Proposal template: technical annex
Research and Innovation actions
Call: H2020-FETOPEN-2018-2020
Topic: FETOPEN-01-2018-2019-2020: FET-Open Challenging Current Thinking

The structure of this template must be followed when preparing your proposal. It has been designed to ensure that the important aspects of your planned work are presented in a way that will enable the experts to make an effective assessment against the evaluation criteria. Sections 1, 2 and 3 each correspond to an evaluation criterion.

Please be aware that proposals will be evaluated as they were submitted, rather than on their potential if certain changes were to be made. This means that only proposals that successfully address all the required aspects will have a chance of being funded. There will be no possibility for significant changes to content, budget and consortium composition during grant preparation.

A proposal that, according to the evaluators' assessments, does not convincingly satisfy all FET gatekeepers as described in the FET Work Programme will be declared out of scope.

Page limit: Sections 1 to 3 of the proposal should consist of a maximum of 15 A4 pages (no cover page). All tables, figures, references and any other element pertaining to these sections must be included as an integral part of these sections and are thus counted against this page limit. There is no page limit for sections 4 and 5.

The page limit will be applied automatically; therefore you must remove the first 2 instruction pages of this template before submitting.

If you attempt to upload a proposal longer than the specified limit before the deadline, you will receive an automatic warning and will be advised to shorten and re-upload the proposal. After the deadline, excess pages (in over-long proposals/applications) will be automatically made invisible, and will not be taken into consideration by the experts. The proposal is a self-contained document. Experts will be instructed to ignore hyperlinks to information that is specifically designed to expand the proposal, thus circumventing the page limit.

The following formatting conditions apply.

Each page should include a footnote with the acronym of the proposal.

The reference font for the body text of H2020 proposals is Times New Roman (Windows platforms), Times/Times New Roman (Apple platforms) or Nimbus Roman No. 9 L (Linux distributions).

The use of a different font for the body text is not advised and is subject to the cumulative conditions that the font is legible and that its use does not significantly shorten the representation of the proposal in number of pages compared to using the reference font (for example with a view to bypass the page limit).

The minimum font size allowed is 11 points. Standard character spacing and a minimum of single line spacing is to be used.

Text elements other than the body text, such as headers, foot/end notes, captions, formula's, may deviate, but must be legible.

The page size is A4, and all margins (top, bottom, left, right) should be at least 15 mm (not including any footers or headers).

Deep Knowledge Ledger Network

Deepklen :: Robhoot

1. Excellence

1.1 Radical vision of a science-enabled technology

Public funded science is highly centralized (Günther, 2018; Inhaber, 1977), prone to errors ((Fang & Casadevall, 2011)), difficult to reproduce (Hardwicke. et al., 2019)), and contains many biases (Ioannidis, 2005). Yet, despite attempts focusing on specific scientific stages are already providing solutions to improve open and reproducible science, automated ledger networks accounting fully for the research cycle to facilitate neutral knowledge-inspired societies and governance are missing. In this project we aim to build an automated ledger knowledge network technology to compactly facilitate open-science, decentralization and security in the science ecosystem (Figure 1). Our final aim is to provide real-time open-access neutral data-rule-knowledge to gain informed decisions and help solve complex social and technological problems.

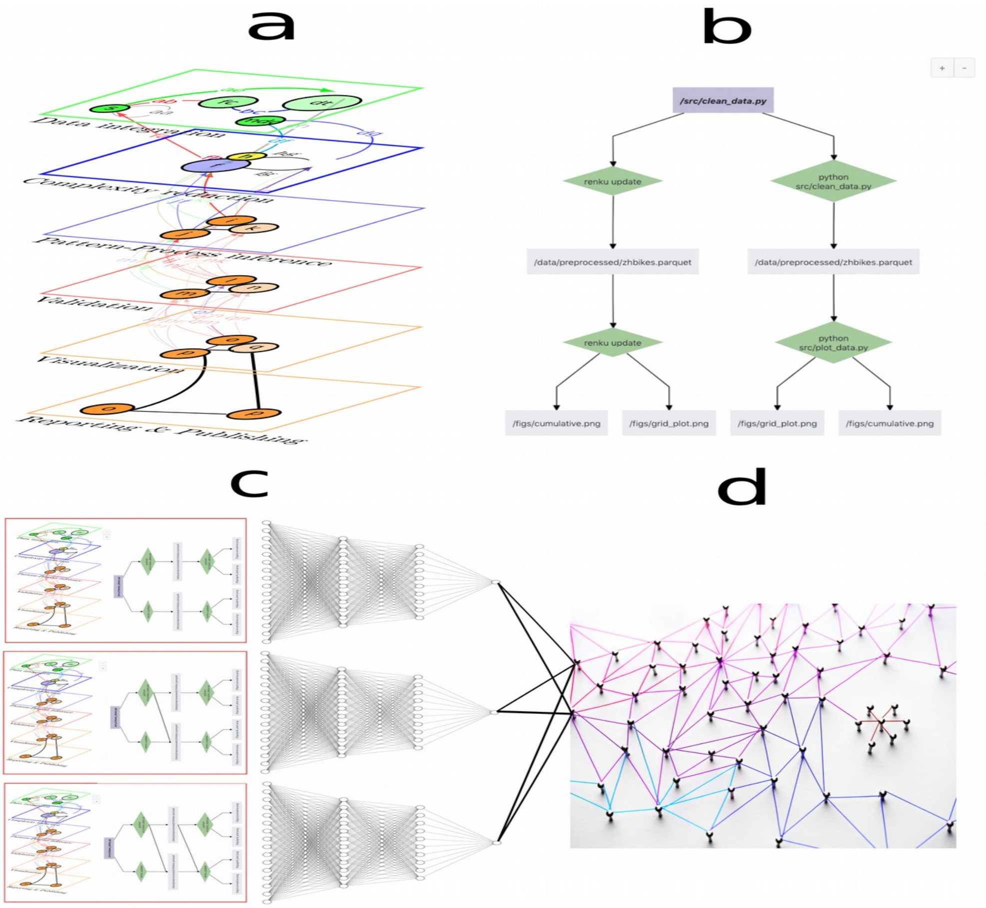

Figure 1: Deep ledger knowledge-based *network technology.* ***a****) End-to-end research cycle from data integration to reporting.* ***b****)* The knowledge graph (KG) tracking one research path of **a** (i.e., *Renku* open-source code). ***c****) Deep knowledge-based network can* automatically explore a population of KGs to gain rule-based knowledge of the data. d) The population of all KGs can be stored in a ledger 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 science ecosystem is error prone with multiple steps requiring information transfer among trusted/untrusted peers. In this ecosystem, immutable peer-to-peer architecture with a mixture of trusted-untrusted peers is key to have open access to full reports to gain informed decisions in complex societal, environmental and technological problems. The overall objectives of the proposal with the specific goals of each one are the following:

G1: Deploy an automated *knowledge-based network technology* accounting fully for the research cycle in a lineage client-tracker to produce a population of *Knowledge Graphs* (KGs) (Figures 1a and 1b.)

G1.1: Intralayer automation of data integration, inference, and validation (Figure 1a.)

G1.2: Intralayer automation of visualization and reporting generation (Figure 1a.)

G1.3: Deep intra- and inter-layer automation with a lineage client-tracker to sample paths in the multilayer network (Figure 1a.)

G2: Deploy an *end-to-end permissioned-permissionless distributed ledger technology* to guarantee decentralization, open-access and security of the KGs in the science ecosystem (Figures 1c and 1d.)

G2.1: Blockchain implementation accounting for consensus algorithms and smart contracts among trusted-untrusted peer-to-peer events.

G2.2: Scenarios to minimize scalability-security-decentralization trade-offs.

G3: *Robhoot Open Network in Biodiversity Research as a c*ase study to facilitate accessibility to open science to provide real-time open-access data-rule-knowledge to gain informed decisions and help solve local and global environmental problems.

G3.1: Biodiversity datasets integration (Figure 1a top layer.)

G3.2: Interaction with G1.2 to test Biodiversity patterns in a deep ledger knowledge network (Figure 1.)

G4: *Launch a testnet* (or prototype) to collect data to explore the security-scalability-decentralization patterns and the robustness of the generated KGs in the deep ledger knowledge network.

G4.1: Interaction with G2.2 to test the scalability-security-decentralization trade-offs.

G4.2: Interaction with other platforms/decentralized networks to facilitate real-time open-access data-rule knowledge to general case studies.

1.2 Science-to-technology breakthrough that addresses this vision

Peer-to-peer interactions composed by trusting and untrusting peers abound in social, economical, natural and technological ecosystems. Many studies in such systems are producing an immense gain in detailed knowledge about scalability, security and decentralization trade-offs (refs; TON network; Fabric ledger OS network). Most studies about these trade-offs have considered one-level networks. Yet, information generation usually comes from the interactions within- and between-layers, and the feedbacks occurring among layers in these systems have provided new unexpected behaviors that are difficult to anticipate when exploring one layer alone (refs). In biological systems, the genetic architecture of functionally important traits feedback throughout the genotype-phenotype map producing variation in phenotypes that are functionally important to understand the evolution of genotypic and phenotypic variation like growth rates and the immune system that ultimately determine the frequency of the phenotypes and the interaction centrality patterns in natural populations (refs). In science and engineering, many steps within- and between-layers occur to generate information (Figure 1a) Similarly to biological systems, interactions including intra- and inter-layer feedbacks are not easy to reproduce if not properly accounted for. One of the main facts when accounting for more than one layer is that the interactions and feedbacks to each other produce a dynamics that significantly differ from the one-layer approach (refs). Accounting for levels and scales in many systems using multilayer networks have provided a framework to explore how the microdynamics of peer-to-peer interactions might connect to the macroscopic properties of the ecosystem like the centralization and and the sensitivity to attacks within and between layers (refs).

There are currently automated platforms mostly in the private domain focusing in specific parts or one layer/one path of the research cycle (BigQuery, Modulos, Google AI, Automated statistician; Ghahramani, 2015, Ease.ml; Li, Zhong, Liu, Wu, & Zhang, 2017): Novelty of reporting generation following one path or resource allocation when exploring many parts of the research cycle. The science ecosystem, however, still lack a framework automating the research cycle from end-to-end into the scalability-security-decentralization trade-offs of digital ecosystems. Many science and engineering projects have failed in reproducibility in public-funded science and technology (refs). Yet, despite public institutions are demanding more reproducibility and openness of the data and the scientific cycle (refs), and overall a shifting towards open and reproducible scientific and engineering landscapes, there are not currently open technologies aiming to compactly facilitate and distribute the scientific and engineering cycle in immutable knowledge networks.

1.3 Interdisciplinarity and non-incrementality of the research proposed

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 a technology will require 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 "*de novo*" technology: the synthesis of automated knowledge generation in a neutral, borderless and immutable network synthesized anew from existing open-source projects like Renku, Fabric and gitchain.

1.4 High risk, plausibility and flexibility of the research approach

We are in need of accounting for the uncertainties, the reproducibility and immutability related to automation in science and engineering. This need is not just for a specific stage of the research cycle, but for the full research cycle, from data acquisition to reporting generation because knowledge-inspired societies and governance will demand full research cycle transparency in solving complex social, environmental and technological problems. This need brings many 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. We will develop a flexible research method focusing more in the *algorithmic robustness of the deep ledger knowledge network* than in the development of *robust automated knowledge generation*. Our motivation will be to provide a first proof of concept of how the technology works: we will sample the KGs using different deep learning algorithms to estimate the uncertainty of the ruled-based inference obtained by fitting predictions to simulated data (Goal G1). 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 (Goal G2). Despite our focus will be bias towards the side of the algorithmic robustness of the deep ledger knowledge network, we will develop a domain-specific case study, our Robhoot Open Network, to test the robustness of the rule-based inference obtained by fitting each of the generated KG to the empirical patterns (Goal G3). The high risk associated to robustly automate the full research cycle for producing immutable open knowledge is buffered to a great extend because the existing ecosystem of tested and reliable open-source tools: We will combine our own algorithms (i.e., data integration and deep learning algorithms for sampling and automating the KGs) with open-source tools like Renku, Fabric and gitchain. This open-ecosystem will allow us to have a flexible l*aunching of a testnet* to collect data to explore the security-scalability-decentralization patterns and the robustness of the generated KGs in the deep ledger knowledge network (Goal G4.)

2. Impact

2.1 Expected impacts

The implementation of an automated end-to-end research cycle within an open ledger will facilitate real-time open-access neutral data-rule-knowledge to gain informed decisions to help solve complex social, environmental and technological problems. This facilitation might occur for local, regional and global problems in many fronts. Specifically, open deep ledger knowledge networks might have an impact in the following five areas:

1. Automated populations of KGs accounting for the full research cycle to help building the science of science by providing data to account for bias, errors, and robustness in rule-based inference and predictions: Increasing the robustness of scientific contributions to new future technologies.

2. The open access of immutable accumulation of knowledge in untrusted digital peer-to-peer ecosystems: Social, environmental and economic impact to facilitate global access to neutral knowledge.

3. The identification of gaps in research paths not explored consequence of lack of synthesis in interdisciplinary research: The creation of new markets opportunities obtained from exploring these gaps.

4. The merging of prediction and explanatory power to gain synergy between AI predictive tools and ruled-based pattern inference creating a more balanced pattern and process inference. Recent examples of AI algorithms playing chess and go using brute force deep learning models or rule-based algorithms: The integration between prediction and understanding power to facilitate explanatory synthesis.

5. The automation of reproducible open knowledge will facilitate the reusability, repeatability, and replicability of research outputs. The open access knowledge for governance transparency.

2.2 Measures to maximise impact

a) Dissemination and exploitation of results

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.

b) Communication activities

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