Robhoot
Knowledge Open Network

January 17, 2020

1 Summary

The Robhoot project aims to connect open-access research and the public in a decentralized network to help taking informed decisions when solving complex social, environmental and technological problems. Current technologies for scientific inquiry and decision-making are highly fragmented and thus only increase robustness, reproducibility, open-access and the interactions with the public marginally. The goal of Robhoot is to propose a hybrid-neutral-technology to lay out the foundation for an open-science research ecosystem aiming to strengthen the robustness and reproducibility of science by fully automating standard reporting in decision-making. Robhoot is not set out to deliver a finished deep knowledge network in the science ecosystem, but to provide a science-enabled technology in establishing a prototype proof-of-principle to connect decentralized and neutral-knowledge generation with knowledge-inspired societies.

2 The Science Ecosystem

The process of building science and technology requires multiple steps of information transfer among trusted/untrusted peers to build solid evidence-based knowledge. Evidence-based knowledge forms the basis of decisions across all human activities. In this regard, reliable, open-access, neutral, and immutable evidence-based knowledge following a secure peer-to-peer architecture storing the open-source knowledge graphs derived from distributed research outputs is far from reality. For example, reproducibility, decentralization, and immutability of knowledge is key to have fully neutral open-access reports when taking informed decisions in complex societal, environmental and technological problems. However, currently public funded science is highly centralized [1**?**], prone to errors [2], difficult to reproduce [3], and contains many biases [4]. Actually, these elements make the connection between the scientific process, open-access and reproducible reporting for decision-making highly improbable. Despite many projects are aiming at making the science ecosystem less centralized and biased while increasing openness and reproducibility a science-enabled technological paradigm connecting open-science to knowledge-inspired societies is not currently in place [5].

Many studies in decentralized ecosystems are producing an immense gain in detailed knowledge about scalability, security and decentralization trade-offs [6, 7, 8, 9, 10]. Automation and AI technologies is the other angle from which many advances are rapidly occurring [11, 12, 13]. Yet, while the existing technological paradigm is rapidly shifting towards science-based decentralization and automation technologies, end-to-end open-source research accounting for decentralized, neutral and automated knowledge-inspired technologies are missing. Rapid advances of automated research platforms facilitating data integration accounting for sections of the research cycle are currently under development[1] but open-source decentralized automated platforms accounting for the research cycle are still at a very incipient stage of development. While conceptual frameworks conceptualizing the required layers in many research fields are well established (Figure 1a), there is currently a lack of integration and development of tools connecting knowledge graphs (Figure 1b) into deep process-based learning networks to explore their robustness (Figure 1c) in fully decentralized ecosystems (Figure 1d).

Features	Science Ecosystem	Robhoot 1.0
Decentralization	No	Yes
Open-access	Mostly No	Yes
Immutability	No	Yes
Robustness	Mostly No	Yes
Reproducibility	Mostly No	Yes
Owner-Controlled assets	No	Yes

Table 1: *Robhoot aims to be designed to resolve desirable properties of science: Robustness, Reproducibility, Decentralization, Open and Direct access to reporting by peers and not-peers.*

3 Robhoot Design Goals

Robhoot aims to build an automated knowledge network technology to connect knowledge-graphs to knowledge-inspired societies (Figure 1). Robhoot will be built on different stages following standard version protocols. The most advanced version is to provide real-time open-access

[1]This is by no means an exhaustive list but it gives an indication of the many projects currently in place: NakamotoT,BigQuery,Automated statistician,Modulos,Google AI,Iris,easeml

and decentralized neutral data-rule-knowledge to gain informed decisions when solving complex social, environmental and technological problems. Figures 1 and 2 show Robhoot stages and the timeline of each stage, respectively.

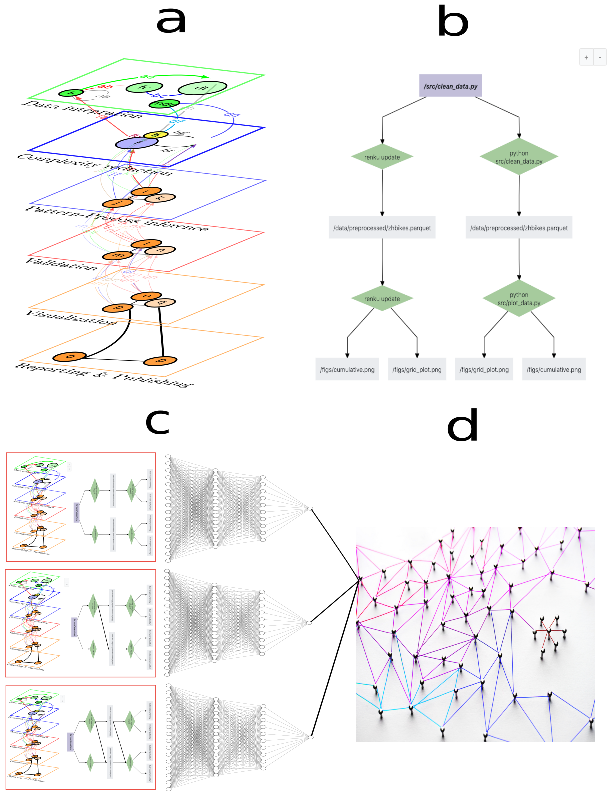

Figure 1: Automated knowledge-based network technology. **a**) **Robhoot 1.0** aims to account for a end-to-end research cycle from data integration (top) to reporting generation (bottom). **b**) **Robhoot 2.0** will integrate a end-to-end research cycle with a knowledge graph (KG) represented as one research path of **a** (i.e., Renku open-source code). **c**) **Robhoot 3.0** will account for deep knowledge-based networks to automatically explore populations of KGs to gain robustness of the process-based patterns contained in the data. **d**) **Robhoot 4.0** will deploy all KGs in a distributed network of mutually trusting/untrusting peers with every peer maintaining the population of the KGs (i.e., decentralized P2P git network like Gitchain.)

The overall objectives for each of the four major Robhoot versions are the following:

3.1 Robhoot 1.0: Automated end-to-end Research Cycle

- Develop, deploy and integrate open-source algorithms to automate an end-to-end research cycle (Figure 1a).
- Robustness within and between layers from data integration, complexity reduction, inference, and validation to visualization and automated reporting (Figure 1a).
- Robhoot 1.0 Open Network in Biodiversity and Global Change Research to connect open science (i.e., citizen and other models) to real-time open-access data-rule-knowledge to gain informed decisions when solving local and global environmental problems.

3.2 Robhoot 2.0: Knowledge Graphs

- Integration between the automated end-to-end research cycle with Knowledge Graphs (KGs) (Figure 1b).
- Robustness and stability of a suite of open-source lineage client-tracker algorithms.

3.3 Robhoot 3.0: Deep learning networks

- Deploy automated deep learning algorithms accounting for a end-to-end research cycle in a lineage client-tracker to produce populations of Knowledge Graphs (KGs) (Figures 1a-c).
- Robustness of the multi-objective optimization by exploring a suite of neural networks algorithms with lineage client-tracker paths in the multilayer network (Figure 1c).

3.4 Robhoot 4.0: Distributed ledger network

- Deploy an end-to-end permissioned-permissionless distributed ledger technology to guarantee decentralization, open-access and security of the KGs populations in the science ecosystem (Figures 1c and 1d.)
- Distributed ledger implementation accounting for consensus algorithms and smart contracts among trusted-untrusted peer-to-peer interactions (Figure 1d).
- Exploring consensus algorithms to minimize scalability-security-decentralization trade-offs when storing the KGs in the knowledge network (Figure 1d).
- Testnet for the interaction between consensus protocols and the scalability-security-decentralization trade-offs when committing the KGs to the distributed ledger.
- Mainnet to cryptographically link each population of KGs to previous KGs-ledger to create an historical KGs-ledger chain that goes back to the genesis ledger. The mainnet aims to connect database real-time open-access citizen data science to knowledge-inspired societies.

4 The multilayer nature of Robhoot

The science ecosystem currently lack technologies automating the full research cycle from end-to-end and its connection to the scalability-security-decentralization trade-offs of digital ecosystems. Despite, public institutions are demanding more reproducibility and openness of the data and the scientific process, and overall a shifting towards open and reproducible scientific and engineering landscapes, there are not currently open and integrated technologies aiming to compactly facilitate and distribute the scientific and engineering knowledge in reproducible and immutable knowledge networks.

Knowledge generation usually comes from interactions within- and between-layers (Figure 1a). The feedbacks occurring among layers in the science and technology ecosystem also provide unexpected behaviors that are difficult to anticipate. Therefore many feedbacks and interactions within- and between-layers are not easy to reproduce if not properly accounted for. Robhoot will take advantage of the multilayer networks framework to explore how knowledge graphs and peer-to-peer interactions might connect to the robustness and reproducibility of the scientific process. The following are the six components required to complete Robhoot 1.0:

4.1 Data integration

Despite open-source ETLs are rapidly evolving towards accounting for many aspects of data integration (formats, historical-real time, storage, dimensions, size, bias and spatiotemporal resolution), there is a missing component in quantifying the robustness of knowledge that integrated data can provide. Automated populations of KGs connecting cutting-edge open-source ETLs to inference classification schemes can provide the quantification of robustness in knowledge-based patters for future predictive technologies.

4.2 Complexity reduction

4.3 Inference

The integration between open-source data integration and inference schemes, the interlayer automation (O1: Multilayer), will allow for the systematic exploration of robust knowledge-based patterns when exploring the population of KGs. This is in sharp contrast to existing AI technologies mostly oriented to prediction without knowledge-based understanding.

4.4 Validation

4.5 Visualization

4.6 Reporting

5 Robhoot in Digital Ecosystems

Technologies with the capacity to compactly account for neutral, borderless, immutable, and open-access information in hybrid, trusted-untrusted peer-to-peer interactions, accounting for the multilayer nature of science and engineering are currently not in place. Producing such technologies will require interactions with the existing and rapidly evolving digital ecosystem in addition to integrating expertise from disparate disciplines like multilayer networks, deep learning, automation algorithmics, and distributed technologies. The integration of these disciplines will require to go beyond domain boundaries. Specifically, we will merge scientists and engineers from data and computer science, the physics of complex systems, artificial intelligence and the biology, ecology and evolution of social, natural and technological ecosystems to develop a hybrid technology: automated knowledge generation in a neutral, borderless and immutable network synthesized anew from existing open-source projects like Renku, Fabric and gitchain. The following are the three main domains from where Robhoot will benefit the most:

5.1 Computing Power

5.2 Decentralization

5.3 Neural Networks

6 How to contribute to Robhoot

7 The Robhoot roadmap

8 Conclusion

Science and technology ecosystems are in need of accounting for the uncertainties, the reproducibility and the immutability related to automation. This need is not just for a specific stage of the research cycle, but for the full research cycle, from data acquisition to automated reporting generation because knowledge-inspired societies and governance will demand full research cycle transparency to solve complex social, environmental and technological problems. This need brings many challenges to our research proposal because obtaining robust knowledge from integrating many layers of the research cycle, each containing its own set of methods and uncertainties, can generate divergent, fragile and contradictory outcomes.

We will develop a flexible research method focusing step by step in different levels of complexity (i.e., from Robhoot 1.0 to 4.0). Our motivation will be to provide a first open-access proof of concept of how the technology works: we will automate reproducible research paths along a multilayer network (Robhoot 1.0) to sample the KGs (Robhoot 2.0) using different deep learning algorithms to estimate the uncertainty of the ruled-based inference obtained by fitting predictions to simulated data (Robhoot 3.0). Accounting for the uncertainties of each of the research stages when sampling the KGs comes from the many distinct paths within and across the layers in the research cycle (Figure 1). We will test a variety of consensus algorithms to explore the degree of security, decentralization and scalability of the ledger knowledge network using the generated population of KGs (Robhoot 4.0).

Despite our focus will be bias towards the algorithmic robustness during the four stages of Robhoot development, we will develop a domain-specific case study, a Robhoot Open Network, to test the robustness of the rule-based inference obtained by fitting each of the generated KG to empirical patterns. The high risk associated to robustly automate the full research cycle for producing immutable open knowledge will be buffered to a great extend because the existing digital ecosystem of tested and reliable open-source tools: We will combine our own algorithms (i.e., add-hoc data integration algorithms and neural-biological networks for deep learning algorithms for sampling and automating the KGs) with open-source tools part of a rapidly evolving open-digital ecosystem.

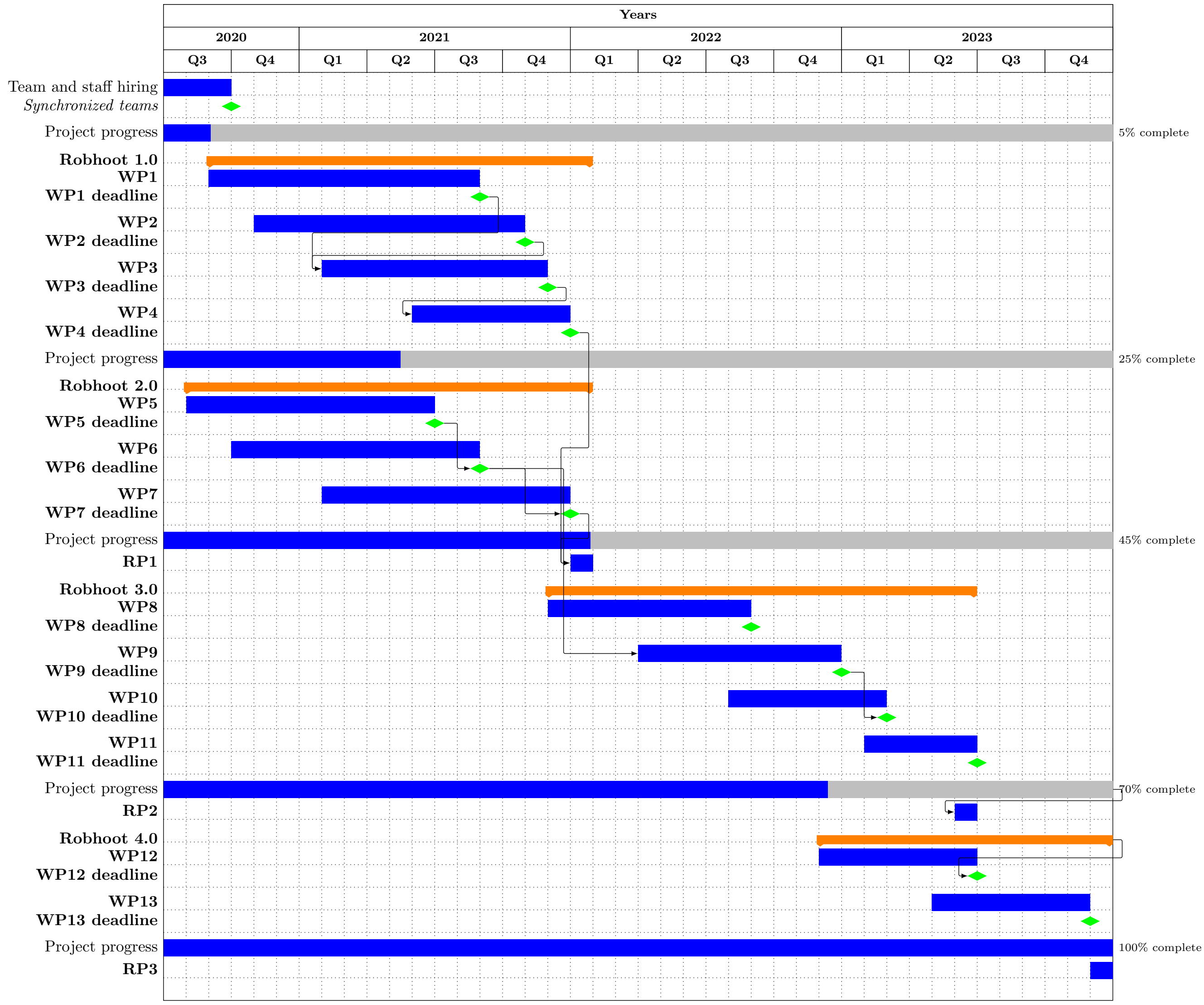

Figure 2: The Robhoot roadmap: a) **Robhoot 1.0** contains four working packages .... b) **Robhoot 2.0** contains three packages to integrate a end-to-end research cycle with a knowledge graph (KG). **c) Robhoot 3.0** contains four packages ... deep knowledge-based networks to automatically explore populations of KGs to gain robustness of the process-based patterns contained in the data. **d) Robhoot 4.0** contains two packages to deploy all KGs in a distributed network of mutually trusting/untrusting peers with every peer maintaining the population of the KGs (i.e., decentralized P2P git network like Gitchain.)

References

[1] 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`.

[2] 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`.

[3] 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`.

[4] 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`.

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

[6] Golem. The Golem Project Crowdfunding Whitepa-

per. *Golem.Network*, (November):1–28, 2016. URL https://golem.network/crowdfunding/Golemwhitepaper.pdf.

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

[8] 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.

[9] 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/.

[10] 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.

[11] J Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[12] 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.

[13] Yolanda Gil, Bart Selman, Marie Desjardins, Ken Forbus, Kathy Mckeown, Dan Weld, Tom Dietterich, Fei Fei Li, Liz Bradley, Daniel Lopresti, Nina Mishra, David Parkes, and Ann Schwartz Drobnis. A 20-Year Community Roadmap for Artificial Intelligence Research in the US Roadmap Co-chairs: Workshop Chairs: Steering Committee. Technical report, 2019. URL https://bit.ly/2ZNVBVb.