



Professionalising Data Science Roles Workshop Summary

Introduction

The 'Professionalising Data Science Roles Workshop' was held on 2nd October 2023 at The Alan Turing Institute. This workshop was organised to bring together UK data science organisations from diverse sectors to feed into the project outputs for the Skill Policy Award Project on '[Professionalising traditional and infrastructure research roles in data science](#)' being run by Dr Emma Karoune and Dr Malvika Sharan.

The event had 48 attendees of which 12 attended online. The attendees came from 27 data science related UK organisations. They represented 11 different groups of research roles including community managers, research software engineers (RSE), data scientists, project managers, policy advisors/managers, research application managers, researchers/senior researchers, data wranglers, heads of programmes or teams, skills and standards, and funders.

The day began with introductions on the [Alliance for Data Science Professionals Standards \(AfDSP\)](#) by Professor Rachel Hilliam in which she gave an overview of the development of the AfDSP standards and recent updates on the progress of accreditation. Then, we heard a talk on the wider research project by Emma Karoune. Emma spoke about the challenges leading to skills gaps in data science, which included research quality, research sustainability, interdisciplinarity, and equality, diversity, inclusion and accessibility (EDIA). For more details on the introduction, please see the [workshop slides](#) and [workshop handout](#).

Emma also presented an example of the further development of a skills framework for a specialist role: Research Community Managers. This role was mapped to the AfDSP standards and details of this example can be found on page 4 and 5 of the workshop handout.

Breakout discussion sessions

Attendees were put into five mixed groups by the organisers that included a mix of roles and types of organisations. They stayed in these groups for both breakout discussion sessions.

The morning breakout discussion session focused on skills mapping for specialist roles and questions given to groups were:

- Discuss and identify core skill areas for specialist roles in your teams



- What areas of the AfDSP standards need further defining for specialist roles in your organisation?
- Are any skill areas missing?

The afternoon breakout discussion session was more flexible with four suggested topics:

- Topic 1: Career pathways in different sectors
 - Example from academic settings are: Junior -> Senior -> Principal Investigator -> Team Director/Leader. What does this look like in other sectors for specialist roles?
 - How can we improve the ability to move between sectors in data science?
- Topic 2: What roles (from the previous discussions) are still missing?
 - Advocating for those not in the room and continuing with the previous discussions
 - Where do we want to upskill our workforce to fill gaps in the ecosystem?
- Topic 3: Optional: Defining skill areas for specialist roles in more detail (like for RCM in the appendix Table 1.3)
- Topic 4: Optional: Future skill needs – what other job opportunities do you think will occur in the future (in 1, 3, 5, 10 years)?

Below is a summary of the breakout discussion sessions based on themes that arose.

General reflections on data scientist as a role

Many groups reflected that 'Data scientist' is not a single role. But rather an umbrella term for many roles that work in data science. Often organisations divide data scientists into more specific roles based on skills and other factors such as project team composition, research or output focus and organisational need or focus. For example, industry attendees reflected that they have very specialised teams with relatively narrow focuses. Particular skill focuses suggested for more focused data scientist teams were statistics, machine learning, deep learning and visualisation.

It was also discussed that there is a need for overlap between data science roles so that members of teams are able to bridge the technology, methodology, domain gaps and skills in interdisciplinary teams.

Most groups reflected on the fact that there are clear differences in the focus of roles even though they may be called the same term. It was mentioned several times that the relative maturity of data science in different domains, sectors and organisations determines the accepted standard practice for roles and therefore the skills needed in roles.

Current changes happening to more established roles

One of the more established roles discussed was Research Software Engineers. It was suggested that this role is now seeing a further divergence of skills into more specialised



roles within this type of role. This theme came up in several of the discussion groups and seems to be happening across different sectors. There was the idea that these diverging roles could be grouped into designing stage versus production, professionalising and market ready stage roles and therefore would require slightly different skill sets.

Members of the Turing [Research Engineering Group](#) spoke about the three defined roles within their teams - Research Data Scientist, Research Software Engineer and Research Computing Engineer. One discussion group also delved into defining a Research Infrastructure Engineer (RIE). They felt this role had a lot of overlap with RSE roles and in some cases you would in fact be called an RSE but be working on infrastructure. A RIE's main tasks were suggested to be production end, maintaining and keeping infrastructure running whereas RSE is working at a software level. RIE technical skills are around hardware, networking, cloud resources and the on premises infrastructure. It is a similar role to that of Systems administrators or Devops in the industry sector.

If the AfDSP standards were to be further developed for the RIE role, they would need the addition of the different technical skills, as mentioned above, but also important skills in area A on risks and security and area E on ethical environment and technical controls such as implementing data safe havens.

Data stewardship was discussed by several groups as a more commonly known research infrastructure role. However, it could have multiple names/synonyms such as data archivist and data librarian. The difference in titles and focuses of the role were thought to be due to differences in domains and sectors. One group mapped this role to the AfDSP standards and felt that much of the skills were covered in areas A, B and C but additional skills/competencies would be needed to fully define this role such as the addition of data management plans and a reference to metadata. Area D is less relevant to this role and area E would be needed and was felt is needed for all data science roles.

Emerging data science specialist roles

Groups discussed a wide range of specialised roles (see appendix 1 for a full list of the roles mentioned in this workshop) and then picked a number of roles to define in more detail in terms of core skill areas and mapping skills to the AfDSP standards. The two most completely mapped examples are below.

Data privacy and compliance expert

Core skills identified are related to AfDSP standards skill area A, particularly evidential requirement 1 - ensuring the protection of personal and sensitive data. Also peripheral skills from the evidential requirements of area C (2) about relationship management particularly open channel communication and area E (2 and 4) about ethical behaviour and reflective practice and ongoing development were thought to be required for this role.

Product owner/manager



Although this role has been around in industry for a long time, it was discussed as an emerging role in the data science domain linked to deployment of data science research outputs such as data, software, tools, and AI models. This role was discussed by more than one group and was seen to have a common growing importance across different sectors (academia, public sector, industry).

Skills highlighted for this role were technical specialist communication (communication at different technical levels), data science knowledge, setting objectives in scrum analysis, responsibility for metrics of project (risk, budget, communicate to decision makers), flagging ethical considerations, multi stakeholder collaboration (communication and engagement with users), operationalising research and sustainability. Other skills listed by groups for this role included:

- Engaging with user community
- Selling concept
- Commercial management signposting
- Facilitating lifecycle of product
- Legal liaising
- Ensuring fit and implementation
- Custodianship/Stewardship

Career pathways for specialised roles

To stimulate discussions, we gave the example of the traditional pathway of an academic as a currently established career pathway. This traditional academic career pathway goes from Junior or early career academic (Research assistant/PHD student/Post-doctoral researcher) to Senior Postdoctoral Associate/Researcher/Fellow to Principal Investigator and then on to Team Director, or Leader or Head of Lab, Group or Department. We wanted to find out what this looks like in academic settings and other sectors for specialist roles.

Many groups started by focusing on barriers to career progression for specialised data science roles. There was much discussion about assessment, promotion and hiring processes in academic roles still being focused on publications. These outputs are often not relevant or needed in many of the specialised roles so this can create barriers moving in and out of these roles if people want to go back to a more traditional academic career path that requires them. It was also pointed out that these specialised roles often are more removed from defining the research questions and being responsible for project publications, therefore, they have other priorities, outputs and outcomes that are important such as enabling others, publishing more diverse outputs (data, software, writing for wider audiences) and also focusing on impact beyond the project.

Following on from this point, it was raised that this means there is an element of risk to being employed in some specialist roles and those in privileged positions might be more likely to take this risk. Risks included low pay for interim positions and fixed term contracts. Some attendees mentioned a lack of role models for some roles and a lack of diversity across the



sector and a lack of understanding by senior management, especially in Universities, as to the career paths and opportunities for these roles.

Many attendees touched on the issue of othering. In this context, othering means not considering or defining specialised roles as research roles or generally part of research teams and often labelling them (in terms of HR processes) as support roles. This practice creates significant and sometimes insurmountable barriers for people in these specialised roles. It was mentioned that this practice is still common in UK Universities and so does not allow people in specialised roles the same benefits as researchers such as access to travel funds for courses and conferences, access to applying for research funding grants, and being considered for authorship of research papers. Not having access to these benefits can significantly hinder career progression for someone in a specialised data science role.

Not being considered as integral to research does happen beyond universities as mentioned by one of the attendees from an independent research organisation. This attendee was from a cultural heritage organisation and reflected that people did specialist tasks at their organisation but they were not formally named or recognised as having a specialist role. They stated that these roles just don't exist currently in their organisation. Contrary to this, several attendees from industry organisations said their organisations have particular pathways for specialised roles meaning that progression was not based on taking on managerial responsibility. Specialisation was therefore encouraged at these organisations. However, this was caveated by it often taking a long time for these people to progress upwards and there being a glass ceiling for specialist only positions.

Management responsibilities are still the way to progress in most roles and this is true in most sectors - public, academic, industry. One attendee from a research organisation said that postdocs fellows with management responsibilities needed five years experience minimum to mentor others. Attendees had differing views on what the need for management experience for career progression meant for specialised roles. Some people mentioned that becoming very specialised may mean you do not get the opportunity to manage people and therefore this restricts your career progression. Others felt that you either need to specialise more to become more senior or in fact the opposite, go broader, to progress further. This was probably a reflection on the different organisations that attended our workshop and so the amount of management experience linked to career progression seems to differ by organisation.

Other reflections were that there are currently limited positions in specialised roles and also silos in terms of organisational silos - places where these positions are more common but not able to move to equivalent positions elsewhere - and sector silos (reaching a ceiling in a sector and don't know where to go next). An attendee from a research organisation reflected that sometimes people get siloed into these positions as there is more income generated in the organisation for professional services and so they have more job security but then they get stuck there.



Another barrier that was discussed was the skills needed for progression in these roles. Many groups mentioned that it is currently an issue knowing what skills are needed for these roles (hopefully something this project can start to address) and therefore there is a lack of specific training for these roles. This means that any training that is available can be costly. But it was also mentioned that employers are expecting candidates to have the needed and often large list of advertised skills and that they also seem to be unwilling to train new employees, so won't take on new staff without all the skills. It was mentioned that many people in these roles at the moment have picked up skills along the way and are good at generalising their skills across different domains.

The last theme around barriers to career progression was definition of the pathway. There were discussions around new roles being hard to imagine, the pathway being unknown and the route to progression being unclear. Another reflection was that at present people are falling into these roles rather than recognising it as a pathway. This is linked to the above points of a lack of role models and the skills needed for roles being undefined.

The other questions that were posed for the afternoon discussions were 'How can we solve these barriers' and "How can we improve the ability to move between sectors in data science?"

The need for stable long term contracts was raised by several discussion groups and linked to this is the need for investing in people. A combination of these two points would mean opportunities for skill development and training for people in specialised roles as well as address many other of the barriers mentioned above such as risk, privilege, lack of role models, lack of training, etc. Upskilling would also improve people's ability to move between sectors, particularly if employers were more willing to train employees for these newly emerging roles rather than insist on fully formed individuals.

Another solution to career path barriers that was seen as important is the organisational recognition that data science, and roles diverse in data science, are valuable and now an integral part of any (research) organisation.

There was also a reflection by one discussion group that our current moment in data science means that people in data science particularly those in junior positions have different skills to the established data scientists or those now established in specialised roles. This therefore means they probably require a different pathway.

There was also a discussion around how concretely we need to define these roles and accreditation - this is given a separate section below.



More emerging roles and future skills needs

There was also much discussion in the afternoon session of the future skills needed in data science because it is still very much an evolving landscape.

Different groups touched on the huge volume of existing data and that this is only going to increase therefore going forward there would be an increased demand for skills in data management and stewardship to cope with this data. There were also discussions about sustainable data archiving in terms of reducing environmental impact and the cost of storage and computing. Therefore, sustainable archiving data specialists may emerge or these skills encompassed into the already emerging data stewardship roles.

Skills in data reuse were also discussed as a future need - people needed to advocate and enable reuse of datasets, bridging domains and specific skills in interoperability of data. This again touches on the need for deployment and collaboration experts such as product managers, data stewards (as already mentioned above) and community managers and this is also linked to points made about the need for skills in navigating regulatory frameworks as data science and AI tools evolve.

There was also much discussion about the constant need for more people with technical skills in areas such as analytics, digital twins, mathematical models, and large language models. Many attendees touched on the need to upskill the existing workforce in technical skills as well as provide training for new/early career employees/trainees. With the growing use of data science and AI, the future workforce across all domains and sectors will need some level of technical data science upskilling.

Ideas of definitions, standards and accreditation for specialist roles in data science

Sparked by Rachel Hilliams introduction to the AfDSP standards, many of the groups discussed the need for more definition of specialised roles and particularly standards and accreditations in data science and for specific roles in data science.

Questions that came out of these discussions were - What are standards for? How do they help us recruit, retain and motivate people? What is the purpose of accreditation and is it needed in data science or for specific data science roles?

There was an overall feeling that standards are useful for defining the skills needed for the overall role of data scientist and they would also be helpful for specialist roles in data science. Generally, more definition of skills would help employers to bring in the right people for the roles but also help to inform people interested in data science roles about what skills they would need for certain roles and positions. Many agreed that greater clarity of the skills needed for specialist roles will help people currently working in data science navigate the extremely varied and fast moving data science jobs landscape. Standards and skills



frameworks to define roles and career pathways in data science more clearly could also be used to inform school children about potential careers in data science.

Other positives for making these roles more concrete are that it improves access to a community by making it easier for individuals in these roles to see others like them. Standards can be used as something to work towards by more junior people and this might help with job retention. It's a positive that standards could be used as a way of implementing continued professional development and a way of incentivising continued training.

Some groups did reflect that there is a less positive side to having concretely defined standards for roles. Although it is important to recognise people and the skills they have, standards could create boundaries around the roles based on required skills. Accreditation of the standards can mean that people who lack one or more of the skills might not be considered for a role - one attendee posed the question - 'is this right when they might be perfect for the role?'.

Contrary to this is that even though regulations can form boundaries, they can make people feel safe and data science is a new field of research that does have some public trust issues. Greater regulation (accreditation) of any kind could therefore be beneficial. A good example that was mentioned in our discussions, and demonstrates this issue of trust well, is in careers such as medicine or engineering where accreditation or chartered status is necessary for safe working and where there is definite need for continued professional development. Is this needed in data science?

There was also a discussion around some people's preference to go into broader roles that they can then self define as they prefer and are comfortable to navigate the uncertainty of this lack of definition. Therefore, some flexibility may be needed to allow these people to flourish.

Next steps

The discussions in this workshop and further interviews of workshop attendees on themes that have emerged during the discussions will be combined with a wider literature review and used in the final publication of this project - a policy briefing note and a shorter post note. These outputs can then be used to advise policymakers and other relevant stakeholders as to the current state of emerging specialist roles in the data science ecosystem and therefore inform future considerations and directions of the skills (training) and career pathway developments needed in this domain.

List of attending organisations (in alphabetical order)

- Accenture
- Ada Lovelace



- Alliance for Data Science Professionals
- Archaeological Data Service
- Biotechnology and Biological Sciences Research Council (BBSRC)
- British Antarctic Survey
- Cardiff University
- Chartered Institute of Archaeologists
- Department of Education
- ELIXIR-UK
- Francis Crick Institute
- Genomic England
- Historic Environment Scotland
- Institute of Mathematics
- National Physical Laboratory
- Newcastle University - Research Software Engineering group
- Newcastle University - National Innovation Centre for Data
- NHS England
- Office for National Statistics
- Open University
- Roche
- Royal Statistical Society
- Software Sustainability Institute
- The Alan Turing Institute
- University of Bradford
- University of Manchester
- University of Southampton

Appendix 1: (Specialist) Data science roles mentioned in this workshop

- Academic lead - HPC
- AI Ethics researcher
- Analytical data scientist
- Bioinformatician
- Business analyst
- Business relationship influencer
- Collections manager
- Community manager x 2
- Corporate governance research ethics manager
- Data accessibility experts
- Data analyst
- Data archivist
- Data curator - FAIR
- Data engineer x 2
- Data ethics advisor
- Data ethicist



- Data librarian
- Database manager
- Data manager
- Data privacy and compliance expert
- Data science advocates
- Data science engineer
- Data science Team Lead
- Data science trainer
- Data sharing expert
- Data steward x 2
- Data steward community manager
- Data training experts
- Data wrangler
- Digital archivist
- Experiment designer
- HPC stakeholders storage trainer
- Lead data steward trainer
- ML engineer
- Modeller
- Partnerships liaison (suggested not RIR from KW)
- Postdoctoral researcher/research fellow (suggested not RIR from KW)
- Product Manager/owner
- Programme manager (suggested not RIR from KW)
- Research application manager x 2
- Research community manager
- Research computing manager
- Real world data scientist
- Relationship manager
- Research and Training funder
- Research infrastructure engineer x 3
- Research sustainability lead
- Spatial data analyst
- Standards manager (sustainability and best practices)
- Statistician
- Scientific data coordinator
- Technical writer (suggested not RIR from KW)
- Training facilitator
- Trainer

Acknowledgements

Thank you to all the attendees of the workshop for their contributions on the day and also to Rachel Hilliam for her presentation. Thank you also to Turing colleagues that facilitated discussion sessions.