ROBHOOT
Automated Discovery-Knowledge Graphs in Federated Cooperation Networks
v.1.0

April 30, 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 in cooperative federate networks targeting robust and rapid discovery 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.
- 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 cooperation, automation, and reproducibility.
- 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 four 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. Yet, despite the rapid evolution of the digital ecosystem around data driven machines, discovery technologies facilitating global access to fully reproducible reports when solving complex governance, social, environmental and technological problems are particularly lacking (Figure 1 and Table 1) [1]. This lack of reproducible technologies is particularly relevant for targeting rapid information access in global emergency and sustainability challenges. How can automated and explainable (Table 1) data- and causal-knowledge graphs be integrated into discovery networks to generate predictive scenarios for global emerging challenges? How can discovery driven intelligent machines help to reach global sustainability goals? The *ROBHOOT* project integrates data- and causal-knowledge graphs, the discovery-knowledge graph, into intelligent networks to provide critical and accesible information to help humanity to take informed decisions in emergency and sustainability challenges (Figure 1 and Table 1).

More than half (i.e., 3.9 billion) of the global population is now online and using the Internet, which represents a more inclusive global information society. People are applying technology for good in powerful ways, from adopting decentralized technologies for humanitarian efforts, to improving agricultural practices and reducing waste in the global food supply chain [2]. Yet, current technological paradigm assisting humans for scientific inquiry is currently based on competitive schemes instead of intelligent global collaborative protocols in federated networks [3]. In addition, technologies for scientific inquiry are highly fragmented, partly solve reproducibility, are mostly developed in close-source software and contain many biases [4, 5, 6, 7, 8, 9, 10]. Thus, despite the importance of global access to discovery to close knowledge gaps for rapid information access in emergency and sustainability challenges, open-source technologies integrating discovery-knowledge graphs in federated, collaborative, intelligent networks are currently not in place.

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 four 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 (section 3.1). *ROBHOOT* **v.2.0** integrates automated and explainable biology-inspired neural-networks to decipher causal-knowledge graphs from data-knowledge graphs (section 3.2). *ROBHOOT* **v.3.0** explores decentralization and sharing data- and causal-knowledge graphs protocols for discovery in federated networks (section 3.3), and *ROBHOOT* **v.4.0** integrates automated discovery-knowledge graphs in federated cooperation networks.

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 multigraph	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.

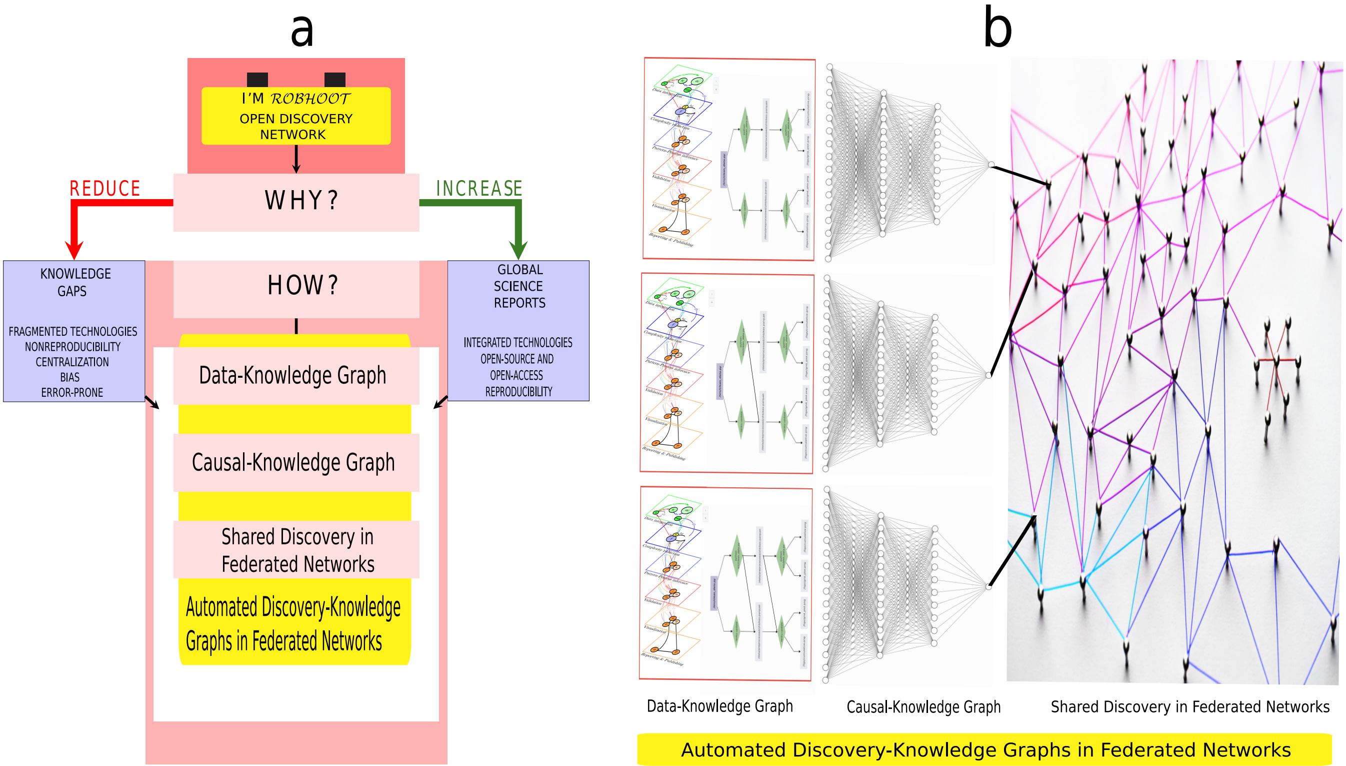

Figure 1: Cooperative Discovery in Federated Networks. *ROBHOOT integrates data- and causal-knowledge graphs, the discovery-knowledge graph, into cooperative federated networks for a sustainable knowledge-inspired society:* **a)** *ROBHOOT targets global knowledge gaps (red path) and open-access reproducible 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.*

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
- 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 particularly 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 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 still present 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.

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.

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).

Knowledge-inspired societies and governance will demand full research cycle transparency, reproducibility and interpretability to inform complex social, environmental and technological problems. The need of transparency, reproducibility and interpretability brings many technical and functional challenges to our research proposal because obtaining robust knowledge from integrating many parts each containing its own set of methods can generate divergent, fragile and contradictory outcomes. $\mathcal{ROBHOOT}$ will have a modular and flexible structure following four main versions each divided in four work packages and milestones (**TO BE CONTINUED**)

2 Impact

2.1 Expected impacts

2.2 Measures to maximise impact

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 $\mathcal{ROBHOOT}$ consortium aims at developing a federated network integrating several technologies into a unified framework. $\mathcal{ROBHOOT}$ 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. $\mathcal{ROBHOOT}$'s goals are developed in four different stages with four main milestones and sixteen deliverables (Figure 4).

3.1 *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, where new questions and scenarios are constantly emerging.
- Data- and Question-Knowledge graphs as solutions for rapid data-driven discovery (defined in Table 1). Which are their features? Which is the state-of-the-art? DKGs will explore similarity patterns of database to discover existing gaps in data availability and patterns. DKGs will complement QKGs to explore poorly explored questions to new pattern discovery.
- 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. Data access with different privacy requirements, formats, heterogeneity, dimensions, bias and spatiotemporal resolution is the norm and many projects are rapidly emerging to facilitate rapid data access [21, 22, 23, 24]. Yet, available automated science-enabled technologies to build data- and questions-knowledge graphs at the global scale to rapidly inform causal-knowledge graphs about the type of interpretable information that can be extracted, while key to make predictive scenarios, still remain at a very early stage of development despite standard protocols to automate data API access, knowledge extraction, and ETFs algorithms are rapidly advancing [25, 26, 27]. The technologies around automated API data-discovery and interpretable question- and data-knowledge graphs remain difficult to compactly link to the automated causal-knowledge graphs for scientific discovery. Milestone one, *ROBHOOT* v.1.0 deploys the following four deliverables: data discovery technology (**D1.1: API discovery (APID)**) aiming to build question (**D1.2: Question-Knowledge graphs (QKGs)**), and data-knowledge graphs (**D1.3: Data-Knowledge graphs (DKGs)**), and a case study merging question- and data-knowledge graphs to gain information about global data-architecture (**D1.4: Global COVID-19 data-architecture (CODA)**).

3.2 *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 biologically inspired neural networks and evolutionary biology techniques [30, 31]. Automatic interpretation of the causal processes underlying empirical patterns will be explored in a series of neuromorphic computing scenarios using neural networks of evolving spiking neurons (**D2.1: Causal Spiking Networks (CSN)**) (Figure 2).

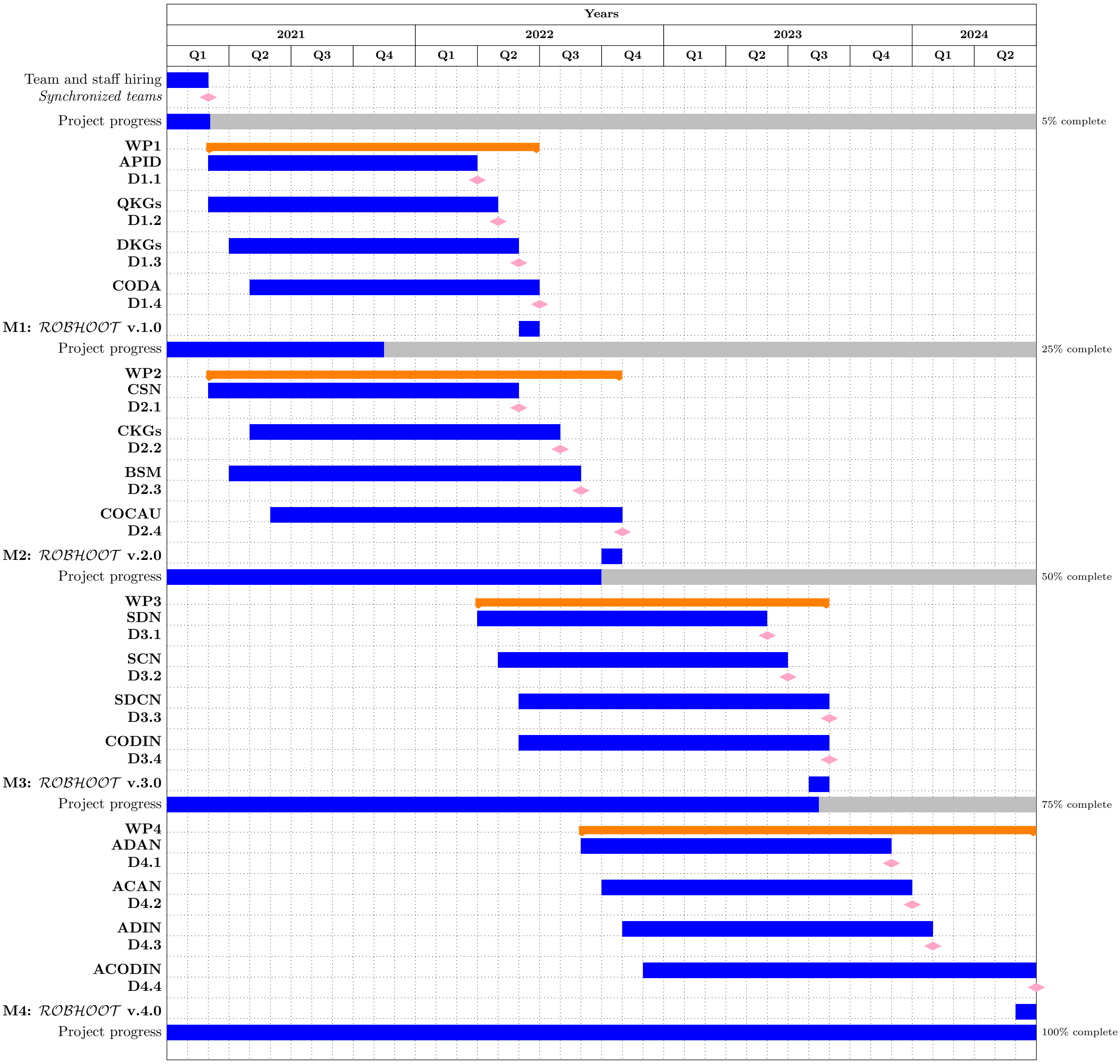

Figure 2: Roadmap: Work package one, **WP1**, introduces Milestone one, *ROBHOOT* **v.1.0** and deliverables one to four (**D1.1** to **D1.4**) to bring question- and data-knowledge graphs into a global data-architecture map. Work package two, **WP2**, introduces Milestone two, *ROBHOOT* **v.2.0** and 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 three, *ROBHOOT* **v.3.0** and deliverables nine to twelve (**D3.1** to **D3.4**) to connect shared discovery-knowledge graphs to federated networks. Work package four, **WP4**, introduces Milestone four, *ROBHOOT* **v.4.0** and deliverables thirteen to sixteen (**D4.1** to **D4.4**) to bring automated discovery-knowledge graphs in federated networks to help closing knowledge-gaps in rapidly emerging global sustainability challenges.

Causal Spiking Neurons 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 automated biologically inspired solutions to complex empirical patterns (**D2.2: Causal-Knowledge Graphs (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 Causal Spiking 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 (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]. *ROBHOOT* v.2.0 deploys 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 (COCAU)**).

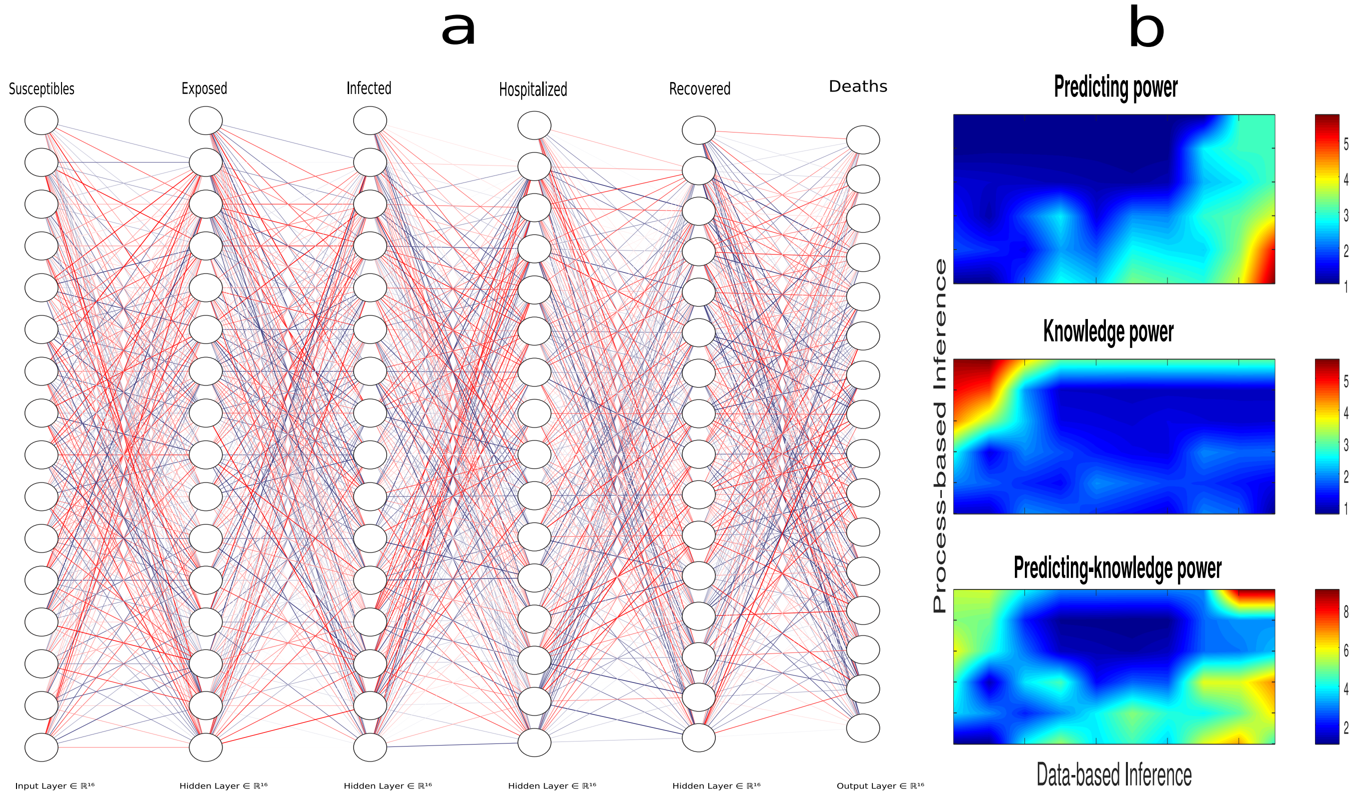

Figure 3: Causal-Knowledge Graphs. **a**) They contain features of deep learning networks and process-based modeling making empirical patterns interpretable. For epidemiology models, for example, all input, hidden and output layers (bottom) can be "converted" to age-class susceptibles (input), exposed, 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.

3.3 *ROBHOOT* v.3.0: Shared Discovery 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 aiming to find robust populations of results in face of rapidly emerging global emergency and sustainability challenges.

A science-based explainable technology is not enough if we aim to globally contrast robustly paths of interpretable causal processes predicting the global data-architecture obtained from the data-and causal-knowledge graphs. In this regard, sharing data- and causal-knowledge graphs are needed to develop scalable cooperative forcasting and strong inference. *ROBHOOT* v.3.0 focus develops protocols in digital networks to embed explainable discovery-knowledge graphs into global cooperation schemes to increase robustness and reproducibility for rapid report generation [3]. Technologies in decentralized digital ecosystems are rapidly advancing in a variety of sectors. Most progress is coming in the scalability, security and decentralization trade-offs [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 networks to facilitate cooperative forecasting is currently not in place.

ROBHOOT v.3.0 deploys the following four deliverables for sharing discovery in networks: sharing question- and data-knowledge graphs, the data-architecture, in federated networks using reproducible-knowledge graphs. **D3.1 to D3.3, Sharing data-architecture (SDN), causal-knowledge graphs (SCN), and discovery-knowledge graphs in federated networks (SDCN), respectively**: They explore scalability and security properties of sharing knowledge graphs using different topologies of federated networks. For example, there can be data-knowledge graphs belonging to specific categories of data but they might be integrated with different data types to gain global understanding of data-architecture (i.e., like letting people on different social networks follow each other).

ROBHOOT v.3.0 includes, as a case study, the sharing of the discovery-knowledge graphs in a federated network. Despite the 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), with these developments inextricably bound up in modern socio-technical developments and processes of globalization, science-based technologies facilitating rapid sharing of information to mitigate risks and enhance response efficiency with globally informed scenarios are particularly lacking [2]. The last deliverable of *ROBHOOT* v.3.0, (**D3.4: Sharing discovery-knowledge graphs in federated networks, COVID-19 as a case study (CODIN)**), deploys sharing data-architecture and causal-knowledge graphs, the discovery-knowledge graphs, into federated networks for the COVID-19 case study.

3.4 *ROBHOOT* v.4.0: Automated Discovery-Knowledge Graphs in Federated Networks

- 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 particularly relevant in the face of rapidly responding to global emergency and sustainability challenges.
- A science-based technology to develop automatically shared cooperative forecasting aiming to find robust populations of results in the 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 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, models and predictions abound, but no one can say with certainty what the course of the virus will be, and much less the impact the pandemic will have on people and societies [40]. In such an scenario, making shared contributions fully reproducible in a tandard protocol containing a mixture of human and automatically generated discovery-knowledge graphs can guide us to improve the robustness of the forecasting. In this regard, shared data- and causal-knowledge graphs, the discovery-knowledge graphs, containing a mixture of human and automatically generated discovery-knowledge graphs, from where cooperative forecasting can estimate robust populations, might help to rapidly generate global reports of future scenarios in real-time.

ROBHOOT v.4.0 connects interpretable discovery-knowledge graphs to automation to study the properties of scalable cooperative forcasting and strong inference in the face of global sustainability challenges. *ROBHOOT* v.4.0 deploys the following four deliverables for automated discovery in federated networks: **D4.1 to D4.3, Automated data- (ADAN), causal- (ACAN), and discovery-knowledge graphs (ADIN) in federated networks, respectively** deploy and share data-, causal- and discovery-knowledge graphs from automatically and cooperatively generated knowledge graphs. The last deliverable of *ROBHOOT* v.4.0, **D4.4: Automated Cooperative Forecasting of Discovery-Knowledge Graphs for the COVID-19 case study (ACODIN)**, deploys and share automatically generated discovery-knowledge graphs in federated networks for the COVID-19 case study to contrast cooperative forecasting under poulations of scenarios.

ROBHOOT v.X.X	Deliver. number	Deliver. name	WP number	Name Lead	Type	Disem.	Delivery date
v.1.0	**D1.1**	APID	WP1	NAME LEAD	OT	PU	15
v.1.0	**D1.2**	QKGs	WP1	NAME LEAD	OT	PU	16
v.1.0	**D1.3**	DKGs	WP1	NAME LEAD	OT	PU	17
v.1.0	**D1.4**	CODA	WP1	NAME LEAD	R,OT,DEC	PU	18
v.2.0	**D2.1**	CSN	WP2	NAME LEAD	OT	PU	17
v.2.0	**D2.2**	CKGs	WP2	NAME LEAD	OT	PU	19
v.2.0	**D2.3**	BSM	WP2	NAME LEAD	OT	PU	20
v.2.0	**D2.4**	COCAU	WP2	NAME LEAD	R,OT,DEC	PU	22
v.3.0	**D3.1**	SDN	WP3	NAME LEAD	OT	PU	29
v.3.0	**D3.2**	SCN	WP3	NAME LEAD	OT	PU	30
v.3.0	**D3.3**	SDCN	WP3	NAME LEAD	OT	PU	32
v.3.0	**D3.4**	CODIN	WP3	NAME LEAD	R,OT,DEC	PU	32
v.4.0	**D4.1**	ADAN	WP4	NAME LEAD	OT	PU	35
v.4.0	**D4.2**	ACAN	WP4	NAME LEAD	OT	PU	36
v.4.0	**D4.3**	ADIN	WP4	NAME LEAD	OT	PU	37
v.4.0	**D4.4**	ACODIN	WP4	NAME LEAD	R,OT,DEC	PU	42

Table 2: List of Deliverables: *ROBHOOT* contains four main work packages each containing four deliverables: Milestone *ROBHOOT* **v.1.0** span from Month 3 to 18. 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 3 to 22. Deliverable **D2.4 (COCAU)**, interpretable causal-knowledge graphs at global scale for COVID-19 depends on all the deliverables of work package two (**WP2**). Milestone *ROBHOOT* **v.3.0** span from Month 15 to 32. Deliverable **D3.4 (CODIN)**, shared discovery-knowledge graphs for the COVID-19 depends on all the deliverables of work package three (**WP3**). Milestone *ROBHOOT* **v.4.0** span from Month 21 to 42. Deliverable **D4.4 (ACODIN)**, automated cooperative forecasting of discovery-knowledge graph for the COVID-19 depends on all the deliverables of work package four (**WP4**).

3.5 Research methodology and work plan – Work packages, deliverables

3.6 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.

Milestone number	Milestone name	Related work package(s)	Due data (in months)	Verification
M1	*ROBHOOT* **v.1.0**	WP1	Months 3-18	OS-Software,Paper/Conf.
M2	*ROBHOOT* **v.2.0**	WP2	Months 3-22	OS-Software,Paper/Conf.,demo-website
M3	*ROBHOOT* **v.3.0**	WP3	Months 15-32	OS-Software,Paper/Conf.,main-website
M4	*ROBHOOT* **v.4.0**	WP4	Months 21-42	OS-Software,Paper/Conf.,official website

Table 3: List of Milestones: *ROBHOOT* contains four milestones: *ROBHOOT* **v.1.0** span from Month 3 to 18 to generate Open-source Software and research paper and/or conference. *ROBHOOT* **v.2.0** span from Month 3 to 22 to generate Open-source Software, research paper and/or conference and public demo-website. *ROBHOOT* **v.3.0** span from Month 15 to 32 to generate Open-source Software, research paper and/or conference and main-website. *ROBHOOT* **v.4.0** span from Month 21 to 42 to generate Open-source Software, research paper and/or conference, and official *ROBHOOT* website.

3.7 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 (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 hybrid-technology. Multi-feature and hybrid imply, from one side, a highly modular network topology (Figure 4). From the other side, there must be connections among the modules to make the multiple-trait and hybrid features a science-enabled functional technology. Each milestone, *ROBHOOT* **v.1.0** to *ROBHOOT* **v.4.0**, have more interactions within than between the milestones. Yet, each milestone connects to the Reproducible-Knowledge Graphs of *ROBHOOT* **v.1.0** to give functionality to the technology as a whole. *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-reproducible knowledge graphs encoding the global data-architecture map (section 3.1 and Figure 4). The first milestone requires a mixture of researchers, from computer-, data-scientists and developers to researchers working in complex networks from the quantitative and epistemological angles. *ROBHOOT* **v.2.0**'s team composed by Guimerà, Baity, Vicente, LEAD, Melián and LEAD fussion automation and interpretable data along causal-knowledge graphs. The team for this milestone add complementarity expertise to the global data-architecture picture: Now the skills focus on data-scientists trained in deep learning networks and automation algorithms, theoreticians with expertise in Bayesian inference, and evolutionary biologists and neurobiologists with expertise in empirical patterns and biology-inspired neural modeling networks (section 3.2 and Figure 4). Despite modules from *ROBHOOT* **v.1.0** and *ROBHOOT* **v.2.0** focus on specific milestones and deliverables (Figure 2 and 4), they keep functional interactions because the data- and causal-knowledge graphs will be fussioned as the discovery-knowledge graphs, to build an interdisciplinarity-hybrid science-enabled technology that can be compactly converted into user-friendy open-software (**To be better captured in Figure 4:: Highlight Discovery-Knowledge Graphs**). Therefore, interdisciplinarity enters not only at the intra-module development stage, but also at the inter-module stage where functionality driven discovery-knowledge graphs can trigger a breakthrough reflected in the highly transdisciplinar and complementarity skills of the consortium (section 4.1). The first two modules in *ROBHOOT* contain researchers from four countries, Estonia, Spain, Switzerland and Sweden (**To be confirmed**), with expertise in Deep Learning networks and AI, Complex networks (quantitative metrics, epistemology and modeling), Theory with Bayesian inference expertise and Modeling in Evolutionary biology, neurobiology and Computer software design, respectively.

The *ROBHOOT* consortium wants to understand discovery broadly accounting for cooperative networks. We want 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 globally contrast robustly interpretable scenarios in the face of global sustainability challenges. To achieve scalability for the discovery-knowledge graphs, sharing is the excellency feature of *ROBHOOT* **v.3.0** (section 3.3). *ROBHOOT* **v.3.0**'s team composed by LEAD, von Waldow, and LEAD develops protocols for sharing data- and causal-knowledge graphs, the discovery-knowledge graphs, and making them reproducible in the context of cooperative federated networks. The team forming *ROBHOOT* **v.3.0** therefore require quite a lot of contrasting skills. First, developers working in P2P and security protocols. Second, social scientists in collaboration to developers aiming to build user-friendly open-access interfaces to explore alternative scenarios of sharing. Third, data-scientists working in reproducible-knowledge graphs to make the discovery-knowledge graphs "reproducible shared objects" (i.e., connecting all the milestones, Figure 4). In summary, a strong complementarity but also a functionality driven cooperation between (*ROBHOOT* **v.1.0, v.2.0 and v.3.0** is required to make the discovery-knowledge graphs reproducible and functionally accessible. *ROBHOOT* v.4.0's team composed by LEAD, LEAD, and LEAD implement the cooperation and automation among discovery-knowledge graphs in federated networks for making cooperative forecasting a standard global property. The last two milestones contain researchers from XX countries, XX, XX, XX, Switzerland and XX, with expertise in protocols in decentralized algorithms, security in computing networks and distributed networks, cooperation and federated networks, and social networks.

3.8 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)

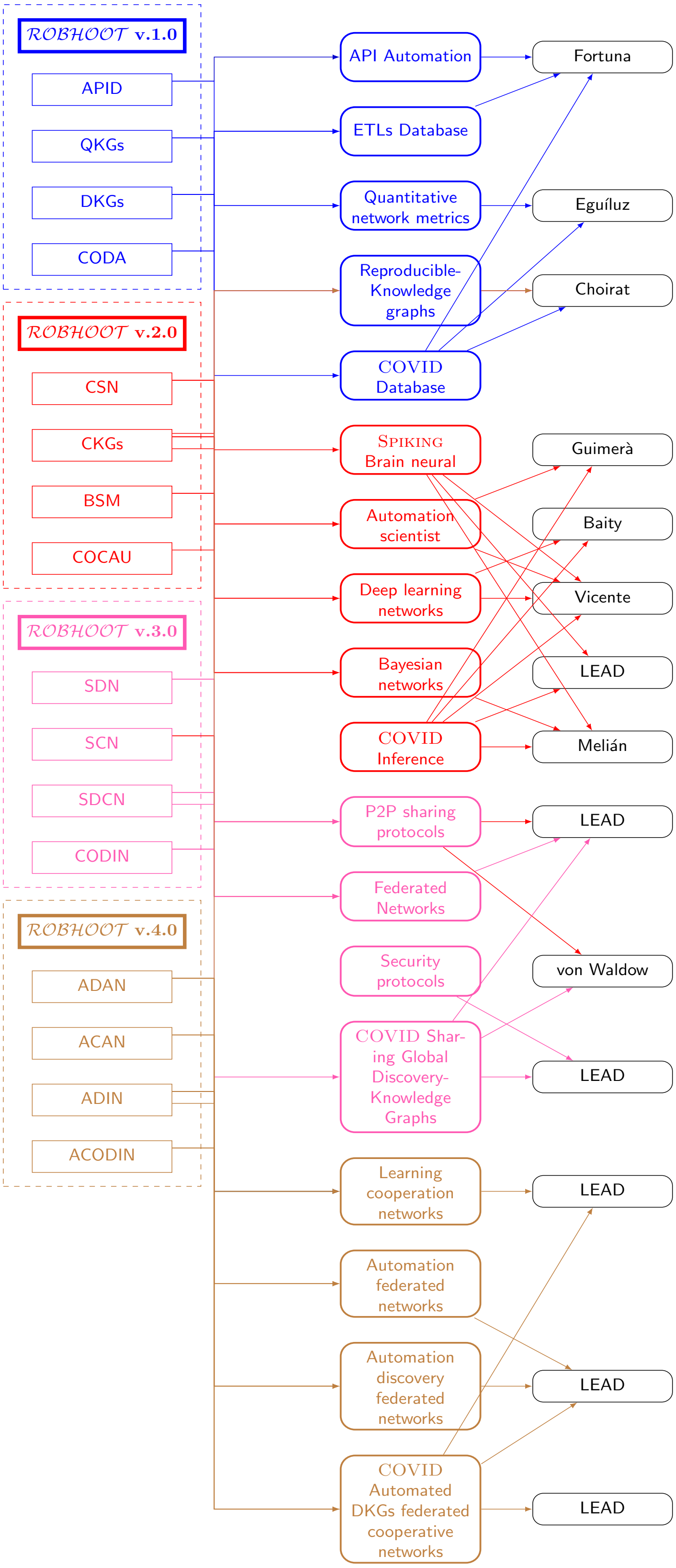

Figure 4: ROBHOOT Team highlighting the four main Milestones, ROBHOOT **v.1.0** *(blue) to ROBHOOT* **v.4.0** *(brown) with acronyms of each deliverable (Left column), main tasks (Center), and Name lead (Right column). Links connect deliverable to tasks and leading groups. The Milestones are modular with the Reproducible-Knowledge Graphs in ROBHOOT* **v.1.0** *and the Discovery-Knowledge Graph in ROBHOOT* **v.2.0** *connecting all Milestones.*

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
- redList 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.1

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] Remarkable times call for remarkable collaboration . (April), 2020.