ROBHOOT
Discovery Knowledge Graphs in Evolutionary Biology-Inspired Federated Networks
v.2.0

May 11, 2020

Contents

Summary

Global sustainability is a major goal of humanity. Many studies have shown global sustainability could be achieved by strengthening transparency, feedbacks 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 global knowledge gaps to reach sustainability and biodiversity conservation goals are at a very incipient stage of development. We fussion data to causal knowledge graph, the discovery knowledge graph, in evolutionary biology-inspired federated networks for a sustainable- and knowledge-inspired society. Discovery knowledge graphs 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 graphs in evolving 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 is 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 the pace dictated by the availability of data and a myriad of data-driven approaches are being developed to extract patterns from data ([2]). Data analytics and the methods developed around it is also being challenged because the diversity of data is growing worldwide (+++). On the other side, AI approaches are rapidly evolving towards more explainable/interpretable pattern description (+++). This forces the digital ecosystem to account for the increasing data heterogeneity and the need of novel methods for finding rapidly interpretable patterns, but still, science-enabled technologies accounting for data heterogeneity and interpretable patterns are scarce [3].Still much bla bla – more flow towards the main point

Describe how our vision surpasses substantially any technological paradigms that currently exist or are under development The transformation of a digital society into a knowledge-inspired society requires solving several gaps: First, science-enabled technological paradigm assisting humans is biased towards a limited range of the "observable" heterogeneity in data-sources limiting the number of interpretable patterns (+++). Second, the AI technological paradigm is rooted in optimization functions (i.e., function loss or reward, similar to fitness optima functions in evolutionary biology +++). Optimization-based technologies limit a broader number of plausible solutions, as usually found in evolving learning capable biological systems (+++), and third, science-enabled technologies (for scientific inquiry) are highly fragmented, partly solve reproducibility and are mostly developed in close-source software ([4, 5, 6, 7, 8, 9, 10]). To leverage the abundance and heterogeneity of data, a science-enabled technology should be able to obtain information from a large pool of heterogeneous data-sources. Second, the analysis of the data should go beyond the identification of patterns and consider approaches delivering processes, the interpretable patterns, to the end-user. Third, the analysis should be performed in 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 finally, the whole process should be automated, reproducible and transparent such that benefit the public.

Describe the vision of a radically-new science-enabled technology that the project would contribute towardsThe project will In this regard, obtaining interpretable information from data collected from many distinct sources combined with science-enabled technologies accounting for evolving interactions among distinct heterogeneous groups are required to gain more robust scenarios in knowledge-inspired societies. Yet, discovery science-enabled technologies delivering interpretable patterns from heterogeneous data-sources applied to gain information of complex governance, social, environmental and technological problems are particularly lacking (Figure 1 and Table 1) [11]. Our project contributes towards a science-enabled technology accounting for data heterogeneity and interpretable patterns, the discovery knowledge graph (Figures 1 and 2 and Table 1), obtained from evolutionary biology-inspired in federated networks. Evolutionary dynamics-inspired technology extracting information from highly heterogeneous and multidimensional groups while minimizing the need of having optimal solutions makes possible the study of within and between groups consensus scenarios to enrich knowledge-inspired societies facing global challenging problems.

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) might be driving functional interactions with other species (i.e., cooperative, antagonistic, competitive, or mutualistic, infectious diseases... etc), but most approaches have neglected the effect of trait dimensionality in heterogeneous populations on interactions being these competition or cooperation **(On neural systems, the vast majority of neurons in the brain show highly differentitated morphological, genetic and phenotypic states? (refs, Wolfgang))**, yet the understanding of functional interactions among such a highly differentiatied states (groups, etc) is not well understood. Taken together, these results suggest that evolved information processing systems 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 and that new science-enabled technologies accounting for diversification, dimensionality and heterogeneity of highly distinct groups in functional information processing federated networks, are required to develop further the increasing demand of data-source heterogeneity and interpretable-knowledge patterns.

Biodiversity data collected by many different countries is a good example for understanding open-problems in federated networks. Many international programs for exploration of the seas (i.e., ICES, ...) involve many countries collecting biodiversity data using, despite efforts of standarization, different protocols and technologies. The data is then used to understand the spatiotemporal dynamics of the ecological communities as a baseline to inform fisheries (+++). Each country might collect data with different gear systems (Figure 2). This is because each country has commercial interests in specific species, despite attempts to dictate standards for the gear systems among all the countries. The result is that countries use different gear systems and collect heterogeneous and biased data about the same species. This makes difficult to obtain accurate distribution maps of species (Figure 2). This situation can be outlined as follows: country own's interest in specific gear systems vs. shared interest using standarized 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., many gear systems and other technologies), and groups (i.e., within and between countries) to mitigate risks and enhance global cooperation scenarios in such a multidimensional ecosystems are not in place ([1], +++).

Describe the overall and specific objectives for the project, which should be clear, measurable, realistic and achievable within the duration of the project. (The details of the project plan belong to the Implementation section) The example about the exploration of the seas teaches us the need of science-enabled technologies facilitating interpretable patterns accounting for heterogeneous data and groups to overcome fragmented and partial responses to a global biodiversity problem. The goal of *ROBHOOT* is to propose a science-enabled technology concept integrating data and causal knowledge graphs into evolutionary biology-inspired federated networks to lay the foundation for a novel scientific discovery technology. *ROBHOOT* contributes towards reproducible cooperative forecasting scenarios in rapidly changing global sustainability landscapes (Figures 1, 2, and 3): *ROBHOOT* **v.1.0** deploys data knowledge graphs obtained from heterogeneous data-sources to decipher global data-architecture maps. *ROBHOOT* **v.2.0** develops explainable evolutionary biology-inspired approaches to fussion data and causal knowledge graphs into discovery knowledge graphs, and *ROBHOOT* **v.3.0** expands discovery knowledge graphs along evolutionary biology-inspired federated networks.

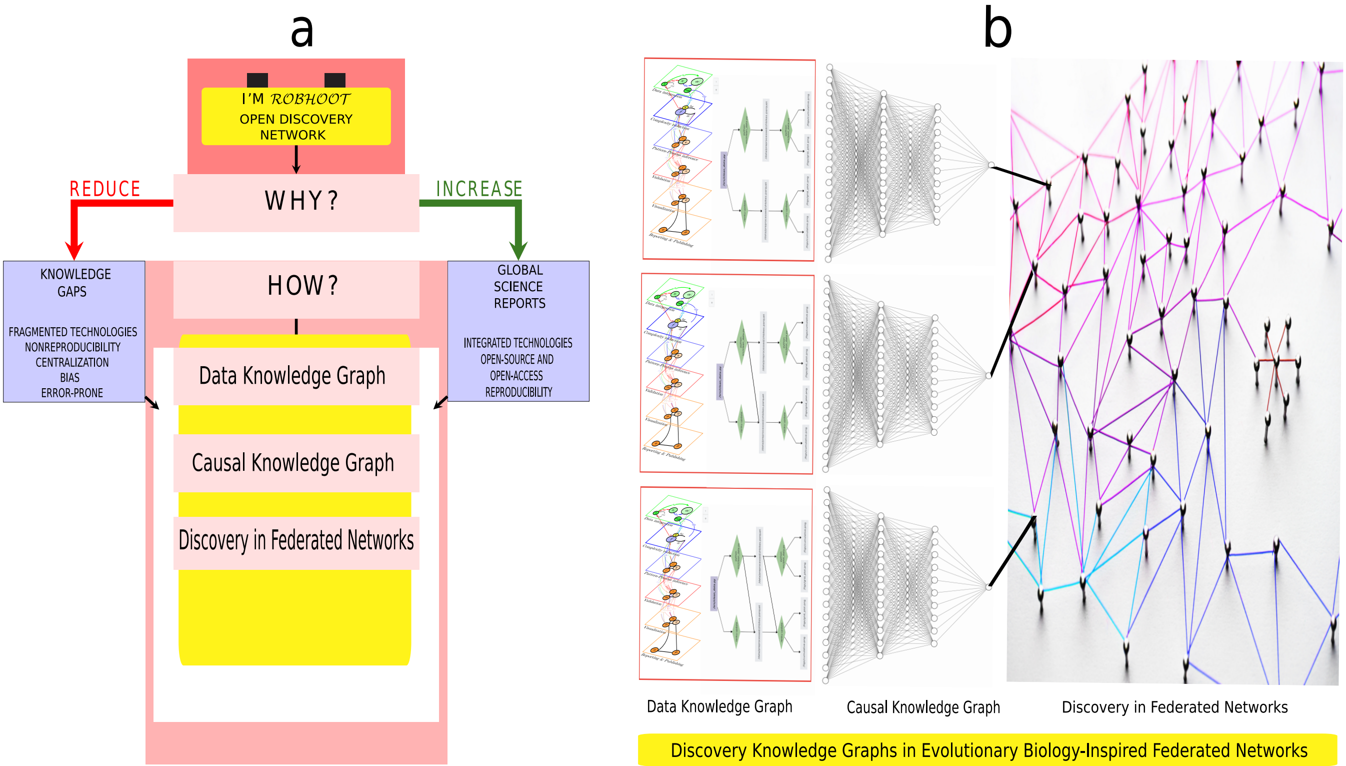

Figure 1: Discovery in Evolutionary Biology-Inspired Federated Networks. *ROBHOOT fussion data and causal knowledge graphs, the "Discovery Knowledge Graphs", into biology-inspired federated networks for a sustainable knowledge-inspired society:* **a)** *ROBHOOT targets global knowledge gaps (red path) and open-access reproducible discovery reports (green path). It integrates three science-enabled technologies:* **a,b)** *Data Knowledge Graphs for heterogeneous source-data discovery.* **a,b)** *Causal Knowledge Graphs to fussion data knowledge to interpretable patterns using "Evolutionary AI automation" and biology-inspired methods, and* **a,b)** *Discovery in biology-inspired federated networks for "Cooperative Forecasting".* **Discovery knowledge graphs in biology-inspired federated networks** *integrate data and causal knowledge graphs into federated networks to generate cooperative forecasting to respond to global and sustainability challenges.*

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, technologies integrating data-driven causal inference into intelligent networks providing rapid and global interpretable information when solving complex governance, social, environmental and technological problems are lacking. Depite rapid advances of research platforms for data analytics in the last decade [13, 14, 15, 16, 17, 18, 19, 20, 21], the integration of science-enabled technologies currently lack knowledge-inspired approaches accounting for data heterogeneoity and interpretable pattern discovery impacting knowledge-inspired societies to help responding to rapidly evolving global sustainability challenges (Figures 1 and 2, and Table 1). Technologies facilitating interpretable patterns from complex systems accounting for large heterogeneity and multidimensionality present still many challenges (+++). This is

Word	Meaning
Data knowledge graph	Technology-driven information extraction from heterogeneous data-sources
Causal knowledge graph	Technology-driven explainable/interpretable patterns to take better informed decisions on global and complex sustainability challenges
Discovery knowledge graph	Fussion of data and causal knowledge graphs for better pattern understanding of heterogeneous data
Automation	Algorithms targeting minimal human-driven interference
Knowledge inspired society	Open-access discovery to facilitate informed decisions in global sustainability challenges
Neutral-knowledge generation	Reproducible reports making transparent the many sources of bias in the discovery process

Table 1: Glossary of terms.

particularly relevant in global emergency or sustainability landscapes, where data heterogeneity properties like availability, data fragmentation, and transparency drive constantly emerging feedbacks among the many data-sources to predict new situations more accurately (Figure 2).

Our understanding of evolved information processing systems driven by multidimensional factors and highly heterogeneous groups is currently 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 biology-inspired explainable knowledge graphs into federated networks accounting for heterogeneity and multidimensionality to make discovery a rapidly 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 discovery in federated networks +++. *ROBHOOT* goes beyond the state-of-the-art of knowledge-graphs by fussioning data and causal knowledge graphs, the discovery knowledge graphs, and scalating these into evolving biology-inspired federated networks to move knowledge-inspired societies towards reaching global sustainability goals when large number heterogeneous groups share resources driven by multiple factors.

ROBHOOT **v.1.0** deploys a data discovery technology to generate data knowledge graphs for an understanding of interpretable patterns accounting for heterogeneous data-sources. Data-architecture alone is not sufficent to outline predictive scenarios in complex sustainability problems. 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 global data-architecture into rapid automated causal knowledge graphs, the discovery knowledge graphs, to facilitate discovery. *ROBHOOT* **v.2.0** integrates automated and explainable evolutionary biology-inspired methods to decipher causal knowledge graphs 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 discovery knowledge graphs to federated networks by connecting heterogeneous-neural inspired networks to learning to obtain cooperative forecasting from heterogeneous collections of data-sources (section 3.3).

1.3 Interdisciplinarity and non-incrementality of the research proposed

Respond more directly! — 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:: Still in a descriptive phase here:: Now it is a disconnected set of ideas *ROBHOOT* is a science-enabled multi-feature technology for interpretable data-driven discovery in federated networks (Figures 1 to 3 and Tables 1 to 3). It contains three milestones 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 algorithms for API data discovery. This module is complemented with scientists from complex networks taking care of quantitative methods in the data knowledge graphs to decipher the existing gaps in data discovery and data-architecture technologies (section 3.1). *ROBHOOT* **v.2.0** team is compossed by data-scientists trained in deep learning networks and automation algorithms, theoreticians and 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 graphs from heterogeneous data-sources (section 3.2). *ROBHOOT* **v.3.0** combines computer scientists and developers targeting sharing and evolutionary biology-inspired models of federated networks, with social scientist, and scientists specialized in ecology and evolutionary biology (section 3.3). The complementarity of the teams in modules one to three strengthen the collaboration for making *ROBHOOT* a science-enable functional technology in a rapidly evolving digital ecosystem [22].

ROBHOOT aims to bring global transparency in knowledge 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). 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 data and causal knowledge graphs into federated networks still require taking risky steps: combining heterogeneous data-sources and evolutionary biology-inspired neural modeling approaches to fill out the existing gaps in the explainable methods arena ato bring them to causal inference of learning with heterogeneous agents sharing resources in complex ecosystems.

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).

2 Impact

2.1 Expected impact

Please be specific, and provide only information that applies to the proposal and its objectives. Wherever possible, use quantified indicators and targets.

Describe how the project will contribute to the expected impacts set out in the work programme under the relevant topic: Scientific and technological contributions to the foundation of a new future technology:: Describe the importance of the technological outcome with regards to its transformational impact on science, technology and/or society:: Building leading research and innovation capacity across Europe by involvement of key actors that can make a difference in the future, for example excellent young researchers, ambitious high-tech SMEs or first-time participants to FET under Horizon 2020:: FET Open combines high scientific ambition with concrete technological implications. It aims to attract interdisciplinary consortia that do not shy away from exploring connections between remote disciplines in order to open-up new and potentially game changing technological directions that FET as a whole aims to develop into the leading technology paradigms of the future, including through FET-Proactive projects and FET-Flagship initiatives. In spite of the high initial risk, the long-term impact can be enormous: these new technologies can become the core for new high-growth companies, for new industries or for radically new ways of tackling societal challenges. We are moving rapidly towards knowledge-inspired societies in need of radically tackling new societal and global environmental challenges. In such a global ecosystem, access to cooperative forecasting and interpretable information is key to generate rapid and robust scenarios when facing complex problems including global sustainability challenges (i.e., global health, ecosystems degradation, biodiversity loss, etc). *ROBHOOT* contributes to **Evolutionary AI Automation**, **Cooperative Forecasting** and **Interpretable Information** for a new science-enabled technology targeting knowledge-inspired societies: First, evolutionary AI automation decipher open-ended search interpretation of complex systems. 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 information accounts for global data-arquitecture and causal knowledge graphs, the discovery knowledge graphs, allowing individuals and companies to access market information to address complex scenarios of future strategies in highly fluctuating local and global conditions.

Describe the empowerment of new and high-potential actors towards future technological leadership. Decision making and governance at local, regional and global scales require access to transparent and reproducible data and interpretable patterns to analyze local solutions benefiting society in real-time emergency situations but also in medium and long term Ecosystem sustainability challenges. Make clear here the Choirat and Guimera group about reproducibility and automation in each of the Milestones, respectively:: any substantial impacts not mentioned in the work programme, that would enhance innovation capacity; create new market opportunities, strengthen competitiveness and growth of companies, address issues related to climate change or the environment, or bring other important benefits for society. Global automated, transparent and reproducible, cooperative, and explainable discovery can have a large impact to knowledge-inspired societies in need to access rapid, robust, and reproducible reports to take informed decisions. It also creates new market opportunities for companies. First, global-data architecture help to build a vision about ..., Second,...., and third.... — Explicitly mention pros about Legal and financial transparency – Reproducibility in Social Governance – Impact to emerging and sustainability challenges :: Novel service for NGO, society and thinktank transparent and reproducible public policies: Advisory boards :: Sustanaibility – check the SDG — This consortium brings together excellent partners from the fields of X, Y, Z and technology development, including one SME, who all exhibit a long-standing experience interdisciplinary research across the boundaries of the individual disciplines. The subsection on related projects shows that this is a novel constellation in Europe (and possibly worldwide). Thus, this consortium is at the leading edge...

2.2 Measures to maximize impact

This section still collection of what to follow ++ random ideas

2.2.1 Dissemination and exploitation of results

- Provide a plan for disseminating and exploiting the project results. The plan, which should be proportionate to the scale of the project, should contain measures to be implemented both during and after the project.
- Explain how the proposed measures will help to achieve the expected impact of the project.
- Where relevant, include information on how the participants will manage the research data generated and/or collected during the project, in particular addressing the following issues:For further guidance on research data management, please refer to the H2020 Online Manual on the Participant Portal.
- What types of data will the project generate/collect?

- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- How will this data be curated and preserved? **Choirat**: Reproducibility: Encode the Data-Knowledge graph into a Reproducible-Knowledge Graph using Renku.

You will need an appropriate consortium agreement to manage (amongst other things) the ownership and access to key knowledge (IPR, data etc.). Where relevant, these will allow you, collectively and individually, to pursue market opportunities arising from the project's results.
The appropriate structure of the consortium to support exploitation is addressed in section 3.3.

- Outline the strategy for knowledge management and protection. Include measures to provide open access (free on-line access, such as the "green" or "gold" model) to peer-reviewed scientific publications which might result from the project. Open access must be granted to all scientific publications resulting from Horizon 2020 actions. Further guidance on open access is available in the H2020 Online Manual on the Participant Portal.

Open access publishing (also called 'gold' open access) means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are usually shifted away from readers, and instead (for example) to the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research.
Self-archiving (also called 'green' open access) means that the published article or the final peer-reviewed manuscript is archived by the researcher - or a representative - in an online repository before, after or alongside its publication. Access to this article is often - but not necessarily - delayed ("embargo period"), as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download/view fees during an exclusivity period.

Strategic dissemination and exploitation will help to explain the wider societal relevance and long-term economic impact of science, build support for future research and innovation funding, ensure uptake of results within the scientific community, open up potential business opportunities for novel products or services, and potentially contribute to better decision-making processes and serve as valuable input for public policies formulation. Dissemination: General dissemination targets are scientists, decision-makers, business community and the public. General dissemination measures will focus on project results and stakeholder engagement (stakeholder consultation processes; workshops to raise awareness, etc.) through:
G1. The project website will be set up within the first three months of the project.
G2. Up to date information material, e.g. brochures, presentation slides, will be distributed at events to increase awareness about our project.
G3. General other publication means will be used such as newspapers, YouTube, TV and radio, social networks (e.g., Facebook) as well as targeted mailing lists (e.g., AI-worldwide).
G4. Scientific publications for the scientific community. We will target high-level journals with open access, like Science, Nature Communication, etc.
G5. The consortium will visit conferences in the related scientific fields in order to interactively present and discuss our results with others. Among other activities, the consortium will organize special sessions at several conferences. Additionally, some targeted, specific dissemination actions will be considered: S1. We need to address mainly multipliers and developers in the ¿??¿? AI community?¿? who engage in data processing. This will be achieved by a "traveling salesman" approach using personal visits and invitations to demonstrate our system.?? S2. Target groups need to be specified and addressed. These are mainly: X departments in relevant companies in the sectors???? S3. At the end of the project we will organize a workshop specifically on X?? approaches for disseminating our results in ??? for assessing future exploitation potential, inviting partners from academia as well as industry.
1. G4 will launch a testnet to help disseminate the main results of the deep ledger knowledge network. The

launch will have invited NGO's and GO across disciplines and social, economical and technological sectors.
2. The Robhoot Open network will be launched as a Biodiversity research network to integrate the existing public databases and crowdsource data collections into the automated KGs and ledger network to facilitate NGOs, GO and other organizations transparency and governance in Biodiversity management.
3. The project aims to publish its main findings in top open scientific journals to communicate the global impact of a deep ledger knowledge network for transparency and governance across social and economical sectors.

2.2.2 Communication activities

- Describe the proposed communication measures for promoting the project and its findings during the period of the grant. Measures should be proportionate to the scale of the project, with clear objectives. They should be tailored to the needs of various audiences, including groups beyond the project's own community. Where relevant, include measures for public/societal engagement on issues related to the project.

Data management and accessibility to community: Other than being constrained by possible IPRs, 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 such as GitHub. Data (measured data), as such, will not be acquired by Robhoot. Open-source framework for delay analysis Standardized inputs 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 delay propagation modeling. 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 discovery technology to tackle global problems related to biodiversity and sustainability challenges is highly informative by itself, but a diverse group of scientists across Europe have decided that merely taking discovery alone is not enough. To understand discovery 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 graphs, the discovery knowledge graphs, into evolving federated networks to achieve cooperative forecasting. *ROBHOOT*'s goals are developed in three main milestones and ten 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 Graphs

Heterogeneous API discovery technologies to build robust and scalable automated interpretable data-driven discovery is an existing need [23, 24]. They are particularly relevant to generate explainable (or interpretable) Artificial Intelligence technologies for global emergency or sustainability landscapes where new questions and scenarios are constantly emerging [25]. Data heterogeneity, however, comprises a series of privacy requirements, formats, dimensions, biases and spatiotemporal resolution [26, 27, 28, 29]. Fortunately,

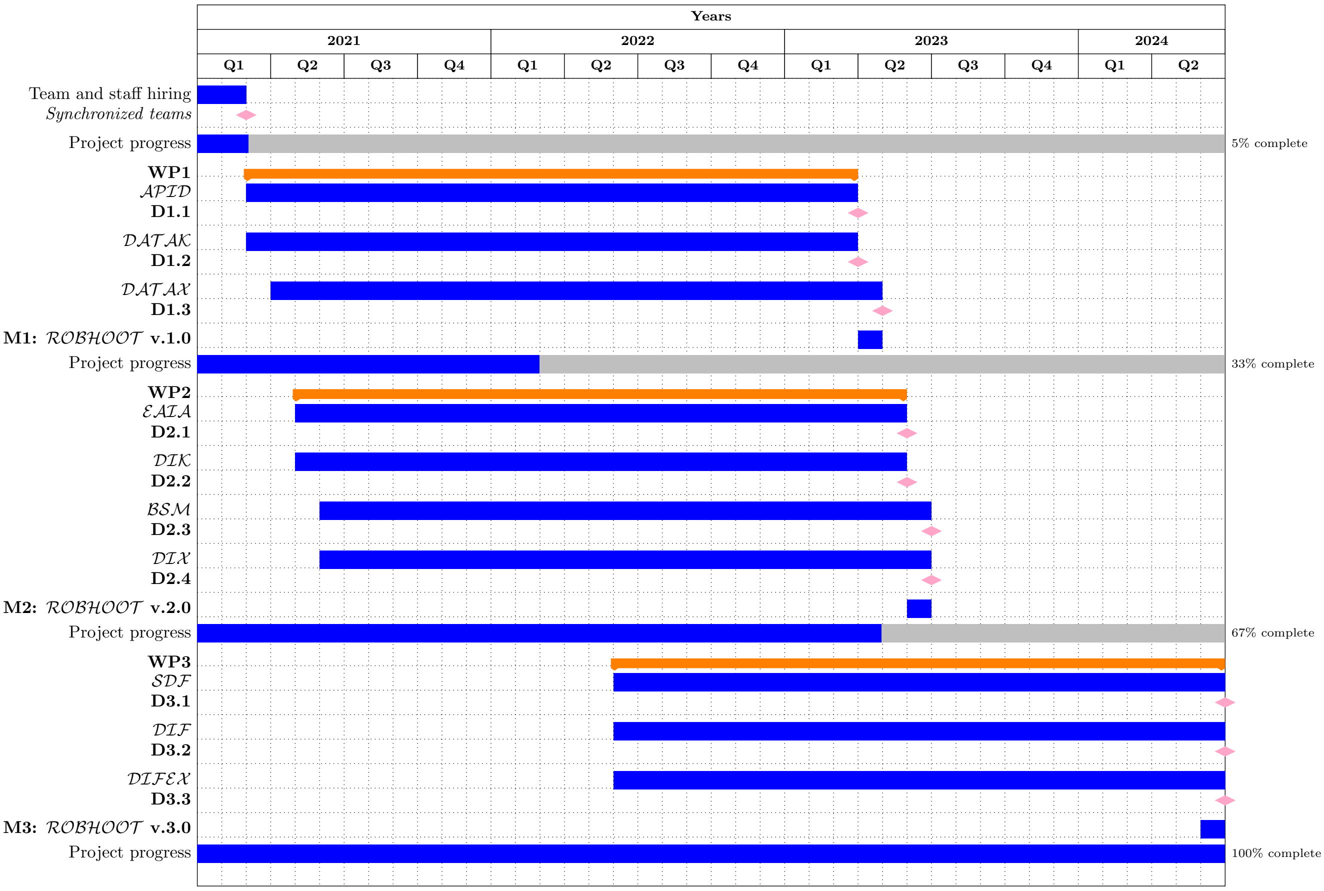

Table 3.1a: *ROBHOOT* **Gantt Chart**: *Work package one,* **WP1***, introduces Milestone ROBHOOT* **v.1.0** *and deliverables* **D1.1** *to* **D1.3** *to generate the data knowledge graph for the Exploration of the Seas Network (Figure 2). Work package two,* **WP2***, introduces ROBHOOT* **v.2.0** *and deliverables* **D2.1** *to* **D2.4** *to fussion data and causal knowlede graphs into interpretable patterns, the discovery-knowledge graphs for the Exploration of the Seas Network. Work package three,* **WP3***, introduces ROBHOOT* **v.3.0** *and deliverables* **D3.1** *to* **D3.3** *to discover knowledge graphs from biology-inspired evolving federated networks.*

standard protocols to automate API access, semantic knowledge extraction, and ETFs algorithms are rapidly advancing [23, 30, 31]. On the other side, new algorithms with different types of semantic technologies are rapidly emerging in the context of building data knowledge graphs from multiple datasets [32]. Yet, technologies focusing on (novel) evolutionary semantic algorithms to build data knowledge graphs from many heterogeneous API and data-sources that can be rapidly integrated into interpretable technologies are not currently in place [25]. *ROBHOOT* v.1.0 explores science-enabled technologies around heterogeneous API data-discovery along three main deliverables (Table 3.1a-b, **D.1.1** to **D.1.3** and Figure 3). *ROBHOOT* v.1.0 generates a data knowledge graph from heterogeneous data-sources for the exploration of the Seas network, a database initially containing 9 million entries, 1612 species, around 20 countries and 11 sampling methods (Figure 2).

3.1.2 WP2: *ROBHOOT* v.2.0: Causal Knowledge Graphs

AI is rapidly advancing in automated discovery (i.e., AutoML [10]) and becoming a more explainable or interpretable technology making more transparent the processes underlying the discovery [25, 33]. While automated and explainable discovery methods are rapidly advancing [16?], evolutionary biology-inspired methods to infer mechanistic understanding of complex problems are not currently in place (ref). Evolu-

Work package		Lead Beneficiary
Title	$\mathcal{ROBHOOT}$ **v.1.0**	
Participants	**Fortuna, Eguíluz, Choirat**	
Person Month per participant		
Start month	**3**	
End month	**27**	
Objectives	Data Knowledge Graph	
Description	Extraction diverse data-sources to infer data knowledge graphs	
Deliverables	**D1.1 (**$\mathcal{APID}$**)**: Heterogeneous API and data discovery **D1.2 (**$\mathcal{DATAK}$**)**: Heterogeneous data knowledge graph **D.1.3 (**$\mathcal{DATAX}$**)**: Data knowledge graph for the exploration of the Seas network	
Title	$\mathcal{ROBHOOT}$ **v.2.0**	
Participants	**Baity, Guimerà, Melián, Vicente**	
Person Month per participant		
Start month	**5**	
End month	**29**	
Objectives	Causal Knowledge Graph	
Description	Interpretable knowledge extraction from data knowledge graphs	
Deliverables	**D2.1 (**$\mathcal{EAIA}$**)**: Automated biology-inspired AI algorithms **D2.2 (**$\mathcal{DIK}$**)**: Discovery evolutionary biology-inspired knowledge graphs **D2.3 (**$\mathcal{BSM}$**)**: Bayesian causal-knowledge graphs **D2.4 (**$\mathcal{DIX}$**)**: Discovery knowledge graphs for the exploration of the Seas network	
Title	$\mathcal{ROBHOOT}$ **v.3.0**	
Participants	**von Waldow, Maass**	
Person Month per participant		
Start month	**18**	
End month	**42**	
Objectives	Biology-inspired Evolving Federated Network	
Description	Automated discovery knowledge graphs in federated networks	
Deliverables	**D3.1 (**$\mathcal{SDF}$**)**: Sharing discovery knowledge graphs in federated networks **D3.2 (**$\mathcal{DIF}$**)**: Discovery in biology-inspired federated networks **D2.3 (**$\mathcal{DIFX}$**)**: Discovery in biology-inspired exploration of the Seas federated networks	

Table 3.1b Work package description: *Work package, Title, Participants, Person Months per participant, Start and End month, Objectives, Description and deliverables of each Work Package.*

tionary biology-inspired techniques offer ruled-based evolution of heterogeneous agents forming network of interactions to disentangling complex sustainability problems (Figure 2). Evolutionary dynamics explores open-ended language of models with varying biologically relevant functions like code insertions, deletions, inversions and other molecular and genotype-phenotype processes to search for biologically inspired solutions in multidimensional systems to decipher complex empirical patterns.

In this regard, causal knowledge graphs connect automated and explainable AI throughout prediction and knowledge power (Figure 2). Automated and interpretable data inference still present many challenges in the context of multidimensional landscapes [34, 35]. This is particularly relevant in Earth and Ecosystem science and Biodiversity and Sustainability science [36], where merging automation to interpretable data might increase human ability to make stronger inferences about future sustainability challenges and solutions. *ROBHOOT* v.2.0 explores science-enabled technologies around multidimensional landscapes developing open-ended automated modeling, evolutionary biology-inspired solutions, AI methods and Bayesian space models along four main deliverables (Table 3.1a-c, **D.2.1** to **D.2.4** and Figure 3). In order to make inference from complex data more robust we contrast predictions from Evolutionary biology-inspired algorithms in the framework of automated Bayesian machines to explore open-ended language of models ensuring the search, the evaluation of models, trading-off complexity, fitting to the data and quantify resource usage [14, 16]. *ROBHOOT* v.2.0 deploys milestone Evolutionary automation (Figure 3.1c) for the discovery knowledge graph for the exploration of the Seas (Figure 2) containing 9 million entries, 1612 species, X countries and Y sampling methods.

ROBHOOT v.X.X	Deliver. number	Deliver. name	WP	Name Lead	Type	Disem.	Delivery date
v.1.0	**D1.1**	*APID*	WP1	**Fortuna**	OT	PU	27
v.1.0	**D1.2**	*DATAK*	WP1	**Eguíluz**	OT	PU	27
v.1.0	**D1.3**	*DATAX*	WP1	**Leads M1**	R,OT,DEC	PU	28
v.2.0	**D2.1**	*EAIA*	WP2	**Melián**	OT	PU	29
v.2.0	**D2.2**	*DIK*	WP2	**Baity/Vicente**	OT	PU	29
v.2.0	**D2.3**	*BSM*	WP2	**Guimerà**	OT	PU	29
v.2.0	**D2.4**	*DIX*	WP2	**Leads M2**	R,OT,DEC	PU	30
v.3.0	**D3.1**	*SFN*	WP3	**von Waldow**	OT	PU	42
v.3.0	**D3.2**	*DIF*	WP3	**Maass**	OT	PU	42
v.3.0	**D3.3**	*DIFX*	WP3	**Leads M3**	R,OT,DEC	PU	42

Table 3.1c List of Deliverables: *ROBHOOT contains three main work packages: ROBHOOT* **v.1.0** *span from Month 3 to 27. Deliverable* **D1.3 (***DATAX***)** *generates the data knowledge for the exploration of the Seas network. Milestone ROBHOOT* **v.2.0** *span from Month 5 to 29. Deliverable* **D2.4 (***DIX***)** *generates the discovery knowledge graph for the exploration of the Seas network, and ROBHOOT* **v.3.0** *span from Month 18 to 42, bringing the deliverable* **D3.3 (DIFX)**, *discovery in biology-inspired exploration of the Seas network.*

3.1.3 WP3: *ROBHOOT* v.3.0: Discovery Knowledge Graphs in Biology-Inspired Federated Networks

Discovery knowledge graphs are not enough if they only represent isolated contributions and can not "learn to learn" among heterogeneous groups in a biology-inspired federated network. In this regard, federated objects can be seen as "neural networks" containing many types of heterogeneous nodes with varying degrees of heterogeneity, connectivity and firing probabilities [38, 39]. Technologies in digital ecosystems around heterogenous federated networks are scarce and mostly focus on decentralization. Decentralization technologies are rapidly advancing in a variety of sectors. Most progress is coming from the scalability and security fronts [12, 40, 41, 42, 43, 44]. In the science ecosystem, only a few applications of open decentralized technologies exist [7]. Yet, sharing reproducible data and causal knowledge graphs, the discovery knowledge graphs, along biology-inspired federated networks to facilitate forecasting when heterogeneous groups interchange information is currently not in place.

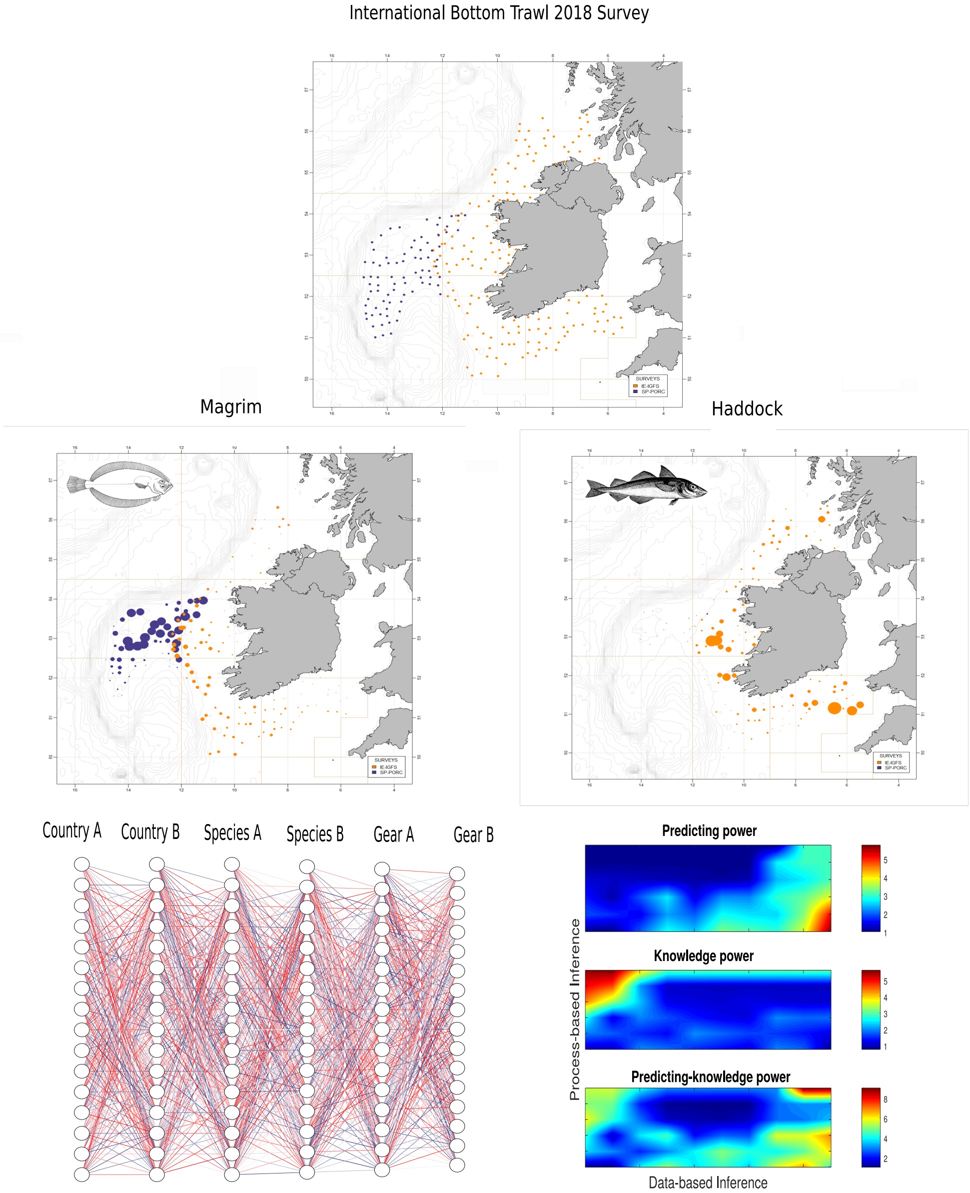

Figure 2: Causal Knowledge Graphs. Top) *The Irish Ground Fish Survey (IE-IGFS, Orange) and the Spanish Survey on the Porcupine Bank (SP-PORC, Blue) were part of the 2018 International Bottom Trawl Survey, coordinated by the International Council for the Exploration of the Sea [37]. Ireland and Spain 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 different 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).* **Bottom left** *Causal-Knowledge Graph representing the 2-countries, 2-species and 2 gears example. The whole data set for 2018 contains 11 countries, 461 fish especies (approx. 200k individuals sampled), and 5 gears.* **Bottom right** *Predictive-knowledge power map. x-axis represents "Data-based inference" (i.e., gradient of non-interpretable ML methods from left (low) to right (high) predicting power). y-axis represents "Process-based inference" (i.e., gradient of process-based methods from bottom (low) to top (high) knowledge power). The gradient of predicting power map (top) shows a hot spot red area in the bottom right highlighting the region where AI best predict the empirical data. The gradient of knowledge power map (middle) shows a red hot spot in the top left highlighting the region where the best mechanistic understanding occur. The predicting-knowledge power map (bottom) shows the sum of the two previous maps highlighting a red hot spot where predicting and knowledge power occur.*

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. $\mathcal{ROBHOOT}$ 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 Table 3.1.a-c). $\mathcal{ROBHOOT}$ v.3.0 deploys sharing protocols of discovery knowledge graphs, **D3.1**, $\mathcal{SFN}$, into biology-inspired federated networks accounting for heterogeneous agents to infer biology-inspired solutions for the exploration of the Seas federated network (Table 3.1a-c, **D.3.1** to **D.2.3**).

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	**Discovery-Knowledge Graph**	WP1-WP2	27	OS-Software,Paper/Conf.
M2	**Evolutionary Automation**	WP2	29	OS-Software,Paper/Conf.,demo-website
M3	**Cooperative Forecasting**	WP3	42	OS-Software,Paper/Conf.,main-website

Table 3.2a: List of Milestones: *$\mathcal{ROBHOOT}$ contains three milestones: $\mathcal{ROBHOOT}$* **v.1.0** *span from Month 3 to 27 to generate open-source software and research papers and/or conferences for the Data-Knowledge Graph. $\mathcal{ROBHOOT}$* **v.2.0** *span from Month 5 to 29 producing the the integration between the Data-, and the Causal-Knowledge Graph, the Discovery-Knowledge Graph, as a open-source software and research papers and/or conferences and public demo-website. $\mathcal{ROBHOOT}$* **v.3.0** *span from Month 18 to 42 to build a prototype of Discovery-Knowledge Graphs in Federated Cooperative Networks as an official $\mathcal{ROBHOOT}$ website.*

3.3 Consortium as a whole

$\mathcal{ROBHOOT}$ is a science-enabled multi-feature technology. $\mathcal{ROBHOOT}$'s consortium is designed with a highly modular structure to gain milestone's functionality (Figure 3, milestones from 1 to 3, blue, red, and pink, respectively). Connections among the modules reflect the emergence of interdisciplinarity technologies, the "Discovery Knowledge Graph", the "Evolutionary AI Automation" and the "Cooperative Forecasting" (Figure 3, green) Is the interdisciplinarity in the breakthrough idea reflected in the expertise of the consortium? How do the members complement one another?. $\mathcal{ROBHOOT}$ **v.1.0**'s team is composed by Fortuna, Eguíluz and Choirat to bring data discovery process, to fully reproducible and heterogeneous knowledge graphs (section 3.1 and Figure 3). Milestone one requires a mixture of researchers: computer-, data-scientists and developers and researchers working in complex networks from the quantitative and epistemological angles. Fortuna's, Eguíluz and Choirat's expertise complement each other's roles: Fortuna's team takes care of data knowledge graphs following evolutionary semantic algorithms for novel data-interactions and API discovery (i.e., $\mathcal{APID}$ and $\mathcal{DATAK}$, $\mathcal{DATAX}$). Eguíluz's team focuses on network modularity, community detection and decentralization metrics for pattern detection in data knowledge graphs (i.e., $\mathcal{DATAK}$ and $\mathcal{DATAX}$, and Choirat's team encodes all the algorithms and procedures from Fortuna's and Eguíluz's teams into reproducible knowledge graphs. Milestone $\mathcal{ROBHOOT}$ **v.1.0** generates a data knowledge graph for the exploration of the Seas (Figure 3, blue).

$\mathcal{ROBHOOT}$ **v.2.0**'s team composed by Guimerà, Baity, Vicente, and Melián fussion Bayesian Machine Scientist to Evolutionary and AI Algorithms, forming the "Evolutionary Automation" approach (Figure 3, green). The "Evolutionary Automation" fussion data to causal knowledge graphs to make patterns interpretable (Figure 2). The team for this milestone add complementarity expertise to $\mathcal{ROBHOOT}$ **v.1.0**'s team: Now the skills focus on data-scientists trained in deep learning networks and automation algorithms, theoreticians with expertise in Bayesian inference, and evolutionary biologists with expertise in evolutionary ecology theory and evolutionary-inspired networks (section 3.2 and Figure 3, red). Despite modules $\mathcal{ROBHOOT}$ **v.1.0** and $\mathcal{ROBHOOT}$ **v.2.0** focus on specific milestones and deliverables (Table 3.1a-c), they connect each other along data and causal knowledge graphs, the discovery knowledge graphs, and evolutionary automation to build a interdisciplinarity science-enabled technology that can be compactly converted into user-friendly open-software (**Discovery Knowledge Graphs** and **Evolutionary AI Automation**, green). Milestone $\mathcal{ROBHOOT}$ **v.2.0** generates a discovery knowledge graph for the exploration of the Seas initially containing 9 million entries, 1612 species using around 11 sampling methods and more than 15 countries (Figures 2 and 3, red). Thus, interdisciplinarity enters not only at the intra-module development stage, but also at the inter-module stage where discovery-knowledge graphs and evolutionary AI automation form the basis for a interdisciplinarity breakthrough reflected in the highly complementarity skills of the consortium (section 4.1). The first two modules in $\mathcal{ROBHOOT}$ contain researchers from Estonia, Spain, Switzerland and Sweden.

The $\mathcal{ROBHOOT}$ consortium wants to advance the rapidly evolving digital ecosystem by making cooperative discovery a fundamental feature of it. For this purpose, a science-based automated and interpretable technology is not enough if each discovery knowledge graph stays isolated from one another. To contrast robustly interpretable scenarios in the face of global sustainability challenges, discovery knowledge graphs should learn to learn from heterogeneous data-sources in the contexxt of evolutionary biology-inspired federated networks. To achieve scalability for the discovery knowledge graphs, neural-inspired protocols in federated networks is the excellency feature of $\mathcal{ROBHOOT}$ **v.3.0** (section 3.3). $\mathcal{ROBHOOT}$ **v.3.0**'s team composed by von Waldow and Maass, develops protocols for sharing discovery knowledge graphs along biology-inspired federated networks. The team forming $\mathcal{ROBHOOT}$ **v.3.0** therefore requires quite a lot of contrasting skills. First, developers working in P2P and security protocols. Second, social scientists, computer scientists, and neurobiologists in collaboration to developers aiming to explore the role of heterogeneous groups of biology-inspired neurons accounting for heterogeneous data-sources in federated networks. Milestone $\mathcal{ROBHOOT}$ **v.3.0** is a fundamental stepping-stone for developing "Cooperative Forecasting": it first guarantees discovery knowledge graphs are reproducible shareable objects. Yet, in the same way than evolutionary algorithms and the Bayesian machine scientist search automatically for open-ended space models to generate the plausible causal knowledge graphs, the discovery knowledge graphs produced in different nodes of a network need to automatically interact and learn from each other to find better forecasting scenarios at a global scale. $\mathcal{ROBHOOT}$ **v.3.0**'s implements heterogeneous groups of (cooperating and competing) neurons in federated networks for making cooperative forecasting a standard global property. Milestone $\mathcal{ROBHOOT}$ **v.3.0** generates a discovery federated network for the exploration of the Seas to provide populations of scenarios satisfying biodiversity maintenance while guaranteeing commercial interest (Figure 3, pink). $\mathcal{ROBHOOT}$ v.3.0 contain researchers from Switzerland and Austria.

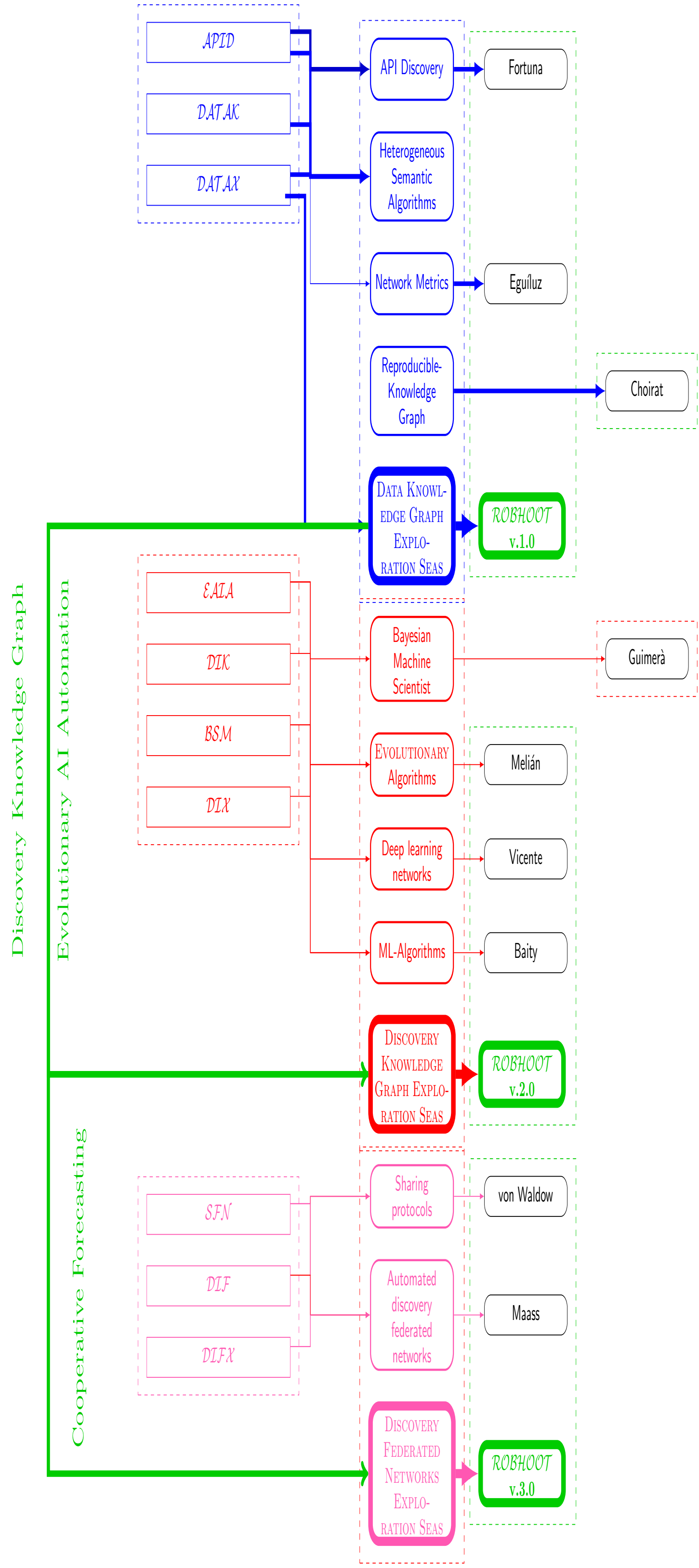

Figure 3: ROBHOOT Consortium: ROBHOOT **v.1.0** *(blue) to ROBHOOT* **v.3.0** *(pink) with acronyms of each deliverable (Left column), tasks (Center), lead and partner names (Right columns). Links connect deliverables to tasks and leading/partners groups. ROBHOOT delivers three interdisciplinarity-driven science-enabled technologies:* **Discovery-Knowledge Graph** *connecting ROBHOOT* **v.1.0** *and ROBHOOT* **v.2.0. Evolutionary AI Automation** *in ROBHOOT* **v.2.0***, and* **Cooperative Forecasting** *connecting ROBHOOT* **v.2.0** *to* **v.3.0**

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] 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.

[14] 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.

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

[16] 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.

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

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

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

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

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

[22] 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.

[23] 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/`.

[24] 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`.

[25] 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.

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

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

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

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

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

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

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

[33] 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.

[34] 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.

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

[36] 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.

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

[38] 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.

[39] 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.

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

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

[42] 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.

[43] 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/.

[44] 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.