Letter of Intent Caixaimpulse

1. General Information

1.1. **Applicant Entity**

Agencia Estatal Consejo Superior de Investigaciones Científicas CSIC

1.2. **Applicant Entity Tax ID Number (NIF/CIF)**

Q2818002D

1.3. **Project title (maximum characters: 100)**

Automated Rapid Discovery and Transfer in Global Emergency Health Situations

1.4. **Project acronym**

AUDAS

1.5. **Name of the project leader**

Víctor Martínez Eguíluz

1.6. **Select the business area**

Digital Health

1.7. **Participation of the project in other "la Caixa" Foundation Calls**

No

1.8. **What is the protection status of the Asset(s)?**

No protection status

1.9. **Time to market**

Less than 2 years

2. Technology Readiness Level (TRL)

TRL 5 – Technology validated in relevant environment

2.1. **Has the Asset(s) been transferred to a Spin-off company?**

No

2.2. **Amount requested from the "la Caixa" Foundation**

The total amount request will be of around 153k €. The following is the list of the four main parts of the budget:

- **Discover phase**:
- 2y Postdoctoral researcher: 78k €
- **Transfer phase**:
- HealtHack workshop to develop reciprocal data-management and technology-transfer protocols at local level: 20k €
- Interface development from the experience gained in the HealthHack: 6m Data or Computer-scientist: 25k €
- International workshop with health managers to explore protocols for data-standards and scalability at the global scale: 30k €

2.3. **Describe the core idea of your application in one sentence**

Automated rapid discovery and transfer of data- and causal-knowledge graphs to facilitate informed decision and complex problem solving in human health ecosystems at local and global scales.

2.4. **Abstract**

Global Health is a major goal of humanity challenged by several pathogens along human history. Many studies have shown global health and sustainability could be achieved by strengthening transparency and feedbacks among social, ecological, technological and governance systems. Global health and sustainability goals, however, strongly depend on rapid access to evidence-, research-, discovery-, and transfer-based knowledge gaps, especially in global emergency situations. Yet, the science-technological-governance ecosystem lack technologies narrowing down global knowledge gaps facilitating rapid access to information to strengthen feedbacks among social, ecological, technological and governance ecosystems. We propose a rapid automated discovery and transfer technology targeting knowledge-gaps and knowledge-causal graphs throughout reproducible and rapid report generation in federated networks. A federated discovery network encompasses a hybrid-automated-technology to lay out the foundation of an open-, cooperative-science ecosystem to automate discovery strengthening rapid solutions and robustness at the local and global scale in emergency and global health challenges. The project aims to provide the architecture of a science-enabled technology to connect global human challenges to easy access neutral-knowledge generation for rapid decisions in knowledge-inspired societies.

3. Scientific and technical feasibility

3.1. What are the scientific fundamentals underlying your idea? Have you performed a proof of concept validation?

The scientific fundamentals of the present project are based on the growing computing capacity, the integration between causal-knowledge graphs and explainable Artificial Intelligence, and the interconection of open-source technologies in the rapidly evolving digital ecosystem. Targeting these fundamentals in automated discovery and transfer requires the integration of different modules, from identification and retrieval of heterogeneous data sources, to the integration of explainable modeling and causal-inference, and visualization and reporting. We have already advanced in the integration of the different modules, from the automated identification, retrieval and data integration to inference and process-based discovery. We have implemented a prototype for the ongoing covid-19 pandemic. Each module includes state-of-the-art developments in computer science, complex systems, and theoretical evolutionary ecology. The proof-of-concept is not fully automated yet and still requires human intervention in module integration and the development of a testnet stage. Nonetheless, we are currently exploring innovative solutions especially in the modules of automated data discovery, causal-knowledge graphs, reporting, and visualization.

3.2. Describe how the Project's Leader knowledge and experience in field of the Asset(s) would contribute to the success of the Project. Discuss it in the context of the composition and the relevancy of the rest of the team members.

The PL has a long tradition in interdisciplinary and computing related projects, with team members coming from different fields. PL has experience 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. Other team members with a broad expertise in scientific computing and open-source development in the context of biodiversity and global sustainability projects include Dr. Melian, ETH-Domain, Switzerland.

3.3. Has your Asset(s) been under evaluation by other funding programmes prior to this call *(an object of research grants or other innovation or acceleration programmes)? Please detail the most relevant ones*

No

4. Transferability and market potential

4.1. According to the stage of development of the Asset(s), what are **the identified needs and determinants that currently condition its successful progression** *to the commercialisation phases?*

During the development process of the prototype (i.e., Robhack workshop during March 2020 and a git repository with five-to-ten contributors per day), we have identified the following needs currently conditioning the progression of the asset.

- From the human resources perspective, team members have volunteered in a coordinated way using collaborative open-source software to develop and integrate the different modules of the asset. We are currently in need of two researchers, a postdoctoral with expertise in Julia computing language and a computer-scientit or developer for interface implementation. The team has recently participated in a intensive one-week Robhack event, i.e., hackaton style for automated technologies, where the different modules have been identified, integrated and tested.

- From the technical data-integration part, while we have been able to integrate state-of-the-art algorithms, there are still some gaps that need to be addressed. For example, data discovery for biomedical knowledge bases requires highly general and fault-tolerance algorithms to access heterogeneous APIs. We were successful in reaching many APIs types for Covid-19 data, but still, more general and fault-tolerance algorithms are required to automate API discovery and data-knowledge graphs.

- From the technical AI and causal-knowledge graph part, we have been able to test a variety of fitting methods to large and complex causal-knowledge graphs to obtain the sequence of events that best fit to the demography parameters in the Covid-19 at city and global scales (i.e., susceptibles, incubation period, infected, recovered, and deaths). Yet, generalizing neural nets and causal-knowledge graphs will require novel implementations of neural ordinary differential methods to rapidly optimize models of thousands of parameters.

4.2. **Describe the specific innovation transfer milestones you wish to achieve** *by participating in this Call in terms of time-to-market*

The participation in this call will allow us to transfer to the local and national health system an automated discovery platform using locally collected and standarized data to show the importance of rapid and reproducible discovery for predicting complex problems at the interface of social and governance systems. Specifically, to reach the market, the following milestones have been identified: At the end of the first year a full operation Asset will be available and tested in a human health case study: the propagation of the covid pandemic at local and global scales (2y Postdoctoral

resercher). During the second year, the Asset will be finished with a user-friendly interface environment aiming to be useful for the non-expert (6m Computer scientist or developer).

4.3. **Who are your end-users?** *What is the market segment and the market size you are addressing?*

We identified as potential end-users of our Asset, a person, organization or governance entity that generate research-based-knowledge and/or requires reports for decision making. Specifically:

- Health Labs and biomedicine institutions: While part of the research is used to analyze data and modeling, a large part of the community are end-users that use packages or software to either analyze data and/or modeling. This sector will benefit from the Asset as it will help to obtain automated robust results.
- Public administration: Decision making and governance require analysis of the risks and emergency situations at local and global scales to make decisions that will be beneficial collectively at these scales. For example, the ongoing pandemic is showing us the need to account for almost-real-time automated analysis of the situation where most of the epidemiological variables are unknown. The Asset will assist decision-makers in this regard by providing rapid and reproducible reporting accounting for the state-of-the-art data and modeling technologies.
- Private companies and corporations: The growth and survival of a company often requires constant innovation and evolution. The Asset will facilitate the discovery cycle in private companies offering them rapid updates of new emerging situations, localy and globally.

5. Societal relevance and potential impact

5.1. **What is the unmet need you aim to address** *and why is it timely and relevant?*

Access to information is key for modern societies facing complex problems including global health. Living in an ultra-connected world where news travels constantly and fast from one point to the other around the globe might produce misinformation and information deluge. The growth of the so-called fake news makes the picture even more complex. This is particularly timely and relevant for anticipating decisions on unexpected and complex health emergency and sustainability situations at local, regional and global scales. To solve these problems, private companies emerge for example as validators of the information. Dentralization and transparency from the source of the data to the reporting is key to facilitate the validation of the information used to solve complex problems. Our asset will contribute to facilitate reproducible reporting generation by providing a the tools and the interface at the local, individual level.

5.2. **What potential impact will your solution have?** *(maximum characters: 1500)*

Automated discovery can have a large impact to people in need to access rapid, robust, and reproducible reportings to take informed decisions, especially in global emergency and/or in sustainability situations.

5.3. **What are the available solutions that currently address the specific problem you aim to solve?**

Most of the available solutions offering rapid and semi-automated reporting generation to solve complex problems in health emergency situations at local and global scales are based in highly fragmented, non-reproducible and black-box technologies. This situation currently inhibits the rapid access of many local companies and governance entities to a technology offering integration, reproducibility and transparency of the reportings to deal with rapid uncertain coonditions in health emergency situations.

5.4. **What is the innovation of your solution? Who will be the main beneficiaries of your solution? (maximum characters: 1500)**

Our innovation to help take informed decisions in emergency and sustainability challenges come from two fronts: First, we will integrate highly fragmented and non-reproducible technologies into a compact reproducible and automated technology. Second, novelty comes from automating data-, and causal-knowledge graphs into an integrated workflow. We aim to release a completely free and open-source prototype for developers to facilitate robustness and scalability of the technology. We will have additional features to offer to research labs and institutions, public administration and private bussiness to facilitate explainable and reproducible reporting generation.

6. Additional Documentation

Copy of the identification document of the Project Leader (DNI/NIF/ID)Letter of commitment from the host institution (template here)