

Skills Policy Awards 2023/2024

Title: Professionalising traditional and infrastructure research roles in data science

Affiliation:

The Alan Turing Institute

Applicants:

Primary applicant: Emma Karoune, Senior Community Manager

Co-applicant: Malvika Sharan, Senior Researcher - Open Research

Contributors and reviewers

Kirstie Whitaker, TPS Programme Director

Arielle Bennett, TPS Programme Manager

Section 2 - Relevant experience

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

I am a trained researcher with a doctoral degree in Archaeological Science and a PGCE in Secondary Science Education. I have a range of experience in both applying data skills in my own research and teaching data science more broadly to other researchers. As a Senior Community Manager within the Tools, Practices and Systems (TPS) Research programme, I am a key member of The Turing Way (TTW) project, a community-led educational project for open, reproducible and ethical data science. As a community management professional, I have extensive knowledge and experience in building and engaging diverse and inclusive communities in data science. Currently, I lead a team of community managers in Health Research, which is part of the Turing-wide community management team led by Malvika Sharan. Additionally, I facilitate the development of many TTW chapters including a catalogue of Research Infrastructure roles in TTW to recognise and professionalise modern data science roles, see:

<https://the-turing-way.netlify.app/collaboration/research-infrastructure-roles.html>.

Beyond the Turing, I hold leadership and advisory roles in international initiatives including Open Life Science, Software Sustainability Institute, ELIXIR-UK and the Open Phytoliths Community that are shaping the landscape of open science policies.

Please provide an example where your expertise helped you achieve impact in the data skills/data science education landscape.

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: During the COVID pandemic, the Turing-RSS Health Data Lab was tasked to help the UK Health Security Agency (UKHSA) by creating modelling data to address policy-relevant questions. However, the organisations hadn't collaborated previously, requiring the rapid establishment of effective ways of working.

Task: To ensure effective knowledge and skills transfer between Lab researchers and data science professionals in UKHSA, my goals were to increase the two-way sharing of ongoing research and expertise, make collaboration more agile and facilitate rapid completion of high-quality reproducible research.

Action: I organised knowledge-sharing opportunities in Lab and UKHSA meetings; promoted events through multiple communication channels; managed project documentation; shared outcomes from the project openly; and connected with other organisations and individuals linking them with each other's research.

Result: The Lab was incredibly successful in producing high-quality reproducible research and producing timely analytics for the UK government throughout the pandemic. In less than two years we completed 7 projects and published 3+ pre-prints and 6 articles in high-impact journals. The senior leaders used the collaborative opportunities to give feedback on ongoing research by UKHSA teams. Through new collaborations, early-career researchers worked with senior leaders and implemented a predictive model developed by the Lab in new contexts.

Section 3 - Project details

Please provide a summary of your personal benefit from 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.*

I am well positioned to deliver on the 'professionalisation' theme of the Skills Policy Awards. I am keen to standardise data skills and role descriptors collaboratively. I have a proven track record of publishing peer-reviewed articles, contributing to open science community policies and making workshop-led reports, such as the [Parliamentary Inquiry to Research Integrity](#) and [UNESCO's Global Call for Best Practices in Open Science](#). Beyond clearly describing skill sets in the data science and AI landscape, my knowledge and experience will amplify the impact of this work, contributing to the development of the future data science workforce.

This award gives me the opportunity to extend my knowledge and experience further into the policy environment. As an established open science expert, I am particularly motivated to work with policy experts to guide this research and my understanding of connecting best practices to policies.

I am keen to expand my professional network by working more closely with the Turing Skills Team and the other awardees as well as benefit from the networking opportunities gained through this project's national and international workshops.

The publishing of a peer-reviewed article and policy note will add to my repertoire of research outputs as well as contribute to the outputs and influence in this sphere of the TPS programme at the Turing. As a TPS researcher, I am committed to operationalising the recommendations in the AI Council Roadmap. TPS team members hold specialised roles that advocate for and guide the adoption of open, reproducible, inclusive and ethical practices in data science and AI research through interdisciplinary collaboration. As a programme we see a gap in the policy landscape to standardise this approach for enabling high-quality ethical research by professionalising dedicated roles at institutional and national levels. This award will enable a significant step forward in this respect.

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

[Professionalisation](#)

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

I hold a senior position within TPS, which represents an interdisciplinary open source infrastructure accessible to all, and that empowers a global, decentralised network of people who connect data with domain experts. We embed equity, diversity and inclusion (EDI) at the very core of our work. One of the missions of TPS is to advance open leadership by professionalising expert roles beyond traditional research scientist roles. To achieve this mission, I have helped to establish a community management team through significant investment from the Turing, namely ASG, RSS-LAB, DECOVID, Turing-Roche Partnership, AIM-RSF, Digital Twins and Data-Centric Engineering. Over the last year, I have worked to successfully position Community Managers as highly visible and professionally recognised roles, exemplified by the community-driven impact achieved in the projects with the support of our team.

This emphasis on supporting data science research through interdisciplinary collaboration, upskilling researchers in data skills and the broader aim of this project of defining traditional and modern data science roles clearly aligns with the professionalisation theme of this award. We are also strongly aligned with the widening participation theme due to our commitment to building diverse and inclusive teams and placing great importance on ethical and EDI-led research.

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

The National Audit Office report on 'Challenges in using data across government' highlights the current gap in data skills at several levels including storage, management, architecture, planning and governance. The majority of these skills are essential for traditional data science research roles, and even more important for professionalising modern roles such as research engineers, data stewards, community managers, research application managers and more.

This skills gap is compounded by a lack of clarity in definitions for data science roles and their skill requirements. The National Data Strategy states that there is no widely agreed definition of data skills and the role descriptors are used inconsistently across different institutions. Therefore, to enable the upskilling of the current workforce, develop more national and international consistency in hiring practices for these roles, and identify the skills needed by the next generation of data professionals, we need to close this knowledge gap by developing clear definitions of roles in data science and the skills needed to perform these roles.

This project seeks to lead in this area in collaboration with experts and diverse stakeholders curating both traditional and modern data science roles, skills and their professionalisation which will be communicated broadly to inform policies.

Please list the main objectives of the proposed research in order of priority.

1. Curating and centralising definitions, role descriptors and skill lists for traditional and modern data science roles in collaboration with the stakeholders from national and international institutions -- for use in policy, hiring and career development.
 2. Publishing a position paper focused on the newly established Community Management team in TPS with comparison to other data science roles professionalised at the Turing.
-

This will serve as a prototype for creating position papers on other modern and lesser established data science roles.

3. Gaining relevant skills from the policy experts, networking with other awardees and collaborators, and utilising resources enabled by the Skills Policy Awards in developing high-quality outputs that strengthen policies for the professionalisation of modern data science roles alongside the traditional roles in the UK.

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

Timeline and methodologies:

1-3 months (March-May 2023)

- Scope project in consultation with the mentor to capture resources and descriptors for both traditional and novel/modern data science roles (D1-3).
- Plan national and international workshops to gather content and community feedback (D1).
- Draft position paper on Community Management alongside comparing to all data science roles supported at the Turing (D2).

3-6 months (June-August 2023)

- Run national and international workshops to achieve the following outputs (D1):
 - Curating traditional and modern data science roles recognised at different institutions in the UK.
 - Centralising job descriptors/role definitions and skills requirements.
 - Inviting community review and feedback on the resource by adding synonyms for roles and career pathways.
- Publish workshop summary (O1).
- Publish the position paper as pre-print and invite reviews (D2).

6-9 months (September-November 2023)

- Write a policy briefing note drawing from the workshops and the position paper for the broader uptake of centralised role descriptors and associated details (D3).
- Edit and finalise the position paper and submit it to a peer-reviewed journal (O2).

9-12 months (December 2023-February 2024)

- Get feedback, and reviews and publish the policy briefing notes (O3).
- Present findings from the research at international conferences and platforms (O3).

Personnel and projects involved from the Turing:

1. Primary applicant: Emma Karoune (30% FTE - March 2023 - February 2024)
 2. Co-applicant: Malvika Sharan (10% FTE - April 2023 - February 2024)
 3. Contributor: Arron Lacey (April 2023 - February 2024) -- Contributor to position paper
 4. All TPS researchers will engage with the project as contributors towards wider data science roles at Turing and the community management team will support the finalisation of the position paper.
 5. The Turing Way project and The Turing Way practitioners hub will be used as additional avenues to engage collaborators from different sectors (external contributors, university, industry, government and non-profit organisations).
 6. Contributions from the Research Engineering Group and Academic Skills team will be invited in creating job descriptions for roles like research engineers, data scientists and training officers.
-

7. Contributors to the “national workshop” will be invited from the Turing University Network, ELIXIR-UK, Software Sustainability Institute (SSI), Royal Statistical Society (RSS), UKRN and Society of RSE.
8. Contributors to the “international workshop” will be invited from international organisations including universities, research institutes, funders and the non-profit sector.

Please outline the expected project deliverables and outputs.

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

Deliverable 1 (D1): Two workshops with national and international stakeholders to scope the data science roles landscape

Hybrid workshop with established national partners including Turing University network, informal partners (SSI, RSS, UKRN, ELIXIR-UK), and relevant funders (UKRI, BBSRC) that support traditional data science roles and have successfully professionalised modern data science roles.

Virtual workshop of international organisations (research institutes, funders and non-profit sector) from established and informal partnerships that host traditional data science roles and have successfully professionalised modern data science roles.

Output (O1): A written summary of workshop outputs ([link to example of TTW bookdash summary](#)), - accessible by CC-BY license.

Deliverable 2 (D2): Position paper on community management role professionalised at the Turing

The paper will highlight the newly professionalised TPS-supported Community Managers and contrast this role with the traditional data science and other modern data science roles at Turing.

Output (O2): An academic paper published as a pre-print for open access and submitted for publication in a peer-reviewed journal.

Deliverable 3 (D3): A policy briefing note

Building on deliverables 1 and 2, a policy briefing note will be created for organisations leading collaborative research and data science. This document will aim at standardising descriptors for data science roles in organisations/sectors in the UK and internationally where specific data science roles are yet to be professionalised.

Output (O3): A comprehensive document summarising all findings from this research published under a permissive license. Additionally, these resources will be presented at international conferences/platforms and shared with all relevant organisations.

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

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.

I am an experienced senior community management professional with an extensive track record of publishing scientific articles, and organising workshops with national and international organisations to consolidate feedback on impactful projects and reports.

I will lead this project (30% FTE) ensuring the timely delivery of the project as described in this proposal with the support of the co-applicant (10% FTE). This dedicated staff time is the main requirement for delivering engagement activities and publication of the position paper, summary of workshops and a policy briefing note as described in this proposal.

Additional costs will be required to run two collaborative workshops with national and international contributors to the policy briefing note and an additional workshop to initiate a review of the position paper from TPS stakeholders. The workshops will be hosted as hybrid meetings with in-person participation at the Turing and international participants joining virtually. I would like to provide travel costs for approximately 20 national participants, including catering and accessibility costs as required by participants.

For the publications, I would like to pay for professionally produced illustrations to visually convey the summary of the skill sets and descriptors for each role and how they overlap.
