ROBHOOT

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

May 8, 2020

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 goals are at a very incipient stage of development. We introduce data- and causal-knowledge graphs, the discovery-knowledge graphs, in 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 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

- Describe the vision of a radically-new science-enabled technology that the project would contribute towards
- The project will contribute towards automated discovery-knowledge graphs accounting for heterogeneous data-individuals-groups-societies interchanging functional information to reach partial solutions in the face of rapidly emerging global sustainability challenges.
- Describe how this vision surpasses substantially any technological paradigms that currently exist or are under development.
- Current science-enabled technological paradigms are rooted in (single or multiple) optimization functions. Yet, biology-inspired technologies targeting learning and cooperation in evolving highly heterogeneous systems is currently not in place. In addition, technological paradigms for discovery targeting global sustainability are currently based on highly fragmented, non-reproducible, and competitive technologies. Discovery-knowledge graphs in cooperative networks instead focus on the need of rapidly responding to global sustainability challenges emphasizing learning, cooperation, and reproducibility in networks compossed by heterogeneous groups.
- 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).

- *ROBHOOT* will be developed in three stages, each containing measurable and achievable goals, work packages, milestones and deliverables (Figures 2 and 4 and Tables 2 and 3).

We are in the midst of the fourth industrial revolution, a transformation revolving around data driven intelligent machines potentially impacting strongly 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 (**Ref needed**). (**more flow here**) People are applying technology in powerful ways, from adopting decentralized technologies for humanitarian efforts to improving agricultural practices and reducing waste in the global food supply chain ([2],++). Big Data analytics is advancing at the pace dictated by the availability of data and a myriad of approaches are being developed to extract patterns from data (review neural networks). Most of these approaches focus on a small number of data sources and groups (refs about heterogeneous and small groups in data collection). AI approaches are also rapidly evolving towards more explainable (i.e., interpretable) pattern description making information about the patterns more accessible (refs). Therefore, data-driven approaches are built on a few sources of data showing a quite constrained, but rapidly evolving (refs), number of interpretable science-enabled technologies (**Maisonobe, M., Eckert, D., Grossetti, M., Jgou, L., Milard, B. (2016). The world network of scientific collaborations between cities: domestic or international dynamics?. Journal of Informetrics, 10(4), 1025-1036**). (this is the gap) In this regard, data collected from many distinct sources together with science-enabled technologies accounting for functional interactions among distinct groups making emphasis on interpretable patterns are required to gain more robust information access in knowledge-inspired societies. Yet, discovery science-enabled technologies accounting for fully reproducible heterogeneous and interpretable data-sources applied to information access of complex governance, social, environmental and technological problems are particularly lacking (Figure 1 and Table 1 (remove table 1)) [1].

Taken together, 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 and therefore a limited number of interpretable patterns (mention large heterogeneity shown in many datasets, despite they are only a few data sets). Second, science-enabled technologies mostly focus on optimization rules (i.e., function loss or reward, similar to fitness optima functions in evolutionary biology, refs). Optimization-based technologies limit a broader number of plausible solutions, as usually found in evolving learning capable biological systems (refs). 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 of data, technology should be able to (automate) data integration from a large number of sources to make it available for analysis. Second, the analysis of the data should go beyond the identification of patterns and consider approaches delivering processes, and the comparison of several processes in an automated fashion, that is, assisting the end-user. Third, to leverage the computing capacity the analysis should be performed in federate way, such that highly heterogeneous populations can learn from each other to reach consensus about the populations of plausible scenarios that best capture the properties of the integrated data, and finally, the whole process should be fully reproducible and transparent such that benefit the public. (only here we show the vision of a radically-new science-enabled technology that the project would contribute towards) Our project contributes towards a science-enabled technology accounting for massive data heterogeneity and interpretable patterns, the discovery knowledge graphs (Table 1 and XX), obtained from evolving learning (neural) information processing systems. (the dream) Evolutionary learning-inspired technology extracts information from highly heterogeneous and multidimensional federate networks minimizing the need of having optimal solutions and making possible the emergence of functional information processing to reach consensus about populations of partial solutions to enrich knowledge-inspired societies facing global challenging problems.

Learning to learn in evolving biological systems have shown multiple plausible solutions can be obtained all expressing highly heterogeneous populations (many solutions are not optimal but suboptimal, maladaptative and require multiple trait dimensions where some are optimized while other are not, refs). Many branches of evolutionary algorithms have attempted to capture the patterns and the mechanisms of the multiple plausible solutions showing functional communities (evolutionary algorithms, genetic algorithms, the evolution strategy, evolutionary programming and swarm intelligence. These techniques form the basis of several disciplines such as artificial life and evolutionary robotics, refs). On ecological systems, intraspecific trait variation and trait dimensionality (i.e., biotic, reproductive, abiotic and migration traits for example) drive population heterogeneity making difficult to anticipate functional interactions with other species (i.e., cooperative, antagonistic, competitive, or mutualistic, infectious diseases... etc). (On neural systems, the vast majority of neurons in the brain show highly differentitated morphological, genetic and phenotypic states, and making functional interactions among such a highly differentiatied states (groups, etc) is yet not well understood.) Taken together, these results show that the understanding of 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 [3]), is currently quite limited and that new science-enabled approaches accounting for diversification in highly distinct forms and functions in functional information processing federated networks, are required to develop science-enabled technologies ...

Biodiversity dynamics in space and time is a good example of an evolving information-processing network of interacting species composed by highly heterogeneous individuals within and between populations – a highly heterogeneous mixed pool of evolving interactions in federated networks (i.e., food webs, mutualistic networks, hybrid networks, interaction with different types, etc — we have expertise in biodiversity and ecological networks – here to show more convincingly we are moving beyond our expertise). Many of these networks are currently loosing species and occur in highly fragmented and disturbed ecosystems. This makes these systems also highly vulnerable to the emergence of new interactions many of them occurring between infectious diseases and humans (i.e., Many zoonosis have occurred recently with a dramatic rise in global pandemics during the last decade, i.e., the SARS pandemic in 2003, to Avian Influenza in 2006, H1N1 in 2009, Ebola in 2014, the appearance of the Zika virus in Latin America in 2015, and the current Covid-19 pandemic (Zoonosis from Eating animals or from farms but in any case it is a new and novel interaction), with these developments inextricably bound up in modern socio-technical developments and processes of globalization (Show ++ evidence about the connection between biodiversity loss and increasing risk of zoonotic diseases – here enter species interactions in heterogeneous populations). Unfortunately, science-enabled technologies facilitating rapid sharing of data and information among highly heterogeneous groups to mitigate risks and enhance global cooperative forecasting to efficiently respond with informed scenarios are particularly lacking ([2], other refs). The current pandemic is teaching us the need of science-enabled technologies facilitating more robust scientific collaboration for more rapid reaction and real-time interpretable (i.e., causal) maps accounting for massive data-heterogeneity to overcome fragmented and partial responses to a global problem. We are responding with many independent webpages offering data (refs), incipient data-knowledge graphs (ref), and many epidemiological models with different assumptions (refs).

All these examples offer an opportunity to build more integrative and innovative technologies, like the emergence of discovery-knowledge graphs in ... The goal of *ROBHOOT* is to propose a new hybrid-technology concept integrating data- and causal-knowledge graphs into intelligent networks to lay the foundation for a novel scientific discovery technology. *ROBHOOT* will contribute towards facilitating governance reproducible scenarios in rapidly changing global sustainability landscapes. *ROBHOOT* will be developed along science-enabled technologies (Figures 1, 2, and 4): *ROBHOOT* **v.1.0** deploys question- and data-knowledge graphs for understanding bias and diversification of information sources when obtaining global data-architecture maps. *ROBHOOT* **v.2.0** integrates automated and explainable biology-inspired neural federated networks to decipher causal-knowledge graphs from data-knowledge graphs, and *ROBHOOT* **v.3.0** explores (automated) discovery-knowledge graphs in federated cooperation networks with a case study mapping the biodiversity and the risk of zoonosis between humans-infectious diseases at the global scale.

1.2 Science-to-technology breakthrough that addresses this vision

- Discuss the relevant state-of-the-art and the extent of the advance the project would provide beyond this state-of-the-art

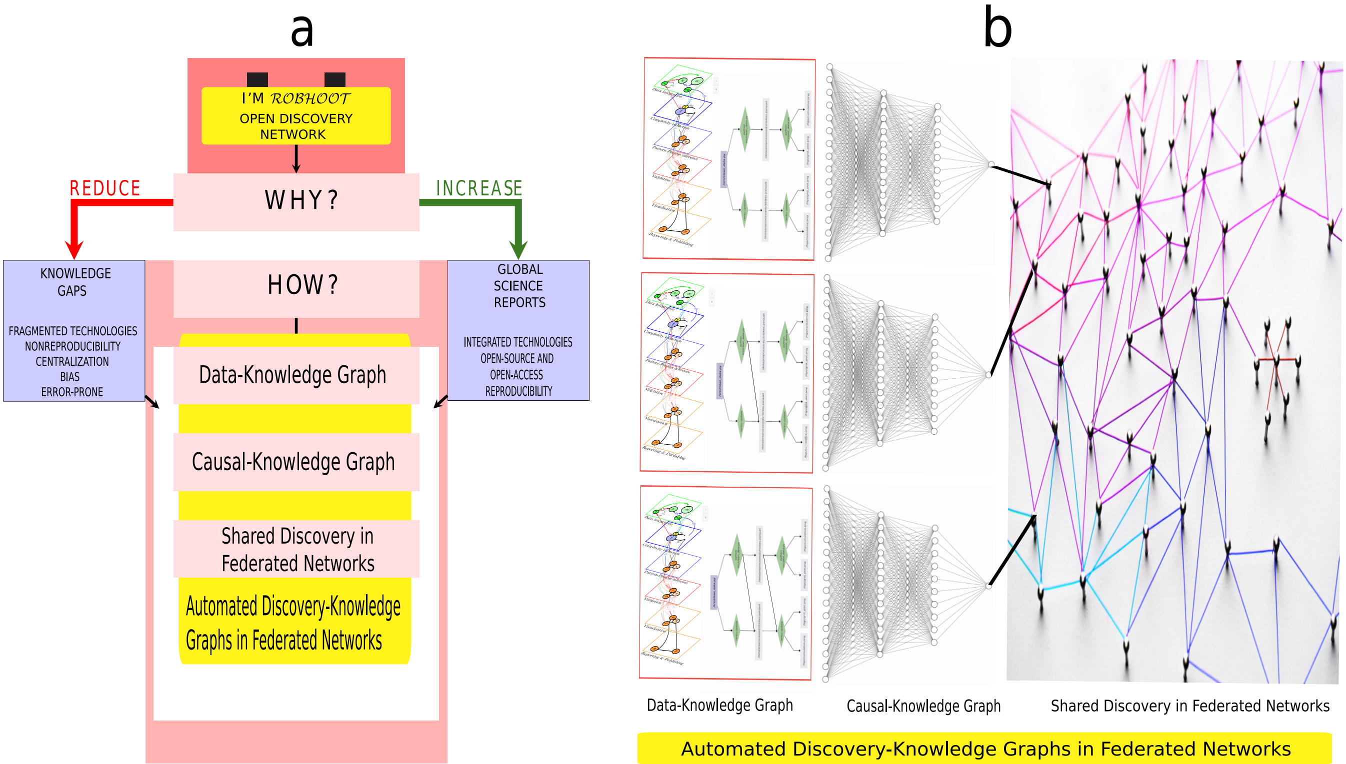

Figure 1: Cooperative Discovery in Federated Networks. *ROBHOOT integrates data- and causal-knowledge graphs, the discovery-knowledge graph, into cooperative federated networks for a sustainable knowledge-inspired society:* **a)** *ROBHOOT targets global knowledge gaps (red path) and open-access reproducible discovery reports (green path). It integrates three science-enabled technologies:* **a,b) Data-Knowledge Graphs** *for discovering global data-architecture.* **a,b) Causal-Knowledge Graphs** *fussion automated and explainable biology-inspired neural networks discovery, and* **a,b) Shared Discovery in Federated Networks** *for cooperative discovery and forecasting.* **Automated Discovery-Knowledge Graphs in Federated Networks** *integrates data- and causal-knowledge graphs into intelligent federated networks to generate robust cooperative forecasting to rapidly respond to global emergency and sustainability challenges.*

- The state-of-the-art of automated and interpretable discovery is currently a fragmented landscape. The result is a slow response to the rapidly growing global emergency and sustainability challenges. *ROBHOOT* will go beyond the state-of-the-art in automated and interpretable discovery: First, it will introduce automated data- and explainable-knowledge graphs into a more compact and robust discovery-knowledge graph technology. Second, it will fussion the discovery-knowledge graph technology within cooperative federated networks to make discovery a rapidly evolving feature responding to the also rapidly evolving global emergency and sustainability challenges.
- 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* will go beyond the state-of-the-art of knowledge-graphs by developing data- and causal-knowledge graphs, the discovery-knowledge graphs, in cooperative intelligent federated networks to move knowledge-inspired societies towards reaching global sustainability goals.

Interconnected global societies constantly face new challenges that need to be rapidly addressed. Yet, 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 [11, 12, 13, 14, 15, 16, 17, 18, 19], the integration of science-to-technology intelligent automation networks currently

Word	Meaning
Question-knowledge graph	Technology-driven information extraction from corpus or similar to detect question gaps in multidisciplinary research
Data-knowledge graph	Technology-driven information extraction from diverse data-sources to infer global data-architecture
Causal-knowledge graph	Technology-driven information extraction to provide explainable/interpretable scenarios on global and complex sustainability challenges
Discovery-knowledge graph	Novel interactions emerging from the integration of data- and causal-knowledge graphs to provide multidisciplinary responses to global sustainability challenges
Automation	Algorithms targeting minimal human-driven interference
Knowledge-inspired society	Open-access discovery to take informed decisions in global sustainability challenges
Neutral-knowledge generation	Open reproducible reports making transparent the many sources of bias in the discovery process

Table 1: Glossary of terms.

lack knowledge-inspired technologies impacting knowledge-inspired societies to help responding to rapidly evolving global sustainability challenges (Figure 1 and Table 1). In this regard, technologies facilitating rapid access to global API-structured data to analyze global data architecture present many challenges. This is particularly relevant in global emergency or sustainability landscapes, where data properties like availability, accuracy and transparency drive constantly emerging feedbacks between questions and scenarios to predict new situations more accurately.

Redundant – *ROBHOOT* **v.1.0** deploys a data discovery technology to generate question- and data-knowledge graphs for a rapid understanding of global data-architecture. Data-architecture alone is not sufficent to outline predictive scenarios in complex sustainability problems. Therefore, data analytics complementing data-architecture discovery and mechanistic inference is desirable to interpret scenarios in rapidly global emerging challenging situations. In this regard, there are also many gaps in connecting global data-architecture into rapid automated causal-knowledge graphs to facilitate discovery that can be transferred to governance decisions. *ROBHOOT* **v.2.0** integrates automated and explainable biology-inspired neural-networks to decipher causal-knowledge graphs from open-ended modeling scenarios. Still, rapidly drawing scenarios from a few labs limit the phase space from where the discovery process is generated. Therefore, the scalability of fully reproducible discovery strongly depend on cooperation and learning in decentralized networks. *ROBHOOT* **v.3.0** explores sharing protocols of discovery-knowledge graphs in federated networks (section 3.3). Finally, *ROBHOOT* **v.4.0** integrates *ROBHOOT* v.1.0 to v.3.0 into automated discovery in federated cooperation networks (Figures 2 and 4).

1.3 Interdisciplinarity and non-incrementality of the research proposed

- Describe the research disciplines necessary for achieving the targeted breakthrough of the project and the added value from the interdisciplinarity.
- Explain why the proposed research is non-incremental.

Respond more directly –
ROBHOOT is a science-enabled multi-feature hybrid-technology for automating interpretable data-driven discovery in intelligent and cooperative federated networks (Figures 1 to 4 and Tables 1 to 3). It will contain four milestones each characterized by a mixture of research disciplines necessary for achieving interdisciplinary breakthrough. *ROBHOOT* **v.1.0** is compossed by computer scientists and developers targeting API discovery

protocols and ETLs algorithms. This module is complemented with scientists from complex networks to develop quantitative methods for question- and 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 will generate a diverse team targeting synthesis between automated and explainable biology-inspired neural-networks to decipher causal-knowledge graphs from data-architecture properties (section 3.2). *ROBHOOT* **v.3.0** is characterized by combining computer scientists and developers targeting decentralized protocols (federated networks, gnunet), with social scientist, and scientists specialized in ecology and evolutionary biology. Team for *ROBHOOT* v.3.0 explores cooperation protocols for discovery in federated networks (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 [20]. *ROBHOOT* **v.4.0** combines computer- and data-scientists working in modules one and two, respectively, with developers, biologists, evolutionary biologists and social scientists working in modules 2 and 3. Such a diverse team will integrate automated and explainable data- and causal-knowledge graphs into federated cooperation networks to generate automated reporting for global emergency and sustainability problems (Figure 4).

ROBHOOT aims to bring global transparency in knowledge generation by acting as an assistant to humans or as a automated and reproducible discovery generator to facilitate sustainability goals of humanity. 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] (Figure 1 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 intergating 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 cooperative networks still require the filling of many existing technological gaps, from identification and retrieval of heterogeneous data sources, to the integration of explainable modeling and causal inference and the learning capabilities of cooperative forecasting accounting for many evolving agents.

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 your 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.
- Decision making and governance at local, regional and global scales require access to transparent and reproducible data-, and -causal-knowledge graphs, the discovery-knowledge graphs, to analyze local

solutions benefiting society in real-time in emergency situations.

- Describe the empowerment of new and high-potential actors towards future technological leadership.
- 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
- 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.
- 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 automated forecasting and interpretable information is key to draw rapid and robust scenarios when facing complex problems including global sustainability challenges (i.e., global health, ecosystems degradation, warming, etc). *ROBHOOT* contributes to **(evolutionary) automation**, **cooperative forecasting** and **interpretable information** for a new science-enabled technology targeting knowledge-inspired societies: First, (evolutionary) automation decipher open-search interpretation of complex systems. Second, global hybrid (i.e., humans and machines) 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. 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....

Legal and financial transparency

Technological Social and Governance

Impact to emerging and sustainability challenges ::

Novel service for NGO, society and thinktank transparent and reproducible public policies:

Advisory boards ::

Sustanaibility – 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 in 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 world-wide). Thus, this consortium is at the leading edge.

2.2 Measures to maximize impact

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:[1]
  - 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?

  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.[2]

  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

[1]For further guidance on research data management, please refer to the H2020 Online Manual on the Participant Portal.

[2]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.

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.

1. The contribution in communication of the Swiss Data Science Center, Switzerland
2. Contribution of the Wyss center
3. Contribution of Ifisc, Spain

3 Implementation

- Describe here the objectives, list of work packages, list of deliverables (Ghentt chart)

Automating the discovery process to tackle rapid global solutions to humanity challenges is highly informative by itself, but a diverse group of scientists across Europe have decided that merely taking discovery alone is not enough. Science is a highly dynamic and global process and there are many paths from where it can be achieved. To understand discovery broadly, these scientists want to advance the automation and cooperative discovery in the global digital ecosystem. To this end, the *ROBHOOT* consortium aims at developing a federated network integrating several technologies into a unified framework. *ROBHOOT* will develop quantitative novel methods such as question-, data-, and causal-knowledge graphs, the discovery-knowledge graph, to understand how cooperative discovery networks might help towards knowledge-inspired societies to provide scenarios in face of global sustainability challenges. This strategy is expected to improve early access to discovery to rapidly act in emergency global situations or sustainability challenges to indentify new emerging targets where automation and global reports can play a key role in knowledge-inspired societies. *ROBHOOT*'s goals are developed in four different stages with four main milestones and sixteen deliverables (Figure 4).

3.1 Research methodology and work plan – Work packages, deliverables

3.1.1 WP1: *ROBHOOT* v.1.0: Data-Knowledge Graphs

- Rapid API access to build robust and scalable automated interpretable data-driven discovery as an existing need. This is particularly relevant in global emergency or sustainability landscapes.
- Data- and Question-Knowledge graphs as solutions for rapid data-driven discovery (Table 1). DKG explores similarity patterns of database to discover existing gaps in data availability
- Global and rapid API access to build robust and scalable question- and data-knowledge graphs as a case study for automated interpretable data-driven discovery.
- See Case Study Figure 4

Global and rapid access to data to build robust and scalable question- and data-knowledge graphs is key for automating interpretable data-driven discovery. This is particularly needed in emergency or sustainability challenging situations at the global scale, where new questions and scenarios are constantly emerging and data access with different privacy requirements, formats, heterogeneity, dimensions, bias and spatiotemporal resolution is the norm [21, 22, 23, 24]. Yet, available automated science-enabled technologies to build data- and question-knowledge graphs to rapidly inform causal-knowledge graphs are missing. Fortunately, standard protocols to automate data API access, knowledge extraction, and ETFs algorithms are rapidly advancing [25, 26, 27], but the technologies around automated API data-discovery and question- and data-knowledge graphs remain difficult to compactly link to the automated causal-knowledge graphs for scientific discovery (Table 3.1.1a and 3.1b, Work packages and deliverables).

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

- Contrasting explainable biologically inspired Causal-Knowledge Graphs for interpretable information when dealing with complex sustainability challenges.
- See Figures 2 and 4
- Contrasting predictions from causal open-ended language of models combining Bayesian networks and optimization methods
- See Figures 2 and 4

AI is rapidly advancing in automated discovery (i.e., AutoML [10]) making more transparent the processes underlying the discovery (i.e., Explainable or interpretable AI [28, 29]). Yet, automated and explainable discovery methods are still at an incipient stage of integration, particularly in open-ended Bayesian machines [14]. This is particularly relevant in the context of biology, brain research, and evolutionary biology techniques where making automatic interpretation of complex systems can provide scenarios to help disentangling complex sustainability problems for humanity. *ROBHOOT* v.2.0 will develop novel causal knowledge graphs integrating automated and explainable discovery accounting for evolutionary biology and AI techniques [30, 31]. Automatic interpretation of the causal processes underlying empirical patterns are explored using evolutionary computing and deep learning networks (Table 3.1.1b, (**D2.1: Evolutionary Artificial Inteligence Algorithms (*EAIA*)**).

Evolutionary dynamics explore 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 automated biologically inspired solutions to complex empirical patterns (**D2.2: Causal-Knowledge Graphs**

Work package		Lead Ben-efi-ciary
Title	$\mathcal{ROBHOOT}$ **v.1.0**	
Participants	**Fortuna, Eguíluz, Choirat**	
Person Month per participant		
Start month	**3**	
End month	**27**	
Objectives	Automated Data-Knowledge Graph: Generalized Algorithms and a case study for COVID-19	
Description	$\mathcal{APID}$ $\mathcal{QKG}$ $\mathcal{DKG}$ $\mathcal{CODA}$	
Deliverables	**D1.1:** $\mathcal{APID}$: Technologies around API data-discovery to obtain global knowledge-graphs remain difficult to automate despite standards and protocols are rapidly emerging [25, 26, 27]. **Fortuna**: Check existing protocols and gaps, ODBMS.org and others to build automated knowledge graphs. Automated workflow to build knowledge-graphs from a large number of API to generate a global data-architecture map. **D1.2:** $\mathcal{QKG}$ **Fortuna**: Question-Knowledge graphs from a large number of corpora datasets using automated extraction algorithms. **Choirat**: Convert Question-Knowledge graph into a Reproducible-Knowledge Graph using Renku. **D1.3:** $\mathcal{DKG}$ **Fortuna**: Data-Knowledge graphs from a large number of datasets using automated API algorithms. **Eguíluz**: Implementation modularity and community metrics to detect gaps in the question-, and data-knowledge graphs. **Choirat**: Encode the Data-Knowledge graph into a Reproducible-Knowledge Graph using Renku. **D.1.4:** $\mathcal{CODA}$ is to produce an Automated Data-Knowledge Graph for the COVID-19, extending the COVID-19 Data-Knowledge Graph built from many data-sources [40]. **Fortuna**, **Eguíluz**, and **Choirat** generate a reproducible Automated Question- and Data-Knowledge Graph for the COVID-19.	

Table 3.1.1b Work package description: *Work package, Title, Participants, Person Months per participant, Start and End month, Objectives, Description and deliverables of each Work Package.*

($\mathcal{CKG}$**)**, Figure 3a). Causal-knowledge graphs enhances the connection between automated and explainable AI throughout prediction and knowledge power (Figure 3b). Automated and interpretable data inference still present many challenges in the context of multidimensional landscapes [32, 33]. This is particularly relevant in Earth and Ecosystem science and Biodiversity and Sustainability science [34], where merging automation to interpretable data might increase human ability to make stronger inferences about future sustainability challenges and solutions. In order to make inference from complex data more robust we contrast predictions from Evolutionary Neural Networks in the framework of Bayesian Space Models to explore open-ended language of models combining Bayesian networks and optimization methods (**D2.3: Bayesian Space Models (**$\mathcal{BSM}$**)**. The Bayesian space models module ensures the search, the evaluation of models, trading-off complexity, fitting to the data and quantify resource usage [12, 14]. $\mathcal{ROBHOOT}$ v.2.0 deploys

Work package		Lead Beneficiary
Title	**$\mathcal{ROBHOOT}$ v.2.0**	
Participants	**Baity, Guimerà, Melián, Vicente**	
Person Month per participant		
Start month	**5**	
End month	**29**	
Objectives	Automated Causal-Knowledge Graph: case study for COVID-19	
Description	$\mathcal{EAIA}$ $\mathcal{CKG}$ $\mathcal{BSM}$ $\mathcal{COCAU}$	
Deliverables	**D1.1:** $\mathcal{EAIA}$: Automated evolutionary AI Algorithms with open-ended (evolving) functions with different degrees of complexity to find expressions (i.e., interpretable models towards causal-knowledge graphs) best fitting the empirical patterns observed in the Data-Knowledge Graphs [12, 14]. Merge Evolutionary algorithms (i.e.,, mutation-selection-migration-recombination algorithms (**Melián**, **Guimerà**) with deep neural networks (**Baity**, **Vicente**). **D1.2:** $\mathcal{CKG}$ **Baity, Guimerà, Melián, Vicente**: Discovery-Knowledge Graphs generated from the fussion of Data-, and Causal-Knowledge graphs (Figure 3) using automated evolutionary AI Algorithms. **Choirat**: Convert Causal-Knowledge graph into a Reproducible-Knowledge Graph using Renku. **D1.3:** $\mathcal{BSM}$ **Baity, Guimerà, Melián, Vicente**: Bayesian Inference to estimate Causal-Knowledge graphs accounting for model complexity. Check link to automated evolutionary AI algorithms via open-ended (evolving) functions. **D.1.4:** $\mathcal{COCAU}$ **Baity, Guimerà, Melián, Vicente** Bayesian Inference to estimate Causal-Knowledge graphs accounting for model complexity for the COVID-19 as a case study. Check link to automated evolutionary AI algorithms via open-ended (evolving) functions.	

Table 3.1.2b Work package description: *Work package, Title, Participants, Person Months per participant, Start and End month, Objectives, Description and deliverables of each Work Package.*

the COVID-19 pandemic as a case study to automatically infer interpretable causal-knowledge graphs at global scale (**D2.4: COVID-19 Causal-Knowledge Graphs (**$\mathcal{COCAU}$**)**).

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

- A science-based explainable-knowledge graph technology is not enough if we aim to globally contrast robustly paths for interpretation of causal process predicting global data-architectures.
- A science-based technology to develop shared cooperative forecasting in federated networks in face of rapidly emerging global emergency and sustainability challenges.
- Science-based explainable discovery-knowledge graphs shared in a federated network is not enough if most contributions can not be reproduced and strongly depend on competitive schemes. This might be

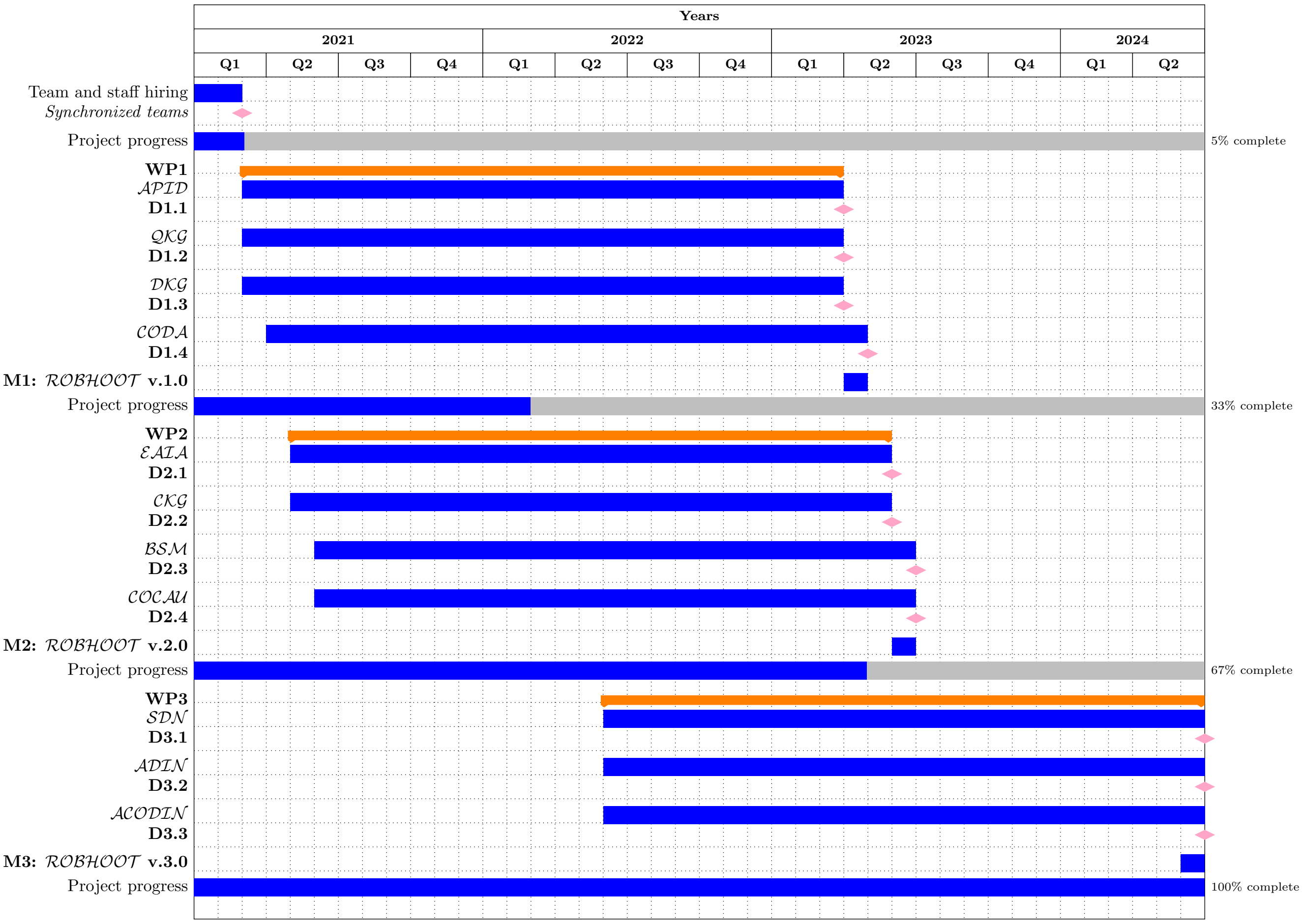

Figure 2: ROBHOOT **Gantt Chart**: *Work package one,* **WP1**, *introduces Milestone* ROBHOOT **v.1.0** *and deliverables one to four* (**D1.1** *to* **D1.4**) *to bring question- and data-knowledge graphs into a global data-architecture map. Work package two,* **WP2**, *introduces Milestone* ROBHOOT **v.2.0** *and deliverables five to eight* (**D2.1** *to* **D2.4**) *to fussion automation and interpretable patterns into discovery-knowledge graphs. Work package three,* **WP3**, *introduces Milestone* ROBHOOT **v.3.0** *and deliverables nine to eleven* (**D3.1** *to* **D3.3**) *to connect shared discovery-knowledge graphs to automated federated cooperative networks.*

particularly relevant in the face of rapidly responding to global emergency and sustainability challenges.

A science-based explainable technology is not enough if we aim to globally contrast robustly interpretable causal processes predicting the global data-architecture obtained from merging data- and causal-knowledge graphs. Similarly, science-based explainable discovery-knowledge graphs shared in a federated network are not enough if most contributions can not be reproduced and strongly depend on competitive schemes. In this regard, federated objects can be seen as networks containing many types of nodes (i.e., like a diversity of neurons) with varying connectivity and firing probabilities [30, 31] sharing causal-knowledge graphs according to given rules to find populations of causal-knowledge graphs that best fit to the data-knowledge graphs patterns. Sharing data- and causal-knowledge graphs need novel (firing, spiking... etc) protocols to enhance forecasting and strong inference in federated networks. ROBHOOT v.3.0 develops protocols in digital networks to embed explainable discovery-knowledge graphs into global cooperation schemes to increase robustness and reproducibility of discovery-knowledge graphs [3]. Technologies in decentralized digital ecosystems are rapidly advancing in a variety of sectors. Most progress is coming from the scalability, security and decentralization fronts [3, 35, 36, 37, 38, 39]. 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 cooperative networks to facilitate forecasting is currently not in place.

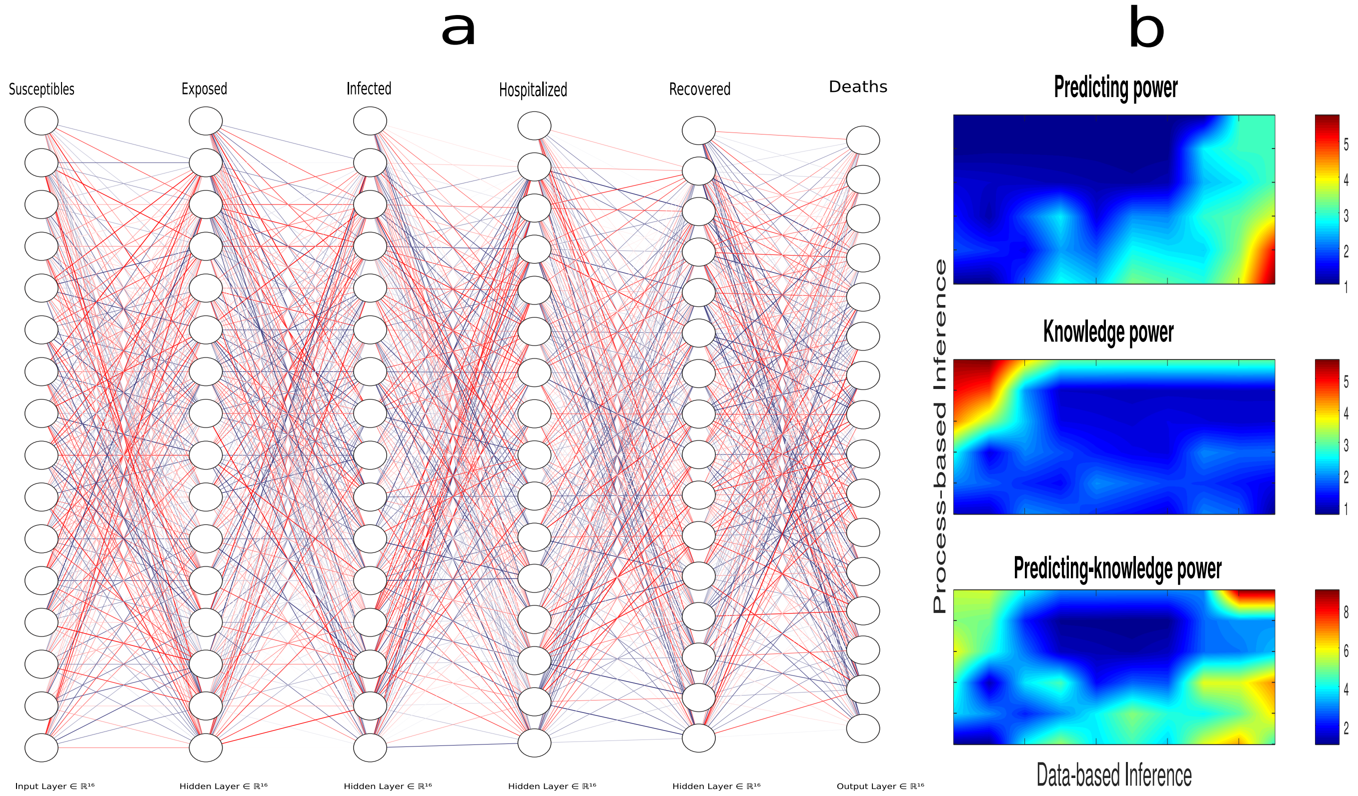

Figure 3: Causal-Knowledge Graphs. **a**) They contain features of deep learning networks and process-based modeling making empirical patterns interpretable. For epidemiology models, for example, all input, hidden and output layers (bottom) can be "converted" to age-class susceptibles (input), exposed, infected, hospitalized, and recovered (all hidden), and deaths (output). The same applies to edges connecting each pair of nodes. Causal-Knowledge Graph recovers the temporal sequence of events that best predict the output layer (i.e., Deaths). **b)** Causal-Knowledge Graphs integrates predicting and knowledge power into interpretable knowledge (bottom, b). x-axis represents "Data-based inference" (i.e., gradient of AI methods from low (left) to high (right) predicting power). y-axis represents "Process-based inference" (i.e., gradient of process-based methods from low (bottom left) to high (top left) knowledge power). The gradient of predicting power map (top) shows a hot spot red area in the bottom right highlighting the region where AI methods 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. Red and blue lines show important and non-important roles in the predicting-knowledge power map, respectively.

This is particularly relevant in the face of rapidly responding to global emergency and sustainability challenges. For example, in the context of the COVID-19 pandemic, and in many other situations challenging global sustainability, models and predictions abound, but no one can say with certainty what the course of the virus or the focus of the causing factors will be, and much less the impact the pandemic will have on people and societies [**?**]. In such an scenario, sharing fully reproducible contributions from a large number of nodes following standard protocols containing discovery-knowledge graphs can guide us to improve the robustness of the forecasting.

$\mathcal{ROBHOOT}$ v.3.0 deploys sharing protocols of discovery-knowledge graphs using reproducible-knowledge graphs, **D3.1, Sharing discovery-knowledge graphs in federated networks ($\mathcal{SDN}$)**. This deliverable explores scalability and security properties of sharing discovery-knowledge graphs using different topologies of federated networks: there can be data-, and causal-knowledge graphs belonging to specific categories and they might be integrated with different data- and causal-types to gain global understanding of data and interpretable patterns (i.e., like letting people on different social networks follow each other). Second,

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**	*QKG*	WP1	**Eguíluz**	OT	PU	27
v.1.0	**D1.3**	*DKG*	WP1	**Choirat**	OT	PU	27
v.1.0	**D1.4**	*CODA*	WP1	**Leads M1**	R,OT,DEC	PU	28
v.2.0	**D2.1**	*EAIA*	WP2	**Melián**	OT	PU	29
v.2.0	**D2.2**	*CKG*	WP2	**Baity/Vicente**	OT	PU	29
v.2.0	**D2.3**	*BSM*	WP2	**Guimerà**	OT	PU	29
v.2.0	**D2.4**	*COCAU*	WP2	**Leads M2**	R,OT,DEC	PU	30
v.3.0	**D3.1**	*SDN*	WP3	**von Waldow**	OT	PU	42
v.3.0	**D3.2**	*ADIN*	WP3	**Maass**	OT	PU	42
v.3.0	**D3.3**	*ACODIN*	WP3	**Leads M1-3**	R,OT,DEC	PU	42

Table 3.1c List of Deliverables: *ROBHOOT contains three main work packages: Milestone ROBHOOT* **v.1.0** *span from Month 3 to 27. Deliverable* **D1.4** (*CODA*)*, the global COVID-19 data-architecture depends on all the deliverables of work package one* (**WP1**)*. Milestone ROBHOOT* **v.2.0** *span from Month 5 to 29. Deliverable* **D2.4** (*COCAU*)*, interpretable causal-knowledge graph for COVID-19 depends on all the deliverables of work package two* (**WP2**)*, and Milestone ROBHOOT* **v.3.0** *span from Month 18 to 42. Deliverable* **D3.3 (ACODIN)***, automated cooperative forecasting of discovery-knowledge graph for the COVID-19 depends on all the deliverables of work package three* (**WP3**)*.*

ROBHOOT v.3.0 connects interpretable discovery-knowledge graphs to fedetated netwoks to study the properties of automated cooperative forecasting and strong inference in the face of global sustainability challenges (**D3.2, Automated discovery in federated networks (***ADIN***)**).

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

3.3 Consortium as a whole

- The individual members of the consortium are described in a separate section 4
- Describe the consortium. Explain how it will support achieving the project objectives. Does the consortium provide all the necessary expertise? Is the interdisciplinarity in the breakthrough idea reflected in the expertise of the consortium?
- In what way does each of the partners contribute to the project? Show that each has a valid role and adequate resources in the project to fulfil that role. How do the members complement one another? Other countries and international organisations: If one or more of the participants requesting EU funding is based in a country or is an international organisation that is not automatically eligible for such funding

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: *ROBHOOT contains three milestones: 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. 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. 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 ROBHOOT website.*

(entities from Member States of the EU, from Associated Countries and from one of the countries in the exhaustive list included in General Annex A of the work programme are automatically eligible for EU funding), explain why the participation of the entity in question is considered essential for carrying out the action on the grounds that participation by the applicant has clear benefits for the consortium.

ROBHOOT is a science-enabled multi-feature technology. ROBHOOT's consortium is designed with a highly modular topology to gain functionality within each milestone (Figure 4, blue, red, and pink). From the other side, connections among the modules reflect the emergence of interdisciplinarity technologies, the Discovery-Knowledge Graph, The Evolutionary Automation and the Cooperative Forecasting (Figure 4, green). ROBHOOT **v.1.0**'s team is composed by Fortuna, Eguíluz and Choirat to bring question- and data-knowledge graphs, the data-discovery process, to fully reproducible-knowledge graphs (section 3.1 and Figure 4). 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 ontology standards and automated API-Discovery (i.e., *APID* and *QKG*, *DKG*). Eguíluz's team focuses on network modularity, community detection and decentralization metrics, to characterize Question-, and Data-Knowledge Graphs (i.e., *QKG* and *DKG*, and Choirat's team encodes all the algorithms and procedures from Fortuna's and Eguíluz's teams into Reproducible-Knowledge Graphs. Milestone ROBHOOT **v.1.0** generates an automated COVID-19 Data-Knowledge Graph (Figure 4, blue).

ROBHOOT **v.2.0**'s team composed by Guimerà, Baity, Vicente, and Melián fussion Bayesian Machine Scientist to Evolutionary and AI Algorithms, the Evolutionary Automation (Figure 4, green), making interpretable data patterns along causal-knowledge graphs. The team for this milestone add complementarity expertise to 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 empirical patterns and evolutionary-inspired networks (section 3.2 and Figure 4, red). Despite modules ROBHOOT **v.1.0** and ROBHOOT **v.2.0** focus on specific milestones and deliverables (Figure 2 and 4), they have functional interactions because data- and causal-knowledge graphs, the discovery-knowledge graphs, will be fussioned using evolutionary automation built on a interdisciplinarity science-enabled technology that can be compactly converted into user-friendy open-software (**Discovery-Knowledge Graphs** and **Evolutionary Automation**, green). Milestone ROBHOOT **v.2.0** generates an automated COVID-19 Causal-Knowledge Graph (Figure 4, 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 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 *ROBHOOT* contain researchers from Estonia, Spain, Switzerland and Sweden (**Computer software design, To be confirmed**).

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-based automated and interpretable technology is not enough if we aim to contrast robustly interpretable scenarios in the face of global sustainability challenges. To achieve scalability for the discovery-knowledge graphs, sharing and automation in federated cooperation networks is the excellency feature of *ROBHOOT* **v.3.0** (section 3.3). *ROBHOOT* **v.3.0**'s team composed by von Waldow and Maass, develops protocols for sharing the discovery-knowledge graphs along federated networks. The team forming *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 build user-friendly open-access interfaces to explore scenarios of automated reproducible sharing in federated networks. Milestone *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 most plausible causal-knowledge graphs, the discovery-knowledge graphs produced in different nodes need to automatically interact and learn from each other to find better forecasting scenarios at a global scale. *ROBHOOT* v.3.0's implements the cooperation and automation among discovery-knowledge graphs in federated networks for making cooperative forecasting a standard global property. Milestone *ROBHOOT* **v.3.0** generates an Automated COVID-19 DKGs in Federated Cooperation Networks (Figure 4, pink). *ROBHOOT* v.3.0 contain researchers from Switzerland and Austria.

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)

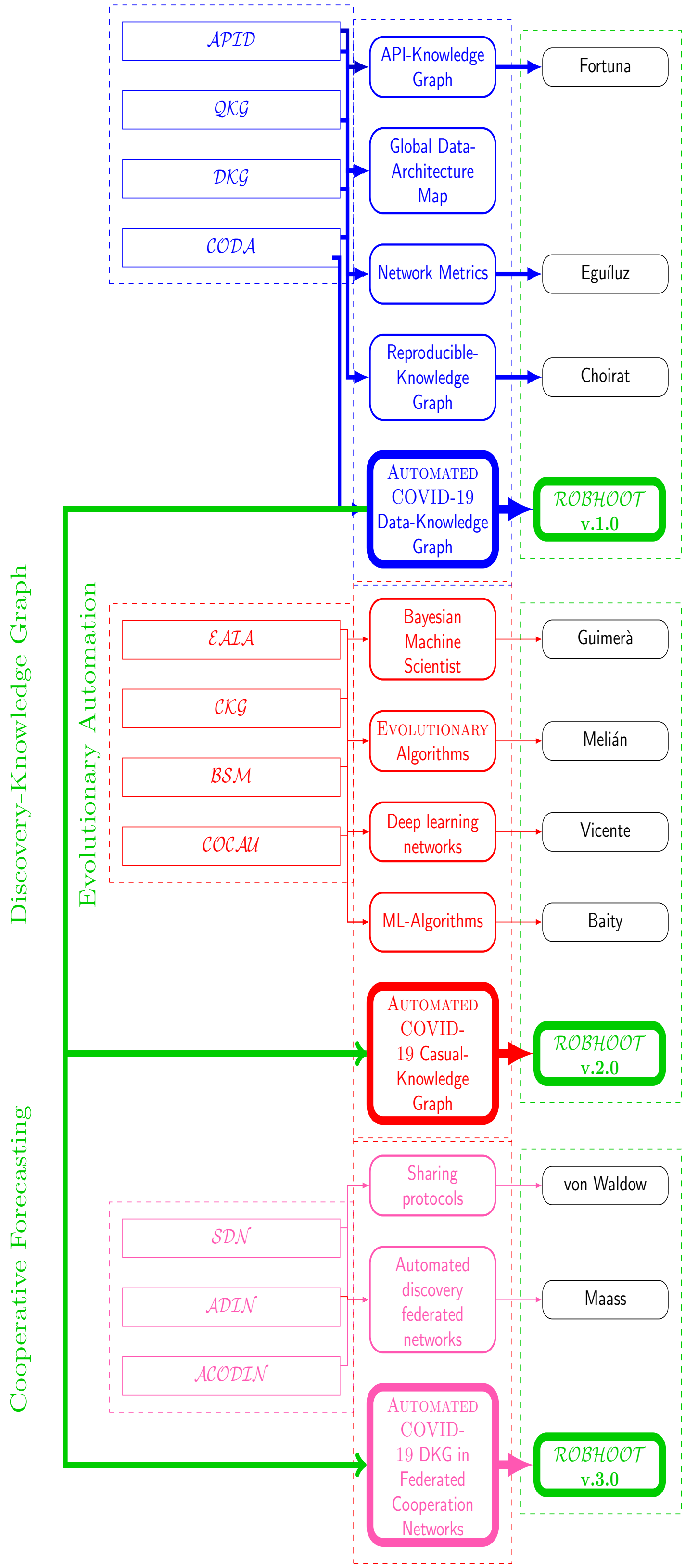

Figure 4: ROBHOOT Consortium: ROBHOOT **v.1.0** *(blue) to ROBHOOT* **v.3.0** *(pink) with acronyms of each deliverable (Left column), tasks (Center), and lead name (Right column). Links connect deliverables to tasks and leading groups. ROBHOOT delivers three interdisciplinarity-driven science-enabled technologies:* **Discovery-Knowledge Graph** *connecting ROBHOOT* **v.1.0** *and ROBHOOT* **v.2.0**. **Evolutionary automation** *in ROBHOOT* **v.2.0**, *and* **Cooperative forecasting** *connecting ROBHOOT* **v.2.0** *to* **v.3.0**

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] 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.

[2] 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.

[3] 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.

[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] 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.

[12] 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.

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

[14] 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.

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

[16] *Iris*. URL https://iris.ai.

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

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

[19] *Aito*. URL https://aito.ai/.

[20] 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.

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

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

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

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

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

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

[34] 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.

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

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

[37] 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.

[38] 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/.

[39] 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.

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