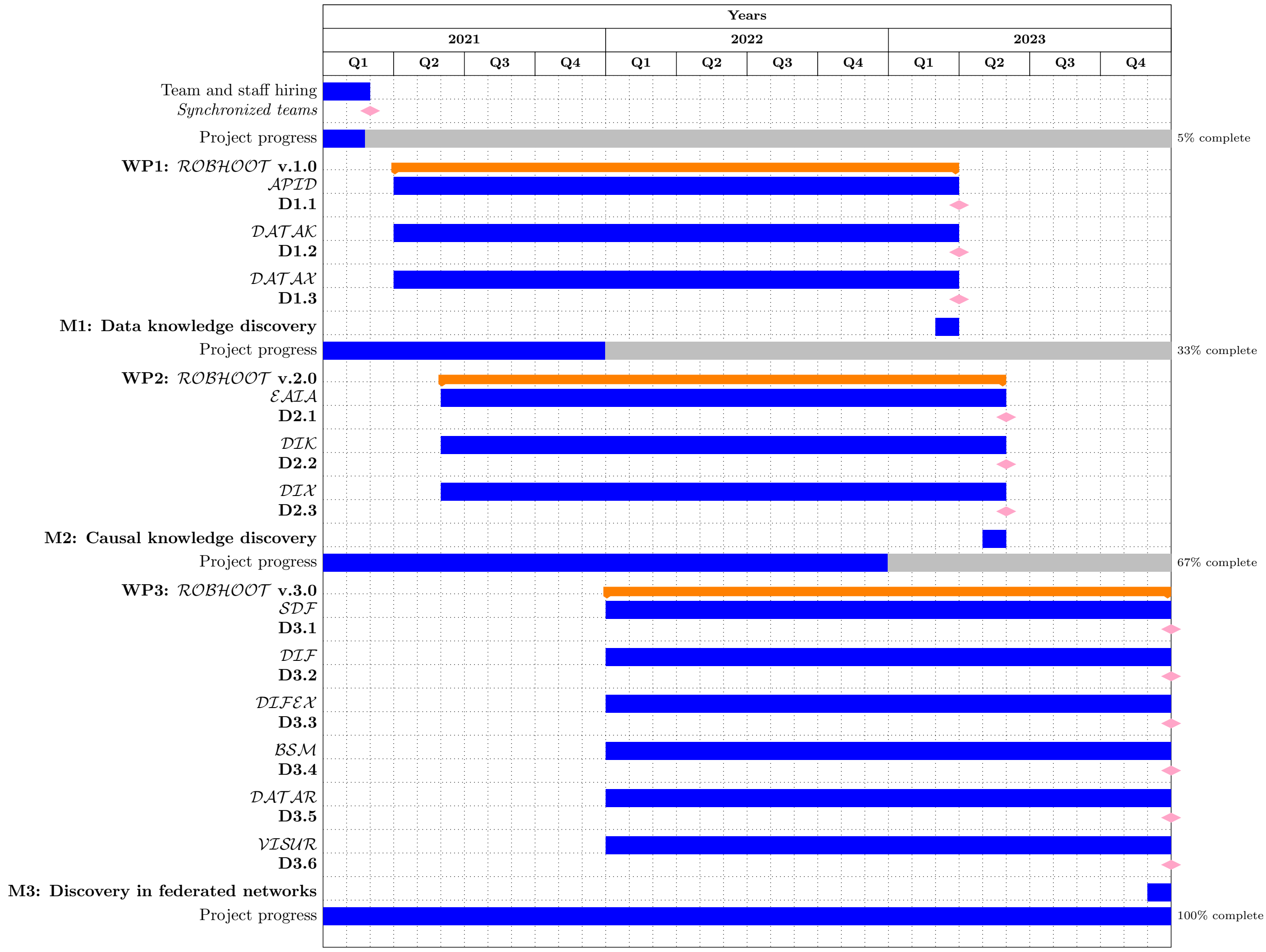

Table 3.1a: *ROBHOOT* **Gantt Chart**: *Work package one,* *ROBHOOT* **v.1.0** *introduces deliverables* **D1.1** *to* **D1.3** *for data knowledge discovery for the exploration of the Seas case study (Figure 2).* *ROBHOOT* **v.2.0** *introduces deliverables* **D2.1** *to* **D2.3** *for causal knowledge discovery for the exploration of the Seas, and* *ROBHOOT* **v.3.0** *introduces* **D3.1** *to* **D3.6** *for discovery in federated networks for the exploration of the Seas. Reproducibility, automation, visualization and reporting deliverables are developed along the full life cycle of* *ROBHOOT*.

chosen to adopt. The main reasons for adopting a graph database to persist the ontological models are: (1) the flexibility that a graph structure provides in implementing any modelling paradigm and (2) the scalability it provides in terms of organizing and accessing massive amounts of data.

Next, we need to integrate the diverse and heterogeneous large-scale data sets that we have semantically annotated previously since they originate from distinct communities of scientists in separate subfields. This will be achieved in large part through the adoption of novel technologies. The few technologies and available tools are still quite limited for many scientists and only a few research groups have tried to apply it on real data (refs +++). It is rather difficult to ask simple questions like, for example, which human diabetes-related proteins are located in the nucleus of the cell and are interacting with proteins related to pancreatic cancer (refs ++ example on sustainability?). Such question requires the integration of data from multiple sources. New approaches are needed to make progress in this direction.

(edited from here) API discovery technologies to build robust and scalable automated data-driven discovery is an existing need [25, 26]. Technologies around building database are particularly relevant to move beyond explainable (or interpretable) Artificial Intelligence technologies and for discovery in global emergency or sustainability landscapes where new questions and scenarios are constantly emerging and new data is constantly being added to a large pool of servers [27]. Building database from a large pool of heterogeneous data, however, comprises a series of privacy requirements, formats, dimensions, biases and spatiotemporal resolution that constraint data integration and discovery [28, 29, 30, 31]. Fortunately, standard protocols to automate API discovery, semantic knowledge extraction, and ETFs algorithms are rapidly advancing [25, 32, 33], and overall different types of semantic technologies are rapidly emerging in the context of integrating many datasets into data knowledge graphs [34]. Yet, technologies focusing on novel evolutionary semantic knowledge extraction algorithms to build data knowledge graphs from many heterogeneous data-sources that can be rapidly integrated into interpretable technologies are not currently in place [27] (Fortuna: make clear how evolutionary biology-inspired semantic knowledge extraction algorithms might work. $\mathcal{RH}$v.1.0 explores evolutionary biology-inspired semantic algorithms around heterogeneous API data-discovery along three main deliverables (Deliverables D1.1, $\mathcal{APID}$, to D1.3, $\mathcal{DATAX}$, Figure 3 and Tables 3.1a-c). $\mathcal{RH}$v.1.0 delivers a data knowledge discovery graph from heterogeneous data-sources for the exploration of the Seas case study (Figure 2). $\mathcal{RH}$v.1.0 explores evolutionary biology-inspired rules using sematic algorithms to discover API and interactions that can be added to the exploration of the Seas database. The exploration of the Seas database started in 1965 and it actually has around 9 million entries, 1612 species, 20 countries and 11 sampling methods (Figure 2). $\mathcal{RH}$v.1.0 will expand it with Fishery data using the (global fishing watch), species interactions and social and stakeholders groups with different interests within each of the countries involved in the international exploration of the Seas.

3.1.2 WP2: $\mathcal{ROBHOOT}$ v.2.0: Causal Knowledge Discovery

Particularly relevant in Earth, Ecosystem and Sustainability science, the rapid progress of AI as an automated and explainable technology ([10, 17, 27, 35, 36, 37],+++) will increase our ability to make stronger inferences about future sustainability challenges and solutions [38]. $\mathcal{RH}$v.2.0 explores evolutionary biology-AI-inspired solutions to infer the causal interactions underlying heterogeneous and complex multidimensional datasets (Deliverables D2.1, $\mathcal{EAIA}$ to D2.3, $\mathcal{DIX}$, Tables 3.1a-c). Baity, Vicente, Melían: $\mathcal{RH}$v.2.0 introduces evolutionary biology-AI-inspired algorithms for the exploration of the Seas case study to ask questions around the drivers and their connections to the sustainability of the Oceans (Figure 2). For this, we need to identify the actors, e.g., species, threats like fishing activity, environental and habitat data, and then infer causal connections among all of them. As an illustration, the use of different Gears across countries affects the spatial and temporal samplings and the distribution of fish species. The consequence of the distribution of abundance/catches of Megrim and Haddock is strongly affected by the choice of gears by the Irish and the Spanish fleet (Figure 2). A group represents a set of actors with multiple traits. In this context, groups can be represented with evolving environmental and technological traits. This can be formally described as a distribution-fishery cooperation-competition matrix, $\mathcal{C}^2$, as follows:

$$\mathcal{C}^2 = \begin{array}{c} \\ \mathrm{D}^i_{\mathcal{A}_g,\mathcal{B}_g}(c) \\ \mathrm{D}^i_{\mathcal{A}_g,\mathcal{B}_g}(nc) \end{array} \begin{array}{c} \begin{array}{cc} \mathrm{F}^i_{\mathcal{A}_g,\mathcal{B}_g}(c) & \mathrm{F}^i_{\mathcal{A}_g,\mathcal{B}_g}(nc) \end{array} \\ \begin{pmatrix} \mathrm{c}(\varphi) & \mathrm{c}(\Phi), nc(\gamma_{A_g}, \gamma_{B_g}) \\ \mathrm{nc}(\Phi_{A_g}, \Phi_{B_g}), c(\gamma) & \mathrm{nc}(\phi_{A_g}, \phi_{B_g}) \end{pmatrix} \end{array},$$

where $\mathcal{D}$, $\mathcal{F}$, i, $\mathcal{A}_g$, $\mathcal{B}_g$, c and nc, represent distribution map and fishery of species i, group g within country $\mathcal{A}$ and $\mathcal{B}$, cooperation and non-cooperation, respectively. We will explore evolutionary biology-inspired functions representing environmental and technological traits with different degrees of complexity in the $\mathcal{C}^2$ matrix: If the two groups within the countries cooperate, $c(\varphi)$, then the environmental and technological rate change, φ, is syncrhonized between groups to evolve towards decreasing Gear bias and make distribution maps and the fishery sustainable. On the other side, if the two groups decide not to cooperate, $nc(\phi_{A_g}, \phi_{B_g})$, then there is environmental and technological rate change, ϕ_{A_g} and ϕ_{B_g} with each group following changes of their own gears, the GOV for the Ireland group and the Baka Gear for Spain group, independently of the other and as a function of their fishery interest. There is no interest in decreasing bias in species distribution maps

Work package	Lead Beneficiary	
WP Title	$\mathcal{ROBHOOT}$ **v.1.0**	
Participant number	**1, 2**	
Short name of participant	$\mathcal{EBD}-\mathcal{CSIC}$ **(1),** $\mathcal{IFISC}-\mathcal{CSIC}$ **(2)**	
Person Month per participant	1 and 2:(12 months $\times$ 0.25 (25% allocated time) == 3 (Provisional)	
Start month	**4**	
End month	**28**	
Objectives	Evolutionary biology-inspired semantic and multilayer networks algorithms for data knowledge discovery	
Description	Patterns from diverse data-sources for novel knowledge discovery	
Deliverables	**D1.1 (**$\mathcal{APID}$, Delivery month: 27, Lead: 1): Interaction data discovery from evolutionary biology-inspired semantic algorithms **D1.2 (**$\mathcal{DATAK}$, Delivery month: 27, Lead: 2): Data knowledge discovery mixing evolutionary semantic and multilayer network algorithms **D1.3 (**$\mathcal{DATAX}$, delivery month: 28, Lead: 1 and 2): Data knowledge discovery for the exploration of the Seas federated network	
WP Title	$\mathcal{ROBHOOT}$ **v.2.0**	
Participant number	**3, 4, and 5**	
Short name of participant	$\mathcal{FISHEC}-\mathcal{EAWAG}$ **(3),** $\mathcal{SIAM}-\mathcal{EAWAG}$ **(4),** $\mathcal{TARTU}$ **(5)**	
Person Month per participant	3, 4, and 5:(12 months $\times$ 0.25 (25% allocated time) == 3 (Provisional)	
Start month	**6**	
End month	**30**	
Objectives	Evolutionary biology-inspired AI-neural networks algorithms for causal knowledge discovery	
Description	Interpretable knowledge extraction for causal knowledge discovery	
Deliverables	**D2.1 (**$\mathcal{EAIA}$, Delivery month: 29, Lead: 3 and 4): Evolutionary biology-inspired AI algorithms **D2.2 (**$\mathcal{DIK}$, Delivery month: 29, Lead: 4 and 5): Evolutionary biology-inspired algorithms for interpretable knowledge discovery **D2.3 (**$\mathcal{DIX}$, Delivery month: 30, Lead: 3, 4 and 5): Causal knowledge discovery for the exploration of the Seas case study	
Title	$\mathcal{ROBHOOT}$ **v.3.0**	
Participant number	**6, 7, 8, 9 and 10**	
Participants	$\mathcal{IT}-\mathcal{EAWAG}$ **(6),** $\mathcal{UNIGRAZ}$ **(7),** $\mathcal{ICREA}$ **(8),** $\mathcal{SDSC}$ **(9),** $\mathcal{DESANTANA}$ **(10)**	
Person Month per participant	6, 7, 8, 9 and 10:(12 months $\times$ 0.25 (25% allocated time) == 3 (Provisional)	
Start month	**12**	
End month	**36**	
Objectives	Evolutionary neural-inspired knowledge discovery in federated networks	
Description	Automated discovery knowledge graphs in federated networks	
Deliverables	**D3.1 (**$\mathcal{SDF}$, Delivery month: 34, Lead: 6): Sharing knowledge discovery in federated networks **D3.2 (**$\mathcal{DIF}$, Delivery month: 34, Lead: 7): Neural biology-inspired discovery in federated networks **D3.3 (**$\mathcal{DIFX}$, Delivery month: 34, Lead: 6 ad 7): Neural biology-inspired discovery for the exploration of the Seas federated network **D3.4 (**$\mathcal{BSM}$, Delivery month; 36, Lead: 8): Automated Bayesian scientist for knowledge discovery in federated networks **D3.5 (**$\mathcal{DATAR}$, Delivery month: 36, Lead: 9): Reproducible knowledge discovery in federated networks **D3.6 (**$\mathcal{VISUR}$, Delivery month: 36, Lead: 10): Visualizing and reporting of knowledge discovery in federated networks	

Table 3.1b $\mathcal{ROBHOOT}$ **work package description**: *Work package, Title, Participants, Person Months per participant, Start and End month, Objectives, Description and deliverables of each Work Package.*

making fishery non sustainable in this case. In the last two scenarios groups enter in cooperation for the distribution map of species i, but not in the fishery $(c(\Phi), nc(\gamma_{A_g}, \gamma_{B_g}))$, or they do cooperate in the Fishery for species i but not for the distribution map of species i $(nc(\Phi_{A_g}, \Phi_{B_g}), c(\gamma))$. The situation for cooperation in the distribution maps follows agreements between the two groups to technological changes in the Gear but still preserving their GOV and the Baka Gears for fisheries. $\mathcal{RH}$v.2.0 search causal knowledge discovery for the exploration of the Seas containing 9 million entries, 1612 species (around 50 variables and traits per species), around 20 countries and 11 sampling methods (Figure 2). The search contrasts scenarios 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 D3.4, $\mathcal{BSM}$)[15, 17]. Causal knowledge graphs connect automated and explainable AI throughout prediction and knowledge power (Figure 2). Baity, Vicente, Melían: Make more clear how the evolutionary-biology-AI inspired modeling will be implemented: The explanation above is a first glimpse to be improved

$\mathcal{ROBHOOT}$ v.X.X	Deliver. number	Deliver. name	WP	Name Lead	Type	Disem.	Date
v.1.0	**D1.1**	$\mathcal{APID}$	WP1	$\mathcal{EBD-CSIC}$	OT	PU	27
v.1.0	**D1.2**	$\mathcal{DATAK}$	WP1	$\mathcal{IFISC-CSIC}$	OT	PU	27
v.1.0	**D1.3**	$\mathcal{DATAX}$	WP1	$\mathcal{EBD-IFISC-CSIC}$	R,DEC	PU	28
v.2.0	**D2.1**	$\mathcal{EAIA}$	WP2	$\mathcal{FISHEC-EAWAG}$	OT	PU	29
v.2.0	**D2.2**	$\mathcal{DIK}$	WP2	$\mathcal{SIAM-EAWAG}$, $\mathcal{TARTU}$	OT	PU	29
v.2.0	**D2.3**	$\mathcal{DIX}$	WP2	$\mathcal{FISHEC-SIAM-EAWAG}$, $\mathcal{TARTU}$	R,DEC	PU	30
v.3.0	**D3.1**	$\mathcal{SFN}$	WP3	$\mathcal{IT-EAWAG}$	OT	PU	34
v.3.0	**D3.2**	$\mathcal{DIF}$	WP3	$\mathcal{UNIGRAZ}$	OT	PU	34
v.3.0	**D3.3**	$\mathcal{DIFEX}$	WP3	$\mathcal{IT-EAWAG}, \mathcal{UNIGRAZ}$	R,DEC	PU	34
v.3.0	**D3.4**	$\mathcal{BSM}$	WP3	$\mathcal{ICREA}$	OT	PU	36
v.3.0	**D3.5**	$\mathcal{DATAR}$	WP3	$\mathcal{SDSC}$	OT	PU	36
v.3.0	**D3.6**	$\mathcal{VISUR}$	WP3	$\mathcal{DESANTANA}$	OT	PU	36

Table 3.1c List of Deliverables: *$\mathcal{ROBHOOT}$ contains three main work packages and twelve deliverables: $\mathcal{ROBHOOT}$* **v.1.0** *span from Month 4 to 28. Deliverable* **D1.3** **(**$\mathcal{DATAX}$**)** *generates the data knowledge discovery for the exploration of the Seas case study. Work package $\mathcal{ROBHOOT}$* **v.2.0** *span from Month 6 to 30. Deliverable* **D2.3** **(**$\mathcal{DIX}$**)** *generates the causal knowledge discovery for the exploration of the Seas case study, and $\mathcal{ROBHOOT}$* **v.3.0** *span from Month 12 to 36, bringing the deliverable* **D3.3** **(**$\mathcal{DIFEX}$**)**, *discovery in neural biology-inspired federated networks for the exploration of the Seas case study. Deliverables* **D3.4** *to* **D3.6**, *Bayesian space models, $\mathcal{BSM}$, reproducible knowledge discovery, $\mathcal{DATAR}$, and visualization and reporting, $\mathcal{DESANTANA}$, respectively, fully guarantee automation, reproducibility and visual reporting for the whole life cycle of $\mathcal{ROBHOOT}$. OT: Other (software/technical diagram, etc), R: Document report, DEC: website, press and media activity, videos and PU: Public fully open.*

3.1.3 WP3: $\mathcal{ROBHOOT}$ v.3.0: Discovery in Federated Networks

Integrating data and causal knowledge graphs provide a mechanistic understanding of how much cooperation vs. competition is occurring 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 in the context of heterogeneity, connectivity and firing probabilities [39, 40].

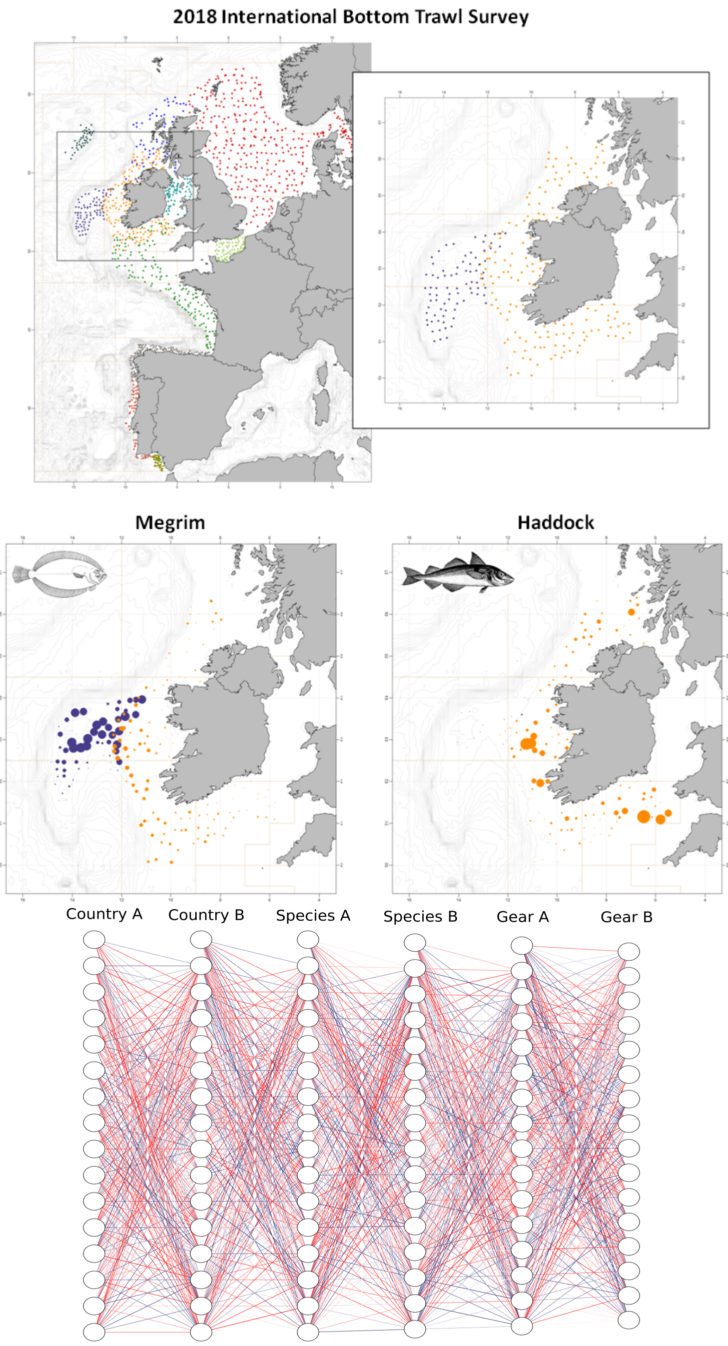

Figure 2: Causal Knowledge Discovery for Sustainable Ecosystems. **Top**) *The Irish Ground Fish Survey (IE-IGFS, Orange) and the Spanish Survey on the Porcupine Bank (SP-PORC, Blue) were part of the 2018 International Bottom Trawl Survey, coordinated by the International Council for the Exploration of the Sea [13]. 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.* **Bottom** *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 especies (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.*

Technologies in digital ecosystems around federated networks are scarce and mostly focus on decentralization, scalability and security fronts [12, 41, 42, 43, 44, 45]. In the science ecosystem, only a few applications of open decentralized technologies exist [7]. 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 ([10, 17]). 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 ([10],+++). $\mathcal{RH}$v.3.0 deploys sharing discovery knowledge graphs, D3.1, $\mathcal{SFN}$, into biology-inspired federated networks accounting for heterogeneous agents to discover novel biology-inspired solutions for the exploration of the Seas federated network (Deliverables D.3.1, $\mathcal{SFD}$ to D.3.3, $\mathcal{FIDEX}$, Tables 3.1a-c). Evolutionary algorithms might trigger novel algorithmic findings, the discovery knowledge graphs, and von Haldow, Maass:: $\mathcal{RH}$v.3.0 introduces "Cooperative Forecasting" as evolutionary biology-inspired neural learning algorithms for the discovery of new solutions in large federated networks (Tables 3.1a-c). $\mathcal{RH}$v.3.0 search for how learning among interacting heterogeneous groups discover evolutionary algorithms and in our exploration of the Seas case study this can be represented as follows: Now the focus is on cooperative learning to discover new solutions. For example, how learning from the most distant strategies in the technological and environmental traits can make distribution catchability maps similar. Groups can now be represented not only as environmental and technological traits, but with evolving learning traits as a function of the distance between each pair of groups sharing resources. This can be formally described as a distribution-fishery cooperation learning matrix $\mathcal{C}^2$, as follows:

$$\mathcal{C}^2 = \begin{array}{c} \\ \mathrm{D}^i_{\mathcal{A}_g,\mathcal{B}_g}(c) \\ \mathrm{D}^i_{\mathcal{A}_g,\mathcal{B}_g}(nc) \end{array} \begin{array}{c} \begin{array}{cc} \mathrm{F}^i_{\mathcal{A}_g,\mathcal{B}_g}(c) & \mathrm{F}^i_{\mathcal{A}_g,\mathcal{B}_g}(nc) \end{array} \\ \begin{pmatrix} \mathrm{c}(\varphi,\mathcal{L}_d) & \mathrm{c}(\Phi,\mathcal{L}_d), nc(\gamma_{A_g},\gamma_{B_g}) \\ \mathrm{nc}(\Phi_{A_g},\Phi_{B_g}), c(\gamma,\mathcal{L}_d) & \mathrm{nc}(\phi_{A_g},\phi_{B_g}) \end{pmatrix} \end{array},$$

where $\mathcal{D}$, $\mathcal{F}$, i, $\mathcal{A}_g$, $\mathcal{B}_g$, c and nc, represent distribution map and fishery of species i, group g within country $\mathcal{A}$ and $\mathcal{B}$, cooperation, and non-cooperation, respectively, as in $\mathcal{ROBHOOT}$ v.2.0. In addition, we introduce learning functions depending of the distance between two groups, $\mathcal{L}_d$. We will search evolving learning functions that can be coupled to environmental and technological traits with different degrees of complexity in the $\mathcal{C}^2$ matrix: If the two groups within the countries are sufficiently distant, then learning functions play a role to cooperate, $c(\varphi,\mathcal{L}_d)$, and the environmental and technological rate change, φ, strongly depend on learning between the interacting groups making distribution maps and the fishery more sustainable.
(Write from here how the learning scenario can enter in the non-cooperative strategies) On the other side, if the two groups decide not to cooperate, $nc(\phi_{A_g},\phi_{B_g})$, then there is environmental and technological rate change, ϕ_{A_g} and ϕ_{B_g} with each group following changes of their own gears, the GOV for the Ireland group and the Baka Gear for Spain group, independently of the other and as a function of their Fishery interest. There is no interest in decreasing bias in species distribution maps making fishery non sustainable in this case. In the last two scenarios groups enter in cooperation for the distribution map of species i, but not in the Fishery $(c(\Phi), nc(\gamma_{A_g},\gamma_{B_g}))$, or they do cooperate in the Fishery for species i but not for the distribution map of species i $(nc(\Phi_{A_g},\Phi_{B_g}), c(\gamma))$. The situation for cooperation in the distribution maps follows agreements between the two groups to technological changes in the Gear but still preserving their GOV and the Baka Gears for Fisheries. $\mathcal{RH}$v.2.0 search discovery knowledge graphs for the exploration of the Seas (Figure 2) containing 9 million entries, 1612 species, 15 countries and 11 sampling methods contrasting predictions from evolutionary biology-inspired algorithms and complements it with automated Bayesian machines ensuring the process is fully automated to facilitate search, and evaluation of models (Deliverable D3.4, $\mathcal{BSM}$) [15, 17], and communication of the full approach in public repositories (Section Impact).

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 [12], 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. $\mathcal{RH}$v.3.0 connects discovery knowledge graphs 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).

3.2 Management structure, milestones and procedures

- Describe the organisational structure and the decision-making (including a list of milestones (table 3.2a))
- Explain why the organisational structure and decision-making mechanisms are appropriate to the complexity and scale of the project.
- Describe any critical risks, relating to project implementation, that the stated project's objectives may not be achieved. Detail any risk mitigation measures. Please provide a table with critical risks identified and mitigating actions (table 3.2b) and relate these to the milestones.

Advisory board covering the weakest parts of the proposal – mention here

Milestone number	Milestone name	Related work package(s)	Due data (months)	Verification
M1	Data knowledge discovery	WP1	28	OS-Software,Paper/Conf.,Main-website
M2	Causal knowledge discovery	WP2	30	OS-Software,Paper/Conf.,Main-website
M3	Discovery in federated networks	WP3	36	OS-Software,Paper/Conf.,Main-website

Table 3.2a: List of Milestones: *$\mathcal{ROBHOOT}$* **v.1.0** *to* **v.3.0** *span from month 4 to 28, 6 to 30 and 12 to 36, respectively, to generate the "Data knowledge discovery", the "Causal knowledge discovery" and the "Discovery in federated networks" for the exploration of the seas case study. Verification for each milestone combines Open-Source software, papers and/or conference and a main website featuring reproducible, automated, visual and interpretable discovery.*

3.3 Consortium as a whole

$\mathcal{ROBHOOT}$ is a science-enabled multi-feature technology designed with a highly modular structure. Modularity allows to gain module functionality while maintaining cross-functional features among the different parts to produce a science-enabled interdisciplinary technology (Figure 3, WP one to three and milestones one to three, red, green and blue, respectively): **Data knowledge discovery**'s team requires skills in evolutionary biology, evolutionary computation, computer science and the physics of complex systems (Section 3.1.1, Table 3.2a and Figure 3). $\mathcal{ROBHOOT}$ **v.1.0** work package mixes complementary expertise in semantic algorithms, evolutionary computation algorithms and multilayer network metrics to create novel evolutionary-biology inspired ontology anotations when merging heterogeneous data-sources into one data knowledge discovery.

Description of risk	Work package	Proposed risk mitigation measures
Medium	**WP1**	Ontology-semantic algorithms as an alternative to evolutionary semantic algorithms to infer interactions among distinct data to build the data knowledge discovery as a general case and for the exploration of the Seas
Medium	**WP1**	Alternative to full automation for data ontologies, which are the algorithms?
Low	**WP1**	Implementation of static multilayer metrics to characterize data knowledge discovery graphs
Medium	**WP2**	Alternatives to evolutionary-biology-AI-inspired algorithms: modular approach
Low	**WP2**	Alternative to full automation for causal discovery, constraints to evolving functions
Low	**WP2**	Alternatives to causal discovery automation for the exploration of the Seas
Low	**WP3**	Alternatives to evolutionary neural biology-inspired algorithms in federated networks
Low	**WP3**	Automation of cooperative discovery and forecasting
Medium	**WP1-WP3**	Modular alternatives to "full reproducibility" connecting the three work packages
Medium	**WP1-WP3**	Modular alternatives to "full automation" connecting the three work packages

Table 3.2b: Critical risks for implementation: *$\mathcal{ROBHOOT}$ contains risks along its three main WPs and milestones. $\mathcal{RH}$v.1.0 with milestone "Data knowledge discovery" has three main risks related to evolutionary semantic algorithms, data discovery automation, and the multilayer metrics to characterize data knowledge graphs. $\mathcal{RH}$v.2.0 with milestone "Causal knowledge discovery" has three main risks related to evolutionary-biology-AI inspired algorithms, causal discovery automation, and discovery automation for the exploration of the Seas case study. $\mathcal{RH}$v.3.0 with milestone "Knowledge discovery in federated networks" has X main risks related to evolutionary neural biology-inspired algorithms, its automation, its interactions to x and y for full reproducibility and visualization and reporting.*

$\mathcal{EBD}-\mathcal{CSIC}$[1] team takes care of data knowledge graphs introducing novel evolutionary semantic algorithms to discover ontologies and interactions among many data-sources (Deliverable D1.1, $\mathcal{APID}$, Tables 3.1a-c). $\mathcal{IFISC}-\mathcal{CSIC}$[2] team focuses on multilayer network modularity, community detection and decentralization metrics for pattern detection in data knowledge discovery (Deliverable D1.2, $\mathcal{DATAK}$, Tables 3.1a-c). $\mathcal{EBD}-\mathcal{CSIC}$ and $\mathcal{EBD}-\mathcal{IFISC}$ teams will join efforts to merge evolutionary semantic algorithms and multilayer network metrics to produce the data knowledge discovery for the exploration of the Seas case study (Deliverable D1.3, $\mathcal{DATAX}$, Tables 3.1a-c and Figure 3 Milestone one in red).

[1]Miguel Fortuna
[2]Victor Eguíluz

ROBHOOT **v.2.0**'s team composed by *EAWAG − SIAM*[3], *TARTU*[4] and *EAWAG − FISHEC*[5] fussion Evolutionary biology, AI Algorithms and deep learning networks, the "Evolutionary biology-inspired AI algorithms" approach (Deliverable D2.1, *EAIA*, Table 3.1a-c and Figure 3, green). The overall goal of this milestone is to connect evolutionary biology mechanisms to deep learning networks to generate a **Causal knowledge discovery** technology to make patterns interpretable (Deliverable D2.2, *DIK*), Section 3.1.2, Table 3.2.a-c and Figures 2 and 3). The team for this milestone add inter-module complementarity expertise to *ROBHOOT* **v.1.0**'s team: Now the skills focus on data-scientists trained in deep learning networks and evolutionary biologists with expertise in evolutionary ecology theory and evolutionary-inspired networks (section 3.1.2 and Figure 3, green). Milestone two generates a causal knowledge discovery for the exploration of the Seas initially containing 9 million entries, 1612 species using around 11 sampling methods and more than 15 countries (Deliverable D2.3, *DIX*, Figures 2 and 3, green). Interdisciplinarity in *ROBHOOT* enters not only at the intra-module development stage, but also at the inter-module stage where causal knowledge discovery and evolutionary biology-inspired AI algorithms might form the basis for the emergence of interdisciplinarity breakthrough ideas reflected in the highly complementarity skills of the consortium. The first two modules in *ROBHOOT* contain researchers from Estonia, Spain, Switzerland.

The *ROBHOOT* consortium wants to advance the rapidly evolving digital ecosystem by making cooperative discovery a fundamental feature of it. For this purpose, a science-enabled data and causal knowledge discovery technology is not enough if they stay isolated from a discovery technology embedded in larger and scalable networks. To discover novel scenarios for ecosystem sustainability, **Discovery in federated networks** should learn to learn from heterogeneous data-sources in the context of evolutionary neural biology-inspired algorithms. To achieve scalability for the discovery in federated networks, neural-inspired protocols in federated networks is the excellency feature of *ROBHOOT* **v.3.0** (section 3.1.3). *ROBHOOT* **v.3.0**'s team composed by *EAWAG − IT*[6] and *UNIGRAZ*[7], develop protocols for sharing data and causal knowledge discovery and neural biology-inspired federated networks, respectively. The team forming *ROBHOOT* **v.3.0** also requires contrasting skills: First, developers working in sharing and security protocols to guarantee scalable transfer of data and causal knowledge discovery. Second, social scientists, computer scientists, and neurobiologists in collaboration to developers aiming to explore the role of evolving neural biology-inspired solutions accounting for heterogeneous data-sources in federated networks. *ROBHOOT* **v.3.0** is a fundamental stepping-stone for developing "Cooperative Forecasting": it first guarantees data and causal knowledge discovery are shareable objects. Then these objects represent the basis for discovery of novel paths that increase sustainability goals produced in the different nodes of a network that interact and learn from each other to find better forecasting scenarios at a global scale. *ROBHOOT* **v.3.0**'s implements heterogeneous groups of (cooperating and competing) neurons in federated networks for making cooperative forecasting a standard global property of *ROBHOOT* (Deliverable D3.2, *DIF*, Tables 3.1a-c). Milestone three generates discovery in federated networks for the exploration of the Seas to provide populations of scenarios satisfying biodiversity and sustainability maintenance while guaranteeing commercial interest of many interacting groups and stakeholders within and among countries (Deliverable D3.3, *DIFEX*, Figure 3, blue). *ROBHOOT* v.3.0 contain researchers from Switzerland and Austria. *ROBHOOT* architecture aims to guarantee strong communication and impact along its whole life cycle and development. The team formed by the *SDSC* (Deliverable D3.4, *DATAR*[8]), *ICREA* (Deliverable D3.5, *BSM*[9]) and **De Santana** (Deliverable D3.6, *VISUR*[10]) will implement reproducibility, automation, and visualization and reporting, respectively, features crossing all *ROBHOOT* milestones to secure dissemination along its life cycle (Figure 3).

[3] Marco Baity
[4] Raul Vicente, to be confirmed
[5] Carlos Melián
[6] Harald von Waldow
[7] Wolfgang Maass
[8] Chistine Choirat, to be confirmed
[9] Roger Guimerà, to be confirmed
[10] partner as a company or institution to be confirmed

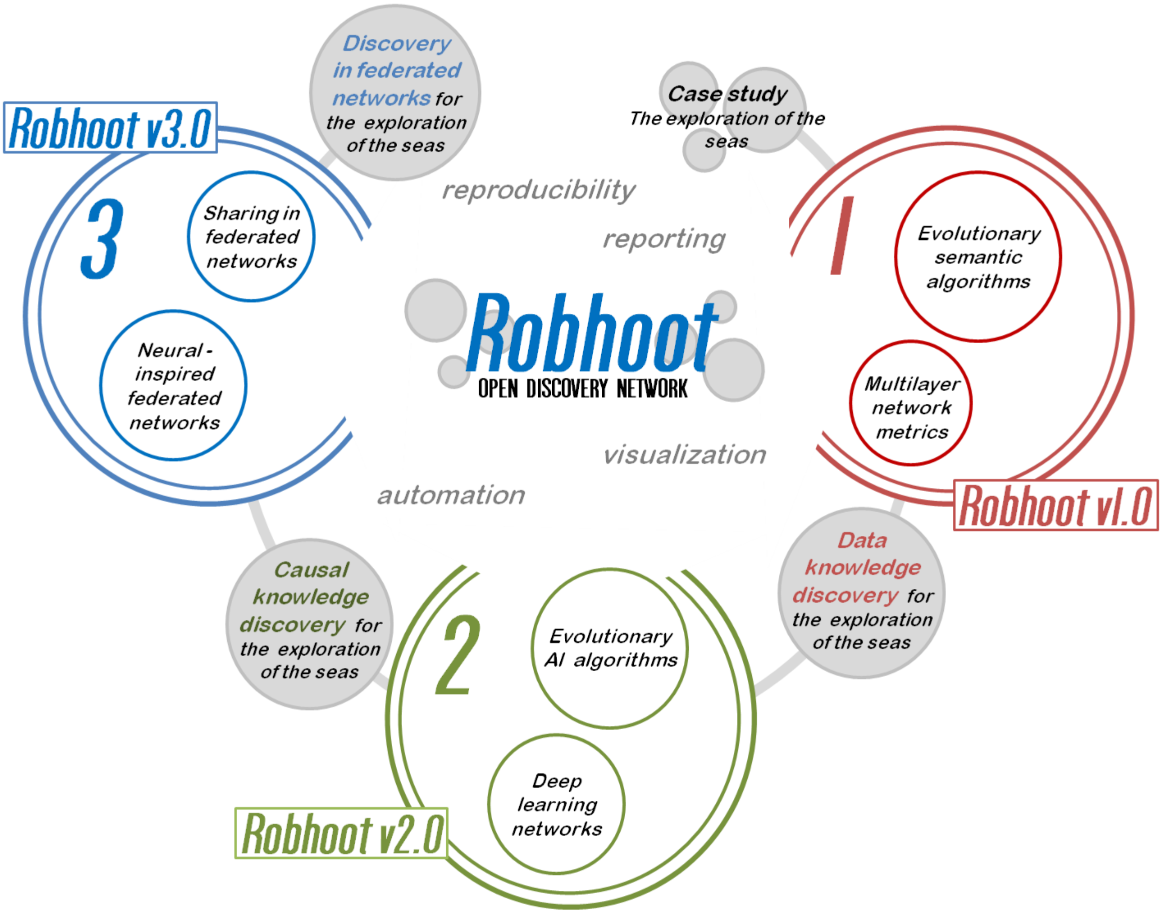

Figure 3: ROBHOOT delivers three work packages along three milestones: **Data knowledge discovery** *for the exploration of the seas in ROBHOOT* **v.1.0**. **Causal knowledge discovery** *for the exploration of the seas in ROBHOOT* **v.2.0**, *and* **Discovery in Federated Networks** *for the exploration of the seas integrating all features into one interdisciplinary science-enabled technology in ROBHOOT* **v.3.0**. *Reproducibility, visualization, automation and reporting cross all the milestones to guarantee transparency and impact.*

3.4 Resources to be committed

- Please make sure the information in this section matches the costs as stated in the budget table in section 3 of the administrative proposal forms, and the number of person months, shown in the detailed work package descriptions. Please provide the following:
- a table showing number of person months required (table 3.4a)
- a table showing 'other direct costs' (table 3.4b) for participants where those costs exceed 15% of the personnel costs (according to the budget table in section 3 of the administrative proposal forms)

4 Members of the consortium

4.1 Participants (applicants)

- For each participant, provide the following: a description of the legal entity and its main tasks, with an explanation of how its profile matches the tasks in the proposal
- a curriculum vitae or description of the profile of the persons, including their gender, who will be primarily responsible for carrying out the proposed research and/or innovation activities. Indicate each person who would be a first-time participant to FET under Horizon 2020
- a list of up to 5 relevant publications, and/or products, services (including widely-used datasets or software), or other achievements relevant to the call content
- List of up to 5 relevant previous projects or activities, connected to the subject of this proposal
- a description of any significant infrastructure and/or any major items of technical equipment, relevant to the proposed work
- if operational capacity cannot be demonstrated at the time of submitting the proposal, describe the concrete measures that will be taken to obtain it by the time of the implementation of the task

- (description legal identity) Dr. Carlos Melián is a tenured researcher in Theoretical Evolutionary Ecology at EAWAG, ETH-Domain in Switzerland, and associate professor at the University of Bern. (CV, gender, responsible research proposed, first time participant FET)
  He is the principal coordinator of the proposal. Dr. Melián has broad expertise in evolutionary algorithms and eco-evolutionary dynamics in ecological communities and biodiversity.
  (5 pubs) • Melián C, et al. 2018. Deciphering the interdependence between ecological and evolutionary networks. Trends in ecology evolution 33,7: 504-512. • Andreazzi C, Guimaraes P, Melián C. 2018. Eco-evolutionary feedbacks promote fluctuating selection and long-term stability of antagonistic networks. Proc. R. Soc. B 285: 20172596. • Melián C, Seehausen O, Eguiluz V, Fortuna M, Deiner K. 2015. Diversification and Biodiversity Dynamics of Hot and Cold Spots. Ecography 38, 393-401. • Melián C, et al. 2015. Dispersal dynamics in food webs. American Naturalist 185, 2: 157-168. • Melián C., et al. 2014. Individual trait variation and diversity in food webs. Advances in Ecological Research. Vol. 50. Academic Press, 207-241.
- **Victor M. Eguíluz (IFISC, CSIC, Spain)**: IFISC is an Maria de Maetzu Excellent center at the UIB, Balearic Islands. Dr. Eguíluz has expertise in health-related topics, in particular he has developed collaborations with Harvard medical school and many biodiversity and sustainability research institutions. The group of the PL has worked in the development of data-driven agent-based networks in social, biological and environmental problems with particular relevance in epidemiological networks.
-
-

4.2 Third parties involved in the project (including use of third party resources)

- For each participant, does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted) Y/N If yes, please describe and justify the tasks to be subcontracted
- Does the participant envisage that part of its work is performed by linked third parties2 Y/N If yes, please describe the third party, the link of the participant to the third party, and describe and justify the foreseen tasks to be performed by the third party
- Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement) Y/N If yes, please describe the third party and their contributions
- Does the participant envisage that part of the work is performed by International Partners3 (Article 14a of the General Model Grant Agreement)? Y/N If yes, please describe the International Partner(s) and their contributions.

5 Ethics and Security

This section is not covered by the page limit.

5.1 Ethics

For more guidance, see the document "How to complete your ethics self-assessment". If you have entered any ethics issues in the ethical issue table in the administrative proposal forms, you must:

- submit an ethics self-assessment, which:
- describes how the proposal meets the national legal and ethical requirements of the country or countries where the tasks raising ethical issues are to be carried out;
- explains in detail how you intend to address the issues in the ethical issues table, in particular as regards: research objectives (e.g. study of vulnerable populations, dual use, etc.) research methodology (e.g. clinical trials, involvement of children and related consent procedures, protection of any data collected, etc.)
- the potential impact of the research (e.g. dual use issues, environmental damage, stigmatisation of particular social groups, political or financial retaliation, benefit-sharing, misuse, etc.)
- If you plan to request these documents specifically for the project you are proposing, your request must contain an explicit reference to the project title.

5.2 Security

- activities or results raising security issues: (YES/NO)
- EU-classified information as background or results: (YES/NO)

References

[1] Christopher Wilson and Maria Gabrielsen Jumbert. The new informatics of pandemic response: humanitarian technology, efficiency, and the subtle retreat of national agency. *Journal of International Humanitarian Action*, 3(1), 2018. ISSN 2364-3412. doi: 10.1186/s41018-018-0036-5.

[2] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, jan 2015. ISSN 08936080. doi: 10.1016/j.neunet.2014.09.003. URL `https://linkinghub.elsevier.com/retrieve/pii/S0893608014002135`.

[3] Marion Maisonobe, Denis Eckert, Michel Grossetti, Laurent Jégou, and Béatrice Milard. The world network of scientific collaborations between cities: domestic or international dynamics? *Journal of Informetrics*, 10(4):1025–1036, 2016. doi: 10.1016/j.joi.2016.06.002. URL `https://ideas.repec.org/a/eee/infome/v10y2016i4p1025-1036.html`.

[4] H. Inhaber. Changes in centralization of science. *Research Policy*, 6(2):178–193, apr 1977. ISSN 0048-7333. doi: 10.1016/0048-7333(77)90024-5. URL `https://www.sciencedirect.com/science/article/abs/pii/0048733377900245`.

[5] John P a Ioannidis. Why most published research findings are false. *PLoS medicine*, 2(8):e124, aug 2005. ISSN 1549-1676. doi: 10.1371/journal.pmed.0020124. URL `http://www.ncbi.nlm.nih.gov/pubmed/16060722`.

[6] Ferric C Fang and Arturo Casadevall. Retracted Science and the Retraction Index. *Infection and Immunity*, 79(10):3855 LP – 3859, oct 2011. doi: 10.1128/IAI.05661-11. URL `http://iai.asm.org/content/79/10/3855.abstract`.

[7] Vlad Günther and Alexandru Chirita. " Scienceroot " Whitepaper. 2018. URL `https://www.scienceroot.com/`.

[8] Tom E. Hardwicke, Maya B. Mathur, Kyle MacDonald, Gustav Nilsonne, George C. Banks, Mallory C. Kidwell, Alicia Hofelich Mohr, Elizabeth Clayton, Erica J. Yoon, Michael Henry Tessler, Richie L. Lenne, Sara Altman, Bria Long, and Michael C. Frank. Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, 5(8):180448, sep 2018. ISSN 20545703. doi: 10.1098/rsos.180448. URL `https://doi.org/10.1098/rsos.180448`.

[9] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A Survey on Bias and Fairness in Machine Learning. 2019. URL `http://arxiv.org/abs/1908.09635`.

[10] Esteban Real, Chen Liang, David R. So, and Quoc V. Le. AutoML-Zero: Evolving Machine Learning Algorithms From Scratch. 2020. URL `http://arxiv.org/abs/2003.03384`.

[11] Matías E Mastrángelo, Natalia Pérez-Harguindeguy, Lucas Enrico, Elena Bennett, Sandra Lavorel, Graeme S Cumming, Dilini Abeygunawardane, Leonardo D Amarilla, Benjamin Burkhard, Benis N Egoh, Luke Frishkoff, Leonardo Galetto, Sibyl Huber, Daniel S Karp, Alison Ke, Esteban Kowaljow, Angela Kronenburg-García, Bruno Locatelli, Berta Martín-López, Patrick Meyfroidt, Tuyeni H Mwampamba, Jeanne Nel, Kimberly A Nicholas, Charles Nicholson, Elisa Oteros-Rozas, Sebataolo J Rahlao, Ciara Raudsepp-Hearne, Taylor Ricketts, Uttam B Shrestha, Carolina Torres, Klara J Winkler, and Kim Zoeller. Key knowledge gaps to achieve global sustainability goals. *Nature Sustainability*, 2(12):1115–1121, 2019. ISSN 2398-9629. doi: 10.1038/s41893-019-0412-1. URL `https://doi.org/10.1038/s41893-019-0412-1`.

[12] Johnny Dilley, Andrew Poelstra, Jonathan Wilkins, Marta Piekarska, Ben Gorlick, and Mark Friedenbach. Strong Federations: An Interoperable Blockchain Solution to Centralized Third-Party Risks. 2016. URL `http://arxiv.org/abs/1612.05491`.

[13] International bottom trawl survey working group (ibtswg). *ICES Scientific Reports*, 2019.

[14] Melnik S., Gubarev A., Long J. J., Romer G., Shivakumar S., Tolton M., and T Vassilakis. Dremel: Interactive Analysis of Web-Scale Datasets. *Proc. of the 36th International Conference on Very Large Data Bases (VLDB)*, pages 159–176, 2010.

[15] Christian Steinruecken, Emma Smith, David Janz, James Lloyd, and Zoubin Ghahramani. The Automatic Statistician. pages 161–173. 2019. ISBN 9783030053185. doi: 10.1007/978-3-030-05318-5_9.

[16] *Modulos*. URL `http://www.modulos.ai/`.

[17] Roger Guimerà, Ignasi Reichardt, Antoni Aguilar-Mogas, Francesco A. Massucci, Manuel Miranda, Jordi Pallarès, and Marta Sales-Pardo. A Bayesian machine scientist to aid in the solution of challenging scientific problems. *Science Advances*, 6(5):eaav6971, 2020. ISSN 23752548. doi: 10.1126/sciadv.aav6971.

[18] *Google AI*. URL `https://ai.google/`.

[19] *Iris*. URL `https://iris.ai`.

[20] *EaseML*. URL `https://github.com/DS3Lab/easeml`.

[21] *Datarobot*. URL `https://www.datarobot.com/`.

[22] *Aito*. URL `https://aito.ai/`.

[23] Cesar Soto-Valero, Amine Benelallam, Nicolas Harrand, Olivier Barais, and Benoit Baudry. The emergence of software diversity in maven central. *IEEE International Working Conference on Mining Software Repositories*, 2019-May:333–343, 2019. ISSN 21601860. doi: 10.1109/MSR.2019.00059.

[24]

[25] J. Fan, A. Kalyanpur, D. C. Gondek, and D. A. Ferrucci. Automatic knowledge extraction from documents. *IBM Journal of Research and Development*, 56(3.4):5:1–5:10, may 2012. ISSN 0018-8646. doi: 10.1147/JRD.2012.2186519. URL `http://ieeexplore.ieee.org/document/6177723/`.

[26] Peter W J Staar, Michele Dolfi, Christoph Auer, and Costas Bekas. Corpus Conversion Service. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 774–782, New York, NY, USA, jul 2018. ACM. ISBN 9781450355520. doi: 10.1145/3219819.3219834. URL `https://dl.acm.org/doi/10.1145/3219819.3219834`.

[27] Giuseppe Futia and Antonio Vetrò. On the integration of knowledge graphs into deep learning models for a more comprehensible AI-Three challenges for future research. *Information (Switzerland)*, 11(2), 2020. ISSN 20782489. doi: 10.3390/info11020122.

[28] *Openstreet*, . URL `https://www.openstreetmap.org/{#}map=5/51.509/-0.110`.

[29] *Blue-cloud*. URL `https://www.blue-cloud.org/`.

[30] *HOT*. URL `https://www.hotosm.org/`.

[31] *Elixir open data*. URL `https://elixir-europe.org/`.

[32] *APIs*. URL `https://apis.guru/openapi-directory/`.

[33] *Open Knowledge Foundation*, . URL `https://discuss.okfn.org/`.

[34] *Knowledge Graph COVID-19*. URL ttp://www.odbms.org/2020/03/we-build-a-knowledge-graph-on-covid-19/.

[35] Anthony O'Hare. Inference in High-Dimensional Parameter Space. *Journal of Computational Biology*, 22(11):997–1004, 2015. ISSN 1066-5277. doi: 10.1089/cmb.2015.0086. URL http://online.liebertpub.com/doi/10.1089/cmb.2015.0086.

[36] Yolanda Gil and Bart Selman. A 20-Year Community Roadmap for Artificial Intelligence Research in the US. aug 2019. URL https://bit.ly/2ZNVBVbhttp://arxiv.org/abs/1908.02624.

[37] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. pages 1–10, 2019. URL http://arxiv.org/abs/1911.01429.

[38] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and & Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature*. ISSN 0028-0836. doi: 10.1038/s41586-019-0912-1. URL www.nature.com/nature.

[39] Wolfgang Maass. Noise as a resource for computation and learning in networks of spiking neurons. *Proceedings of the IEEE*, 102(5):860–880, 2014. ISSN 00189219. doi: 10.1109/JPROC.2014.2310593.

[40] Wolfgang Maass. To Spike or Not to Spike: That Is the Question. *Proceedings of the IEEE*, 103(12):2219–2224, 2015. ISSN 00189219. doi: 10.1109/JPROC.2015.2496679.

[41] Golem. The Golem Project Crowdfunding Whitepaper. *Golem.Network*, (November):1–28, 2016. URL https://golem.network/crowdfunding/Golemwhitepaper.pdf.

[42] Nikolai Durov. Telegram Open Network. pages 1–132, 2017.

[43] Elli Androulaki, Artem Barger, Vita Bortnikov, Srinivasan Muralidharan, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Chet Murthy, Christopher Ferris, Gennady Laventman, Yacov Manevich, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukolić, Sharon Weed Cocco, and Jason Yellick. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. *Proceedings of the 13th EuroSys Conference, EuroSys 2018*, 2018-Janua, 2018. doi: 10.1145/3190508.3190538.

[44] Ocean Protocol Foundation, BigchainDB GmbH, and DEX Pte. Ltd. Ocean Protocol: A Decentralized Substrate for AI Data & Services Technical Whitepaper. pages 1–51, 2018. URL https://oceanprotocol.com/.

[45] BigchainDB GmbH. BigchainDB: The blockchain database. *BigchainDB. The blockchain database.*, (May):1–14, 2018. doi: 10.1111/j.1365-2958.2006.05434.x. URL https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf.