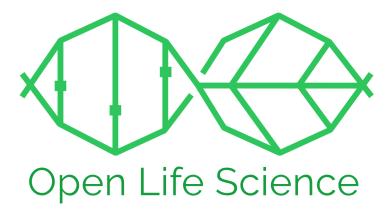# Submission by OLS team for the the Skills Policy Awards 2023/2024

Open Life Science

Funded by The Alan Turing Institute - Skills Team, February 2023

Announcement: https://www.turing.ac.uk/skills-policy-awards-20232024

Call document:
https://www.turing.ac.uk/sites/default/files/2022-11/skillspolicyawards2023_calldocument_0.pdf

Section 1 - Personal details

Please enter your full name.

Yo Yehudi

Co-applicants

Malvika Sharan

Finance contact

Emmy Tsang

Reviewers

Mayya Sundukova

Paz Bernaldo

Section 2 - Relevant experience

Please summarise your relevant data skills/data science education background and/or active research interest in the field. (200 words)

I am an experienced data science professional with over 20 years of experience as a computer programmer and research engineer, and 7 years of working closely with open data communities while holding leadership and advisory roles in related domains.

Currently, I am the Executive Director of Open Life Science (OLS), an organisation dedicated to growing open science ambassadors who create open equitable communities in their own scientific and local domains. OLS offers a well-recognised 16-week training and mentoring program, enabling researchers to develop and lead research projects and become culture-change drivers.

My research and career goals combine grassroots activism with policy-level engagement - my doctoral research covered open source community sustainability, and barriers to open pathogen data sharing and access. The latter was presented at the Data For Policy conference. I also engaged heavily with the open source scientific community and funders, during my time as Open Source Technology Lead at the Wellcome Trust's Data for Science and Health team.

Before this, I was a Research Software Engineer at the University of Cambridge, and in industry and academic spinouts in various software engineering/data science roles. I also taught data science and coding for CodeFirst: Girls and other open science initiatives.

Please provide an example where your expertise helped you achieve impact in the data skills/data science education landscape.(250 words) *Please respond to this question using the STAR method, referenced below.*

- *Situation: Describe the situation and when it took place.*

- *Task: Explain the task and what the goal was.*

- *Action: Provide details about the action you took to attain this.*

- *Result: Conclude with the result of your action.*

Situation: As a data science trainer and community builder, I have noted that most early career researchers from minority groups, low-income backgrounds and the Global South contexts struggled to engage with open source projects after their training, missing out on future opportunities. Reasons included systemic barriers such as disability, discrimination/exclusion from tech spaces, lack of a diverse relatable workforce, limited time for volunteer/free work and lack of personalised support for newcomers to data science.

Task: Since 2019, through Open Life Science, I set out to create a collaborative and inclusive open science training and mentoring infrastructure to retain engagement from diverse learners, connect them with new opportunities (education, employment, career) and collaborate with the international data science network through both grassroots and organisational efforts.

Action: As a founder and Executive Director of OLS, I oversee the operations of the Open Life Science organisation, centring on equity, diversity and inclusion and accessibility (EDIA). I ensure that OLS reaches an international audience and provides support infrastructure: "no-quibble" microgrants; honoraria and stipends for open science work; leadership roles through collaborative grant proposals and fellowships; and integrating considerations for disability often deemed "too difficult" to implement.

Result: Through six cohorts, OLS has engaged >400 participants from 6 continents who launched 174 projects across low-, middle-, and high-income countries. Through open science training, upskilling researchers and retaining this diverse community of professionals, OLS has provided a blueprint for what support infrastructure is needed to widen participation in data science.

Section 3 - Project details

Please provide a summary of your personal benefit for receiving an award and what you hope to gain in terms of professional development. (300 words) *Your response can explain why you are ideal for the award, your motivation for applying and scope for learning/development.*

Under my leadership, OLS has successfully worked towards its mission to make open research and data science practical, achievable, and equitable for the global research

community. Our flagship training and mentoring program has successfully built a diverse workforce who apply open science and modern data skills to build technical projects that benefit their communities. Along with my team, I have personally supported many of them to develop proposals and applications to receive funding, contracts, fellowships, internships, employment, awards and professional collaborations internationally, helping them establish themselves as stakeholders in tech and open science spaces.

To further amplify the impact of my work, I see an opportunity to engage effectively with policymakers. As OLS' executive director, I seek to use the Skills Policy Award to integrate policy literacy into the OLS' curriculum as well as lead the OLS team to transfer insights from their work to policy space. This work will directly translate into helping grow a new generation of open scientists who can effectively engage with policy and policymakers, voicing interests of stakeholders that are often underrepresented, marginalised, or unable to gain a platform to be heard.

Part of this budget is dedicated to supporting OLS Resident Fellows - diverse changemakers who work with me and the OLS team on open science goals closely aligned with UNESCO open science recommendations and AI Council Roadmap. They act as multipliers, conducting research and sharing their knowledge while establishing themselves in leadership roles. As experts, these Fellows will broaden my team and contribute to research (see methodology for details) helping produce all outcomes -- ensuring diverse perspectives and accountability throughout the research.

Finally, I will benefit from networking with other awardees doing similar work or have collaborators in their network who might be interested in becoming expert consultants for the planned research.

Please select which of the four thematic areas your proposal aligns best with.

Widening Participation

Please describe how your proposal aligns with one or more of the four data skills thematic areas. (200 words)

Diversity amongst staff and students in higher education is [demonstrably unequal](#), both in gender balance and ethnic balance. Similarly, open source software developers self-report to be over 90% male and <[15% self-report as BME](#). Numerous national and international studies highlight that women, certain ethnic minorities, people with disabilities and those from disadvantaged socioeconomic backgrounds [are underrepresented in education, training and employment related to STEM](#). With growing applications of data science and AI technology in decision making and automation, as per [ONS analysis](#), women and young people are most at risk of unemployment. It is unsurprising that lack of diversity among tech workforce directly reflects on the technology and continues to disadvantage populations from marginalised communities who already experience systemic barriers to participation.

OLS aims to address this challenge by diversifying skills for the next generation of diverse workforce in areas utilising open data practices, providing them with the opportunity to take new career directions and gain funding through collaboration, leadership and entrepreneurship.

OLS' goals and operations directly align with the theme of 'widening participation'. OLS embodies principles of EDIA and is deeply committed to achieving AI and data literacy for everyone, operationalising the [AI Council Roadmap](#) recommendations mentioned under this theme.

Please identify a gap/challenge in the data skills landscape and clarify how the proposal seeks to address it/them. (200 words) 198/200

The open source software community has significant potential to strengthen data science, allowing people to build upon each other's work and use standardised tools and data formats, enabling ethical technology through data science methods (AI/ML models) that are open, accessible, and beneficial for the broader society. However, it is extremely important to upskill the next generation of diverse tech professionals by equipping them with data skills aligned with the principles of open science.

OLS plays an important role in the open data science movement by making open data science best practices accessible and creating effective support structures for individuals, projects and communities from both majority and underrepresented groups. People from technical and non-technical backgrounds develop skills in community-oriented data science and open leadership.

We see a gap in the policy landscape that operationalises the values exemplified by OLS – centring EDIA principles in higher education and data science, and strengthening organisations' infrastructure to provide diversified career and funding opportunities. This proposal will enable OLS to engage with the policy ecosystem – proposing open science projects on policy interventions, transferring knowledge from data science communities to inform policies and upskilling the next generation of a diverse workforce in effectively engaging with policymakers.

Please list the main objectives of the proposed research in order of priority. (200 words)

1. Integrating policy skills and resources in OLS: Gaining relevant skills from the policy experts and resources offered in the Skills Policy programme, which will expand my skill set and directly impact the work of the OLS team and our capacity-building efforts.

2. Conducting research and community consultations: Landscape analysis to identify national policies and connecting with stakeholders from organisations in the UK and internationally for expert interviews and consultations relevant to the Skills Policy Awards theme "Widening Participation".

3. Publishing a policy briefing note to broader access: Drawing insights from previously listed activities, I will lead the development of a policy briefing note that will comprehensively describe support systems and infrastructures such as funding, career pathways, incentive structure, accessibility considerations, coaching and mentoring. This document will aim to enhance existing institutional, national and international policies to implement these support infrastructures to widen participation in data science and technology from diverse participants with an emphasis on members from marginalised and underrepresented groups.

Please describe your proposed project methodology. (400 words) *Please also include a draft timeline for delivery and details regarding your objectives in relation to project milestones.*

Timeline and methodologies, deliverables (D1, D2, D3) and outputs (O1, O2, O3, O4) described in the next section:
1-3 months (March-May 2023)
- Scope the research with guidance from the expert/mentor in the programme (D2)
- Plan for the expert interview and consultation by designing the questionnaire and expected outcomes of the consultation reports (D1)
- Invite the experts from national and international organisations from different sectors (as listed in this proposal) to become contributors to this proposal - additional stakeholders may be invited from the OLS community and open science network (D1)
- Start the policy landscape analysis research (D2)
3-6 months (June-August 2023)
- Drawing from ethnographic methodologies, conduct interviews and consultations with the "national" experts (D1)
- Summarise insights and findings from the expert consultation (O1)
- Share a draft of the policy landscape analysis structure with the OLS community for feedback and contributions (O3)
- Start scoping the policy briefing note drawing insights from the work carried out for deliverables 1 and 2 (D3)
6-9 months (September-November 2023)
- Repeat the interviews and consultations process to gather inputs from the "international" experts (D1)
- Summarise insights and findings from the expert consultation in report that combines insights from the national report published in the previous quarter (O2)
- Submit the landscape analysis research report as pre-print and invite peer review (O3)
- Develop the first draft of the policy briefing notes (Led by me, contributions from the two resident fellows on this project and the OLS team) (O4)
9-12 months (December 2023-February 2024)
- Finalise the policy briefing note (O4)
- Present the outcomes from this project at national and international platforms and conferences

- Provide a roadmap and the next stages of engaging organisations nationally and internationally with the resources and research outcomes from this project

Personnel and resources involved:
1. Executive Director: Yo Yehudi (10% FTE)
2. OLS Director 2 (to be confirmed) (5% FTE)
3. Resident Fellow 1 - Mayya Sundukova (25% FTE)
4. Resident Fellow 2 - Flavio Azevedo (FORRT) (20% FTE)
5. The OLS community will be involved at various stages of this research. Members will be invited to contribute to the development of the resources as co-authors, reviewers, expert consultants or in other roles.
6. National stakeholders identified from UKRI, BBSRC, Software Sustainability Institute, ONS, Society of RSE, OpenAIRE and ELIXIR-UK (O1).
7. International stakeholders identified from UNESCO, Wellcome Trust, NASA-TOPS, European Commission, Pollicy (Uganda), MBOA-Lab(Cameroon), RSE-Asia Association (O2).
8. Other Skills Policy Awarded projects will be invited to collaborate on overlapping areas of interest.

Please outline expected project deliverables and outputs. (250 words) words

Your response can be a list and should refer to specific times and project milestones.

Deliverable (D1): Expert consultation reports - Interview and consultation with policy experts from national and international organisations to consolidate data, evidence and case studies on creating successful support systems.

Output (O1): The report from national consolidation will be shared to engage local authorities and organisations (UKRI, BBSRC, Software Sustainability Institute, ONS, Society of RSE, OpenAIRE and ELIXIR-UK).

Output (O2): The report from international consolidation will engage stakeholders from UNESCO, Wellcome Trust, NASA-TOPS, European Commission and stakeholders from the Global South organisations. This report will be designed to strengthen governmental policy efforts, such as the White House Year of Open Science in the USA: https://open.science.gov/ and the NWO's Open Science funding policies: https://www.nwo.nl/en/calls/open-competition-domain-science-m-0.

Deliverable (D2): Landscape analysis research report - Research will be carried out to document policy and support systems deployed by the various organisations, followed by their impact on widening participation in data science. The main thematic focus will remain in the OLS' areas of work as described in this proposal.

Output (O3): An academic report summarising the research from this deliverable.

Deliverable (D3): Policy Briefing note - Building on previous outputs, a briefing note will provide comprehensive recommendations and a roadmap for creating funding routes and accessible career pathways, alongside describing requirements for mentoring and upskilling in open data practices.

Output (O4): A briefing note aimed at policymakers and existing support infrastructure in organisations in the UK and internationally.

All resources will be presented at international conferences by the Resident Fellows, as well as shared with relevant organisations.

Please provide a statement to demonstrate that the resources requested are appropriate for the research proposed. (200 words)

Please explain how your project is deliverable with respect to your time commitment, project duration and any additional costing, if applicable. All items requested in the proposal must be justified to the best of your ability.

This staff time is the main requirement for delivering the project as described in this proposal. I will lead this project (10% FTE) ensuring the timely delivery of the project as described in this proposal, with additional advisory support from a second OLS director (5% FTE). Two research fellows (20% and 25% FTE) will contribute to conducting the research, and draft reports and policy briefing notes with the involvement of the OLS team.

I am an experienced researcher with an extensive track record of publishing scientific articles in both quantitative fields such as data science, as well as qualitative methods such as interviewing and thematic analysis. Resident Fellows bring research expertise from working in open research, policy for mental health and political science.

The team involved in the project will be designing and conducting interviews with national and international experts to consolidate feedback to help inform the briefing notes (D4).

Additional costs will be required to pay hourly for admin support for interview transcription.