

Trustees Annual Report

Ersilia Open Source Initiative

Period: 10th of November 2020 – 30th of November 2021

Trustees of the Charity during this period:

- Dr. Edoardo Gaude
- Dr. Miquel Duran-Frigola
- Mr. Akash Rungta (from 1st July 2021)

In addition, Ms. Alacia Armstrong was appointed Trustee of the Charity post year end (14th December 2021).

The Trustees of the Ersilia Open Source Initiative (the Charity) are pleased to present their report together with the financial statements of the Charity for the year ended 30th November 2021. The trustees have complied with the duty to have due regard to guidance issued by the Charity Commission on public benefit and have adopted the provisions of the Statement of Recommended Practice (SORP) for this report.

OBJECTIVES AND ACTIVITIES

The Charity's objectives for the public benefit, as set out in their Constitution are:

1. The advancement of research in relation to the discovery and development of new medical treatments.
2. The generation of open source machine learning tools for drug discovery to be used by the global community.

Focusing in the Sustainable Development Goals (SDGs), as defined by the United Nations, the Charity is contributing to the objectives:

- SDG 3: Good Health and Well Being: we aim to deliver new and improved medical options to those living in low-resourced areas
- SDG 9: Industry, Innovation and Infrastructure: we disseminate new technologies and focus on making them usable by the broader scientific community
- SDG 10: Reducing Inequalities: we work to increase the research and innovation of LMIC

These objectives are translated into the following activities for the public benefit:

Addressing unmet medical needs in the fields of infectious and neglected tropical diseases. Disease burden distribution remains highly unbalanced worldwide. Six of the top ten causes of death in Low and Middle Income Countries (LMICs) are still due to infections, including Malaria, Tuberculosis and HIV/AIDS. In contrast, in High Income Countries (HIC), where less than 20% of the population resides, non-communicable diseases are more prevalent, including cardiovascular complications, dementia and cancer. Strikingly, only 10% of the drugs currently in development are targeting infectious diseases, effectively neglecting the endemic health problems of LMICs (WHO Data, 2021). The main reason for this disparity is the

high cost of developing new medicines, estimated to be 1.5 Billion USD per drug, which has led the pharmaceutical industry to overlook therapeutic areas with expected low return on investment. By emphasising the development of research projects targeting these disease areas, and translating assets developed in the context of non-communicable diseases to the field of infectious diseases, the Charity aims to deliver new cost-effective solutions to treat conditions suffered by a large proportion of the world population, which are currently orphan of treatment.

Reducing the imbalance in research productivity between countries. LMIC research output (measured in the number of science and engineering publications) is less than 10% of the global research output. This has severe implications for the living conditions in low-resourced countries, as they depend upon scientific and technologic solutions devised in the Global North, and that, oftentimes, do not meet the real needs of their populations. Many researchers, both from the Global North and the Global South, claim that the global trend in scientific collaborations (also referred to as helicopter research, where scientists from HIC benefit from collaborators in LMIC to collect data and coordinate logistics, but leave them out of the actual scientific discussions and authorship) is harming science by a) focusing on goals that might not be relevant to the actual problem b) underusing the potential of researchers located in underserved areas and c) perpetuating established power dynamics. A central part of the Charity's activities for the public benefit, thus, are focused on empowering institutions in LMICs and augmenting their research capacity through a sustainable adoption of digital assets. Contrary to the classical North-South collaboration, where endemic countries are mere facilitators of field research, our project puts the focus on the research activities and ideas of the local institutions.

Implementing Open Source Artificial Intelligence and Machine Learning (AI/ML) models. The Charity chose to fulfil the above mentioned activities by using AI/ML and other computational assets for three reasons. First, AI/ML holds the potential to revolutionise the drug discovery field. By leveraging millions of data points produced in laboratories around the globe, AI/ML provides valuable predictions about the therapeutic efficacy of new drug candidates, putative side-effects or co-administration issues with other drugs. Simply put, AI/ML allows the screening of larger libraries of compounds in less time. This is particularly important in drug repurposing efforts, or in the exploitation of natural products, which are key strategic interests of LMICs. Second, AI/ML is a low-cost solution that may drastically reduce the number of necessary experiments to find a new drug, reducing the prohibitive costs of drug discovery in low-resourced settings. Third, there is a growing community of open-source contributors and enthusiasts, which facilitates the disclosure of technological developments in AI/ML through permissive licences, fostering Open Science, fair attribution and collaboration.

Distribute and disseminate the tools, research outputs and training materials following the principles of Open Science. To have a meaningful impact on the work of researchers worldwide, and ultimately, improve the quality of healthcare for the general public, the Charity is a strong supporter of Open Science. We believe that making the code public (open source) is only the first step, but not enough, since it still requires expertise in computer science and programming and many scientists cannot implement those assets in their daily experimental

pipelines. To this end, we focus on deploying the tools we develop in a user-friendly manner that does not require previous coding skills to lower the barrier to access these new technologies. Moreover, we publish all findings in Open Access journals, to ensure everyone can read them regardless of their institutional affiliation. In addition, we include local scientists in project development when we collaborate with other institutions, as the most straightforward way to deliver capacity building and training. Finally, we emphasise the importance of raising awareness among the general public, for which we use social media and blog posts to disseminate, in a non-technical manner, our findings.

Ersilia, the invisible city: the Trustees wish to explain the chosen name of the Charity, as it reflects its mission and vision and might help those who read the report to better understand our work. Ersilia is one of the cities imagined by Italo Calvino in its book *Invisible Cities*... *"In Ersilia, to establish the relationships that sustain the city's life, the inhabitants stretch strings from the corners of the houses, white or black or gray or black-and-white according to whether they mark a relationship of blood, of trade, or authority, agency. When the strings become so numerous that you can no longer pass among them, the inhabitants leave: the houses are dismantled; only the strings and their supports remain. From a mountainside, camping with their household goods, Ersilia's refugees look at the labyrinth of taut strings and poles that rise in the plain. That is the city of Ersilia still, and they are nothing. They rebuild Ersilia elsewhere. They weave a similar pattern of strings which they would like to be more complex and at the same time more regular than the other. Then they abandon it and take themselves and their houses still farther away. Thus, when travelling in the territory of Ersilia, you come upon the ruins of the abandoned cities, without the walls which do not last, without the bones of the dead which the wind rolls away: spiderwebs of intricate relationships seeking a form."* Like Ersilia the city, our Charity aims at laying strings, starting relationships and building a network of scientists working together for the benefit of those who suffer diseases for which there is yet no cure.

ACHIEVEMENTS AND PERFORMANCE

In its first year of existence, the Charity's activities have been focused on:

- Setting up the necessary infrastructure to provide service to its beneficiaries
- Establishing partnerships with key stakeholders that will help advance the activities of the charity
- Releasing the first Open Source tools
- Raising funds to sustain our activities.

In the following paragraphs, the trustees aim to summarise the main achievements of the Charity in these areas.

Development of the Ersilia Model Hub. The Ersilia Model Hub (EMH) is a free, online, open-source platform where scientists can browse through a catalogue of AI/ML models, select the ones that are relevant to their research and run online predictions without the need to write a single line of code. The EMH gathers, in a single resource, two classes of models. On the one hand, we collect models developed by third parties and available in scientific publications. On the other hand, we develop models in-house and/or in collaboration with research groups in LMICs. Similar platforms exist in different contexts, and are having an immense impact on their

fields, like Hugging Face in Natural Language Processing, but none is focused on drug discovery or neglected diseases. The EMH is the main asset of the Charity and will be instrumental to deliver its mission. During this year, we have successfully developed and tested an alpha version of the EMH, which features >30 models and works in a command-line interface (CLI). Models are fetched from our GitHub repository and can be run locally. A minimally viable product (MVP) will be officially released in January, 2022, but it can already be accessed on our GitHub repository (<https://github.com/ersilia-os/ersilia>) with associated documentation. We expect to reach a hundred users in the first months of its release. To our beneficiaries (the broad scientific community), having access to the EMH represents the possibility of speeding research projects by applying AI/ML tools, accessing this expertise on their daily tasks and, in general, benefiting from third party research results following the principles of Open Science. By improving scientific research and collaboration we hope to benefit the wider society by facilitating the improvement of healthcare with safer and more affordable drugs.

Development of an automated pipeline for AI/ML models in chemistry. Over the past year, we have realised that automatizing the process of training new AI/ML models using collaborator's data would be key to speed up these partnerships, increase our capacity to work with our beneficiaries and generate more assets for the whole scientific community. To this end, we have developed a pipeline, named ZairaChem, that takes molecular data as input and trains a predictive model using state of the art technology. Thanks to it, we have been able to engage in a 3-month collaboration with the H3D Centre in Cape Town and develop a pipeline that includes over 15 AI/ML models (see below). Moreover, ZairaChem is not only used by the Charity, but it has also been released to the public, so any scientist with interesting data can now test their own models. We also encourage model deposition in the EMH of third-party assets, to increase its visibility and benefit the larger scientific community. ZairaChem is a milestone in the Charity's mission, and we expect it to become a main tool for our beneficiaries.

Generation of new antimalarial leads. The Charity has participated in the Open Source Malaria Consortium (OSM), a worldwide effort to identify patent-free drugs against malaria. The OSM initiative, led by Prof. Matthew Todd from the University of London, works fully on the open, and engages equally academic and industry partners in their search of new chemical structures with activity against *P.falciparum*. Selected molecules are synthesised and tested in laboratories around the globe, and progression to more advanced clinical stages will be carried out by new non-profit pharma companies such as M4Id. In this context, the Charity used experimental data available on a chemical series of interest (Series 4) to train a reinforcement learning algorithm that generates new molecules from the same chemical series with enhanced activity against *P. falciparum*. In short, we generated over 300.000 new compounds and shortlisted 10 of them to be synthesised in the laboratory. In addition to the molecules, we also provided an online tool to predict activity against *P.falciparum*, a visualisation tool of the chemical space based on TMap and predicted the structure of ATP4, the *P.falciparum* target of Series 4, using the recently released AlphaFold2. All the assets generated are available on the Charity's GitHub, and the scientific discussion arising from these results can be found on the OSM public GitHub as well. The participation in this project has served the Charity's beneficiaries in several ways. First, by proposing new molecules to target malaria, a disease that kills half a million people a year. Second, by developing and releasing new AI/ML protocols

that can be adapted by any scientist to their needs. Third, and most importantly, by strengthening novel research models like the one proposed by OSM, which rely on open science as opposed to the traditional patent-based pharma industry model. This allows faster collaborations, reduces duplication of efforts and releases a large amount of data that can be used by other researchers, and, most importantly, aims to deliver cheaper drugs to the market.

Implementation of a virtual screening cascade at H3D. The Charity has partnered with the Holistic Drug Discovery and Development (H3D) Centre from the University of Cape Town, South Africa. The H3D centre is focused on drug discovery for malaria and tuberculosis, and is the lead institution in the field in Africa, also championing capacity building and networking amongst other African countries. The research community at H3D and its extended network represent one of the main groups of beneficiaries of the Charity, and thus this partnership is a key step towards broadening the impact of the Charity's activities. In the context of the partnership, Charity volunteers performed a 3-month research stay at H3D, during which they developed a pipeline of AI/ML models that mimics the experimental screening cascade. These models, based on in-house experimental data, will enable selecting the best molecules to progress in the drug discovery process without the need of running the experiments, therefore reducing the cost and time to find new drugs. Together, malaria and tuberculosis are responsible for the loss of millions of lives each year and cause billions of dollars of economic loss to countries, becoming a cause rather than a consequence of poverty. Malaria first-line treatment resistance is rising, and no new drugs against tuberculosis have been developed in the last thirty years, emphasising the need to strengthen the research into these diseases. By working with H3D, the Charity aims to speed up this research and improve the likelihood of finding new treatments. Moreover, these assets will also be incorporated to the EMH to the benefit of other scientists working in the field.

Delivering capacity building and outreach activities. A central pillar of the Charity is not only to release new AI/ML models and computational tools for research but also to ensure scientists can use them appropriately. To this end, and while the Charity is not an education institution, we partner with higher education institutions, research institutes and schools to ensure sustainable transmission of knowledge about the digital assets we develop, and to encourage the new generations of scientists to explore the field of computer and data science. In this direction, we have delivered over 10 talks in international conferences, like the Joint Roadmap for Open Science Tools (JROST), the Bioinformatics Open Science Conference (BOSC) and the BioAfrica Convention. We have also delivered talks to schools (mid-grade), university undergraduates and MSc students to disseminate the Charity's work and encourage students, particularly those underrepresented in STEM disciplines, to consider a career in these fields. In addition, we have created training materials and content to introduce experimental scientists to Python (the most widely used programming language) and tailored the exercises to real life applications. Finally, we prioritise on-site project development over remote work in order to work closer with our collaborators, include their staff in the project and facilitate knowledge sharing. This approach has been a success during the H3D research stay, in which we have worked hand-in-hand with 5 H3D researchers, delivered talks to the whole staff and delivered our introduction to Python and AI/ML to a selected group of 13 scientists. The Charity Trustees feel that the impact of the capacity building activities has been larger than expected, and has directly benefited over 50

scientists who now have better knowledge and tools to start incorporating state of the art computer methods into their research.

Future plans

The Charity will continue and expand the projects described above in the next year, and engage in novel collaborations as well. The Trustees would like to outline the major advances and plans since the end of the reporting period (30.11.2021). First, the partnership with the H3D Centre will continue with the development of a second project related to African pharmacogenomics, funded by GSK and Novartis (grant conditionally approved at the moment of writing). Second, EOSI has partnered with the University of Buea, Cameroon, to develop a 5-year project related to the use of natural products as an antiviral source. This project is funded by the Bill and Melinda Gates Foundation. Third, the Charity has been selected as one of the six projects participating in the first cohort of the Digital Infrastructure Incubator, offered by the Code for Science and Society and focused on community engagement and governance models for open source initiatives. Also, the Charity has been accepted into the Fast Forward Accelerator, a 5 month incubation program designed specifically for tech nonprofits. These two Incubators will set the basis for the sustainable growth of the Charity and provide the necessary resources to improve fundraising activities, community engagement and impact reporting. Finally, we have been given a grant to access the Barcelona SuperComputing Center - MareNostrum resources, which will be used in the period November'21 - February'22, increasing the Charity's AI/ML model production capacity

FINANCIAL REVIEW

Financial position

The financial statements have been presented in the Receipts and Payments format and submitted for independent account examination by Streets Chartered Accountants (Cambridge, UK). All income and expenditure is expressed in GBP, and income or expenditure in other currencies has been converted to GBP as per official conversion rate at the moment of receiving or emitting the expense. Donations and grants for purposes specified by the donor/funder are treated as restricted funds. The rest are treated as unrestricted funds, which are expendable at the Trustees' discretion to further the charitable objectives and support the associated costs of running projects.

Incoming resources

The Charity has received voluntary donations, grants to further its activities for the public benefit and income in return for charitable services. All income from the 2020-2021 period falls within the category of charitable income and is thus exempt from corporation tax as detailed by HMRC. The major income source is a grant from Merck Healthcare to develop a privacy-preserving tool to encrypt AI/ML models (25010 GBP). This grant will be expended in the financial year 2021-2022. The second major source of funding comes from the crowdfunding campaign FundOSS, which provided 9829 GBP in unrestricted funds. Minor income comes from charitable services offered to non-profit research institutions (ISGlobal and H3D). The total income for the reporting period is 41553 GBP.

Expenditure

The major expenditure, 3602 GBP, relates to the purchase of a computer workstation featuring a GPU unit (a graphical card to enable machine learning model training), which was paid with the money received from the OakDale Trust, the Astor Foundation and the FundOSS campaign. 2579 GBP was spent to develop the Charity's website and branding material, essential as a newly incorporated organisation. 853 GBP were paid as a service for the development and delivery of programming courses for our beneficiaries. 143 GBP were spent in various licences for the daily work of the Charity (Zoom, Google Colab, Website Hosting, AirTable and Loom). The total expenditure for the reporting period is 7177 GBP.

Trustee remuneration

The trustees received no remuneration. The design studio, Forma.co, for which a connected person to the trustee Dr. Miquel Duran-Frigola works, has received a fee for service of 2579 GBP, which has been approved by the Board of Trustees (1st of July, 2021) in accordance with the Foundation Constitution.

Reserves policy

The Charity holds cash at bank of 34376 GBP, 6941 GBP in unrestricted funds and 27435 GBP in restricted funds. The Trustees aim to maintain free reserves in unrestricted funds at a level which equates to approximately three months of unrestricted charitable expenditure. Based on this year's expenditure, the free reserves target amount is 1750 GBP. There is currently an excess free reserves because the Charity aims to significantly grow its activities, and therefore expenditure, in the year 2021-2022. The Trustees aim to use these free reserves to better support the increase in activities in the coming months and cover expenditures related to equipment, software licences, and invest in the quickstart of new projects and partnerships.

Going concern assessment

The Charity has raised 41553 GBP in its first year of operation, and has achieved major success in securing funding for the next year from the Digital Infrastructure Incubator, the Fast Forward Accelerator, the Bill and Melinda Gates Foundation and the African Gradient project. The Trustees believe these are more than sufficient to ensure the continuation of the Charity's activities in the next two years.

In addition, the Trustees acknowledge the major role the Charity volunteers have played in setting up the infrastructure, developed projects and a track record for the charity and managed the fundraising activities. The Trustees are also aware that the Charity's operations will require at least three full time employees in the next two years. The expected funds in the period 2021-2022 will support the incorporation of these.

STRUCTURE, GOVERNANCE AND MANAGEMENT

The Charity is governed by its Trustees, who are also its voting members, according to the Foundation Constitution Model. The Trustees set the mission and vision of the Charity, oversee that all the activities undertaken help further such mission and vision and take care of strategic

decisions and viability of the project. They delegate the day-to-day decisions and implementation of the strategy to the Chief Executive Officer, currently a volunteer position covered by Dr. Gemma Turon, and the scientific direction is delegated to the Chief Scientific Officer, currently also filled in in a voluntary manner by Trustee Dr. Duran-Frigola. The Board of Trustees meets four times a year, and monthly one or more Trustees meet with the CEO and CSO of the Charity to discuss strategic decisions and provide guidance.

The legal structure of the Charity is a Charitable Incorporated Organisation, registered with the Charity Commission for England and Wales with number 1192266.

New Trustees are elected by current Trustees. The criteria for choosing new Trustees is based on the eligibility conditions set out by the Charity Commission as well as: a) their knowledge of the Charity mission and vision, b) their alignment with the values represented by the Charity c) the skills and experience that they can bring onto the Board that are currently lacking. The Charity cannot have less than two trustees and there is no limit to the number of Trustees in the board, though the average number should be around 6 members at any given time. Since this is the first year upon incorporation, the current trustees are looking for suitable new candidates, and plan on advertising the role internationally in the coming year. Currently, there is no limitation on the time a trustee can serve.

New Trustees are appointed at a general trustee meeting and briefed on their legal obligations, the charity commission guidance on public benefit and general guidelines for Trustees. They also receive the Foundation Constitution document, the strategic plans of the Charity, the business plan and an account of the financial situation of the Charity.

REFERENCE AND ADMINISTRATIVE DETAILS

The Administrative Details of the Charity are as follows:

Registered Name: The Ersilia Open Source Initiative

Other names: the Charity is also referred to as EOSI or Ersilia

Charity Registration Number: 1192266

Address of its registered office: 28 Belgrave Road, Cambridge, CB1 3DE, United Kingdom

Trustees who served during the reporting period:

- Dr. Edoardo Gaude
- Dr. Miquel Duran-Frigola
- Akash Rungta, Msc (from the 1st of July 2021)

Ms. Alacia Armstrong was appointed Trustee of the Charity post year end but before the Accounts were approved by the Trustees

EXEMPTIONS FROM DISCLOSURE

There is no exemption from disclosure to report.

FUNDS HELD AS A CUSTODIAN



There are no custodian trustees.

The Annual Report has been revised and approved by:

Miquel Duran-Frigola

04 / 29 / 2022

Edoardo Gaude

04 / 29 / 2022

Akash Rungta

05 / 02 / 2022

Ersilia Open Source Initiative**Financial Statements for the period from 10th November 2020 to 30th November 2021**

Receipts and Payments accounts

Section A: Receipts and Payments (to the nearest GBP)

	Unrestricted Funds	Restricted Funds	Endowment Funds	Total Funds	Last Year
A1 Receipts					
General Donations	12,443	0	0	12,443	0
Research Grants	0	25,010	0	25,010	0
Receipts from charitable activities	1,675	2,425	0	4,100	0
A2 Asset and investment sales	0	0	0	0	0
Total Receipts	14,118	27,435	0	41,553	0
A3 Payments					
Licenses	143	0	0	143	0
Cost of charitable activities	853	0	0	853	0
Cost of generating voluntary receipts	2,579	0	0	2,579	0
A4 Assets and Investment Purchases					
GPU Workstation	3,602	0	0	3,602	0
Total Payments	7,177	0	0	7,177	0
Net of Receipts/(payments)	6,941	27,435	0	34,376	0
A5 Transfers between Funds	0	0	0	0	0
A6 Cash Funds last year end	0	0	0	0	0
Cash Funds this year end	6,941	27,435		34,376	0

Section B: Statement of assets and liabilities at the end of the period (to the nearest GBP)

	Unrestricted Funds	Restricted Funds	Endowment Funds
B1 Cash Funds			
Bank current account	6,941	27,435	0
Total Cash funds	6,941	27,435	0
B2 Other Monetary Assets	0	0	0
B3 Investment Assets	0	0	0
B4 Assets retained for the charity's own use	Fund to which asset belongs	Cost	Current value
GPU Workstation	Unrestricted	3,602	3,000
B5 Liabilities	0	0	0

Notes:

The financial statements have been prepared in accordance with the CC16a template using the Receipts and Payments basis

The fixed assets retained for the Ersilia Open Source Initiative use is the computer workstation with a GPU unit, purchased in August, 2021.

Signed and approved by the Charity Trustees:



Miquel Duran-Frigola

04 / 29 / 2022



Edoardo Gaude

04 / 29 / 2022



Akash Rungta

05 / 02 / 2022

Ersilia Open Source Initiative

Independent Examiner's Report to the Trustee of Ersilia Open Source Initiative

Period from 10 November 202 to 30 November 2021

I report to the trustee on my examination of the financial statements of Ersilia Open Source Initiative ('the charity') for the period ended 30 November 2021.

Responsibilities and basis of report

As the trustee of the charity you are responsible for the preparation of the financial statements in accordance with the requirements of the Charities Act 2011 ('the Act').

I report in respect of my examination of the charity's financial statements carried out under section 145 of the 2011 Act and in carrying out my examination I have followed all the applicable Directions given by the Charity Commission under section 145(5)(b) of the Act.

Independent examiner's statement

I have completed my examination. I confirm that no material matters have come to my attention in connection with the examination giving me cause to believe that in any material respect:

1. accounting records were not kept in respect of the charity as required by section 130 of the Act; or
2. the financial statements do not accord with those records; or
3. the financial statements do not comply with the applicable requirements concerning the form and content of accounts set out in the Charities (Accounts and Reports) Regulations 2008 other than any requirement that the accounts give a 'true and fair' view which is not a matter considered as part of an independent examination.

I have no concerns and have come across no other matters in connection with the examination to which attention should be drawn in this report in order to enable a proper understanding of the accounts to be reached.



Shane Tharby FCA
Independent Examiner

For and on behalf of
Streets Chartered Accountants
3 Wellbrook Court
Girton
Cambridge
CB3 0NA

3 May 2022
