Software Sustainability Institute

Software and skills for research computing in the UK

Authors

Michelle Barker, Elena Breitmoser, Philippa Broadbent, Neil Chue Hong, Simon Hettrick, Ioanna Lampaki, Anthony Quinn, and Rebecca Taylor.

November 2023

CONTENTS

Contributor names are listed alphabetically, roles are starred (*) where organisational leadership was taken.

Michelle Barker: Writing—original draft (*); writing—review and editing (*); formal analysis; investigation: interviews (*).

Elena Breitmoser: Project administration (*); data curation; survey; investigation—interviews; writing—review and editing.

Philippa Broadbent: Software (*); visualisation; formal analysis—survey; writing—original draft.

Neil Chue Hong: Conceptualisation (*); funding acquisition (*); writing—review and editing; formal analysis; methodology; project administration; supervision (*); investigation: survey; investigation—interviews.

Simon Hettrick: Conceptualisation (*); writing—review and editing; formal analysis; validation; supervision; survey.

Ioanna Lampaki: Software (*); methodology (*); writing—original draft; visualisation; formal analysis; data curation; resources; formal analysis—survey (*).

Anthony Quinn: Software; methodology; formal analysis; data curation; resources; investigation—focus groups (*); writing—original draft.

Rebecca Taylor: Conceptualisation (*); methodology; investigation—focus groups; formal analysis; supervision; writing—review and editing.

This study has been funded through the UKRI Digital Research Infrastructure programme, and will contribute to the development of national programmes. It was undertaken between December 2021 and August 2022 by the Software Sustainability Institute, with researchers based at the University of Edinburgh and the University of Southampton, in collaboration with Dr Michelle Barker. The study personnel are detailed in Appendix A.

EXECUTIVE SUMMARY

Software and the people who produce it have revolutionised the way that research is conducted, pervading all aspects of the research lifecycle. These emerging tools and techniques require new skills and, often, new forms of research collaboration that combine a variety of professional capabilities. This report delivers a better understanding of the software and skills required in order for research computing in the UK to respond to the challenges it faces over the next five years across three overlapping levels: people, infrastructure, and policy. It arrives at the following key findings:

People

- RSEs perform an essential role in the research computing ecosystem that must be supported with appropriate career paths.
- Some research computing-adjacent roles, such as Research Librarian and Data Steward, are not sufficiently visible or utilised by researchers engaged in scientific computing.
- Challenges in accessing adequate training may inhibit professional development, particularly for those in earlier career stages. These challenges relate to areas including embedding skills in curricula and recognising that much learning is on the job/self-taught.
- There are still gender inequalities in scientific computing, particularly concerning skills acquisition. More research is needed to obtain a more nuanced understanding of opportunities and challenges, and to enable similar analysis of other under-represented groups.

Infrastructure

- 97% of survey participants see software as important as their own research, with 85% citing it as essential.
- The large disparity in the number of researchers using larger scale computing infrastructure compared with those using university-level resources needs to be addressed.
- 27% of participants believe their research will be limited by access to computational/data infrastructure in the next year, rising to 35% of participants at senior career stages.

Policy

- Organisational processes that support RSEs and research software initiatives are beneficial at the institutional level, and would also be valuable at the national level.
- Funding for research software and its personnel is limited, with the most common research software funding mechanism being standard research grants. There is some perception that including research software development costs in funding applications is detrimental.

The report concludes with recommendations that can assist the UK in maintaining its world-leading position with regards to research software.

People recommendations

1. All parts of the research community must understand that a wide variety of roles support the research ecosystem. Better recognition and appropriate career pathways are vital to ensuring that there are enough skilled people in the UK to support research. This is the joint responsibility of a number of key stakeholders, including Principal Investigators, universities, funders, industry, and government.

   1.1. Enable detailed analysis of how to professionalise RSE roles, building on existing initiatives such as the international RSE survey. This analysis should profile RSE as a career through case studies, careers fairs, etc., and it should occur in tandem with legitimisation of other roles such as Data Stewards, etc.

   1.2. Facilitate collaboration between government, funders, and employers (particularly universities and Principal Investigators, and potentially industry) to create national policies aimed at improving standards of employment for RSEs in academia around length of contract, pay standards, mobility between academia and industry, and professional development, by means of standard assessment criteria. Investigate whether contract terms such as a minimum of two years (or over) for RSEs improve job satisfaction and retention.

   1.3. Support further research by UKRI and relevant professional bodies (RLUK, CILIP, etc) to identify why some research computing-adjacent roles, such as Research Librarian and Data Steward, are not as visible or utilised by researchers engaged in scientific computing. Consider whether this suggests changes are required to make these roles more relevant to researchers whose work is computationally intensive.

   1.4. Facilitate collaboration between government and universities to create a training strategy that empowers all research roles and enables them to take the time to learn the skills needed for modern research. It must build a pathway through data and software training that starts at the undergraduate level, builds through Masters and PhD study, and continues during employment. The strategy should:

      - Incentivise professional development opportunities for early career researchers.
      - Identify and address any gender disparities among the people who receive training.
      - Identify and address disciplinary differences, such as the lower software development undertaken by MRC and BBSRC researchers.
      - Cultivate the most in-demand skills through teaching and training, such as honours/Masters courses.

   1.5. Enable regular information collection exercises to be undertaken to allow UKRI to track current training provision and highlight skills that will be vital in the future.

   1.6. Conduct a large-scale study of gender inequalities in scientific computing alongside an analysis of EDI initiatives, with the goal of producing a set of recommendations for funders on reducing inequalities. This could assist in addressing challenges and frame recommendations on how to use existing best practice within the UK and internationally to improve EDI outcomes. Leaders in this field, such as Advance HE and the Royal Society, should be brought in as partners in this study.

Infrastructure recommendations:

2. Develop a national roadmap for a coordinated access to, and training in, all levels of UK research computing infrastructure, involving universities, funders, industry, and government. This should include a focus on enabling personnel to transition between levels, both in their use of research computing and in their confidence/competency with the software.

   2.1. Provide unified guidance across UKRI Research Councils that support recognition of software as critical digital research infrastructure, and implementation of international standards on software citation and FAIR Principles for Research Software.

   2.2. Undertake further analysis to understand what the barriers are to researchers running their experiments on Tier 2 and larger research computing infrastructure. This should include consideration for increasing availability of RSEs to help researchers in bridging this gap, to ensure that a wide range of researchers benefit from investments in exascale, HPC, and cloud computing.

   2.3. Recognise the need to support users of both less advanced and more advanced infrastructure through access to RSEs and training, and encourage international cooperation in order to improve access to larger scale resources and different architectures and technologies.

Policy recommendations:

3. Provide unified funding requirements across UKRI Research Councils that align with international standards, including relevant UNESCO and OECD recommendations, and lay out the following international vision:

   - Research software must be recognised as a key element of research.
   - The development and maintenance of research software must be supported.
   - Research software must be as open and/or as FAIR as other components, so that the research it enables can be trusted and replicated.

   3.1. Incorporate into these unified funding requirements a framework in line with the 10 Simple Rules for Funding Scientific Open Source Software, and include:

   - Specific programmes for maintenance.
   - Encouragement of reuse and/or contribution to existing platforms.
   - A variety of sizes of funding.

   3.2. Continue to lead and/or contribute to international efforts to develop standards and practices that solve challenges faced by the UK research community.

Many of the recommendations contained in this report focus on areas where there are no exemplars available elsewhere to learn from, and thus where the UK has both the privilege and the challenge of being able to break new ground. Other recommendations relate to areas where there are examples of best practice, either locally or internationally, which are not being implemented in a coordinated way in the UK.

This study has been funded through the UKRI Digital Research Infrastructure programme, and will contribute to the development of national programmes. It was undertaken between December 2021 and August 2022 by the Software Sustainability Institute, with researchers based at the University of Edinburgh and the University of Southampton, in collaboration with Dr Michelle Barker. The study personnel are detailed in Appendix A.

LIST OF ACRONYMS

ASCL: Astrophysics Source Code Library

AHRC: Arts and Humanities Research Council

AI: Artificial intelligence

ARDC: Australian Research Data Commons

BAME: Black, Asian, and minority ethnic

BBSRC: Biotechnology and Biological Sciences Research Council

BCS: British Computer Society

CCP: Collaborative Computational Project

CEng: Chartered Engineer

CILIP: Chartered Institute of Library and Information Professionals

CPU: Central processing unit

CSci: Chartered Scientist

CSSI: Cyberinfrastructure for Sustained Scientific Innovation

DARE: Data and Analytics Research Environments

DFG: German Research Foundation

DiRAC: Distributed Research using Advanced Computing

DORA: Declaration on Research Assessment

EDI: Equity, diversity, and inclusion

EOSC: European Open Science Cloud

EPCC: Edinburgh Parallel Computing Centre

EPSRC: Engineering and Physical Sciences Research Council

ESRC: Economic and Social Research Council

FAIR: Findable, Accessible, Interoperable, Reusable

GDPR: General Data Protection Regulation

GIS: Geographic information system

GPU: Graphics processing unit

HESA: Higher Education Statistics Agency

HPC: High performance computing

HTC: High-throughput computing

IMA: Institute of Mathematics and its Applications

INTERSECT: INnovative Training Enabled by a Research Software Engineering Community of Trainers

IoP: Institute of Physics

LERU: League of European Research Universities

ML: Machine learning

MRC: Medical Research Council

NERC: Natural Environment Research Council

NIH: National Institutes for Health

NIHR: National Institute for Health and Care Research

NLeSC: Netherlands eScience Centre

NSF: National Science Foundation

NWO: Dutch Research Council

OECD: Organisation for Economic Co-operation and Development

OSPO: Open Source Program Office

PPTAP: Particle Physics Technology Advisory Panel

REF: Research Excellence Framework

ReSA: Research Software Alliance

RLUK: Research Libraries UK

RSE: Research Software Engineering

RSEs: Research Software Engineers

SSI: Software Sustainability Institute

STFC: Science and Technology Facilities Council

TREs: Trusted research environments

UKRI: UK Research and Innovation

UNESCO: United Nations Educational, Scientific and Cultural Organisation

1. INTRODUCTION

Software has transformed every stage of the research lifecycle. From machine-learning and big data to modelling and simulation, from management of experiments to record-keeping and reproducibility, research is powered by software. Digital technologies have become vital tools and techniques for the modern researcher, and their application requires new skills and interdisciplinary collaborations.

This report lays out the policies and actions required for research computing to continue thriving in the UK. "Research computing" refers to the range of computer resources—including high performance computing (HPC), high-throughput computing (HTC), cloud data storage, and network resources—used to support research.

Research computing is a critical element of modern research in terms of achieving the UK Research and Innovation vision of an outstanding research and innovation system. Analysis of the quantitative impact of digital research techniques and tools shows an average two- to fourfold increase in research impact (Dietrich, 2018).

This report will contribute to the evidence base informing the development of a national research software strategy, providing recommendations on how the UK can maintain its world-leading position with regards to research software. Two UK reports have recently laid the foundations for understanding the requirements for software and skills to support modern research: Large-scale computing: the case for greater UK coordination (Government Office for Science, 2021) and ExCALIBUR Research Software Engineer Knowledge Integration Landscape Review (Parsons et al., 2021). Both strongly recommend investment in software and skills.

This report provides an overview of the UK software and skills landscape along with recommendations on how policies and support for these could be structured. It is based upon an 8-month, mixed methods study that provides additional insight into perceptions of and attitudes to research computing. Section 2 begins with a landscape analysis of international initiatives to contextualise the UK's approach. Section 3 introduces the large-scale study of current software and skills requirements in the UK research community that this project undertook, which involved a survey, group interviews, and stakeholder interviews. Section 4 details the analysis of this study, and section 5 presents key findings and recommendations.

2. LANDSCAPE ANALYSIS

A number of governments have strategies and/or major initiatives to support and recognise research software and its personnel. In addition to the UK, these include Australia, Canada, the European Union, France, Germany, the Netherlands and the US (Barker et al., 2021). This section analyses these efforts in order to understand UK approaches in an international context.

This analysis is undertaken across three overlapping levels: people, infrastructure, and policy. They include the following elements:

People	Infrastructure	Policy
Career paths	Software usage	Enabling environments
Skills and training	Computational infrastructure	Funding
Diversity of personnel		

2.1 PEOPLE

Employment in the research sector is evolving in terms of career paths, skills and training, and diversity. Improving practices around research personnel is increasingly a top priority as a way to improve research outcomes. For example, the G7 Research Compact encourages G7 nations to collaborate on global challenges through actions that focus first on people: "Explore incentives, including enhancements to research assessment that foster recognition and reward collaboration across all disciplines and topics to drive a culture of rapid sharing of knowledge, data, software, code and other research resources" (G7, 2021).

National examples of this focus on research software personnel include Canada's Research Software Current State Assessment, which identifies challenges including:

- The lack of software development skills at all stages of the education and research pipeline needs to be addressed.
- Incentives related to metrics, funding, reward and recognition, and career progression are critical in catalysing engagement with research software.
- The lack of diversity in the research software community suggests we need a more effective approach to equity, diversity, and inclusion (EDI) in the research software context (Digital Research Alliance of Canada, 2021).

Examples of this focus also come from within the research community. For example, A survey of the state of the practice for research software in the US identifies three primary classes of concerns that have prevented research software from achieving maximum impact, the third of which relates to the functioning of the individual and the team. This work examines "issues such as training and education, ensuring appropriate credit for software development, enabling publication pathways for research software, fostering satisfactory and rewarding career paths for people who develop and maintain software, and increasing the participation of under-represented groups in research software engineering" (Carver et al., 2022). This focus is also evident at a disciplinary level: a US National Academies report on Next Generation Earth Systems Science at the National Science Foundation (NSF) publicises the need for research software, RSEs, and the integration of diversity, equity, inclusion, and justice in all aspects of next generation Earth systems science (National Academies of Sciences, Engineering, and Medicine, 2021).

A range of initiatives in the UK are concerned with software skills, retention, and personnel demographics. In terms of research staff in general, the UK Research and Development People and Culture Strategy states that "[p]eople are at the core of research and development, so there is nothing more important than how we attract, develop, and retain enough people ... We need to unleash a new wave of talent: attracting, developing and retaining diverse people with the right skills, working in an environment that nurtures and gets the best out of everyone" (Department for Business, Energy & Industrial Strategy, 2021).

This focus is also applied explicitly to research software personnel. Large-scale computing: the case for greater UK coordination highlights the "strong case for continued public investment in large-scale computing", while noting that "investment in computer hardware alone will not be sufficient ... demands outweighs supply and the workforce lacks the diversity required to produce software that encompasses the needs of society". The report makes recommendations on investment in software and skills, including software engineers, and the need to increase diversity and foster inclusion, noting that diversity in the large-scale computing workforce is understudied, with limited data available (Government Office for Science, 2021).

2.1.1 Career paths

A significant proportion of this work on research software personnel aims to create career paths that both support open science practices and broaden research personnel demographics. This section explores efforts to map the roles needed, and strategies for supporting new professional roles such as RSEs.

International landscape

The need for a greater understanding of research careers is recognised worldwide. Major international initiatives aimed at improving the ways in which researchers are recognised and valued include the Declaration on Research Assessment (DORA) and the Hong Kong Principles for assessing researchers, and many papers call for a rethink of the divide between academic and support/technical staff (Teperek et al., 2022; VSNU et al., 2019). National strategies pursuing this agenda include:

- Australia: the National Research Infrastructure Roadmap notes that the attraction and retention of skilled staff can be challenging, and cites scientific software development as an example of system-wide skills shortages (Department of Education, Skills and Employment, 2022).
- Europe: European Open Science Cloud (EOSC) has convened a Task Force on Infrastructure for Quality Research Software. The third of the three objectives of this task force is to "Increase recognition to software developers and maintainers of research software as a valuable research result, on a par with publications and data, in the open science landscape" (EOSC, 2021a).
- France: the Second French Plan for Open Science includes strategies to better recognise software development in research. These include the "objective to achieve better recognition of software development in career evaluation for researchers and engineers" (Ouvrir la Science, 2021).

This focus on research careers extends to emerging digital research roles, including research software personnel such as RSEs. It is a prevalent issue from a disciplinary perspective as well: one report on RSE in HPC noted that the role of HPC RSEs still lacks attractive career paths in academia, and that the expertise of RSEs is often not acknowledged (Ferenbaugh et al., 2022). The development of the RSE community, which originated in the UK, has been immensely valuable in making research software personnel visible (Akhmerov et al., 2019; Brett et al., 2017; Cohen & Woodbridge, 2020; Hardey & Leng, 2020; Woolston, 2022).

UK profile

The international commitment to improving research careers paths is shared by many UK initiatives, including those focused on artificial intelligence (AI). For example, the UK National AI Strategy identifies a need for enhanced UK capacity in key AI professional skills related to research and innovation, such as data scientists and software engineers, noting that this applies to software personnel more generally and not just specialist research software roles (Office for Artificial Intelligence et al., 2021). The Talent Commission provides a more research-specific example, highlighting the importance of technical staff and skills in academia, including RSEs (MI Talent et al., 2022). The UK Data Research Infrastructure Landscape also highlights the need to support the career structures of individuals creating or engaging with digital research infrastructure (DARE UK Consortium, 2021).

There are also UK initiatives that are research software-focused. Large-scale computing: the case for greater UK coordination highlights the "strong case for continued public investment in large-scale computing" while noting that "investment in computer hardware alone will not be sufficient". Its recommendations include a focus on software engineers, and it highlights the fact that "[i]n academic fields, RSEs do not have a clear career path, progression opportunities are often limited, and the use of temporary contracts is common". The report concludes by urging that "career pathways should be developed to help attract, support, and retain these skill sets" and includes recommendations on how to achieve this (Government Office for Science, 2021).

2.1.2 Skills and training

Digital skills and training are essential for ensuring the development of high-quality and sustainable software to maximise research impacts. It is necessary to identify the issues that need to be addressed in order to ensure that the research community has the skills to maximise the use of research software, and to enable provision of the necessary training.

International landscape

There has been a burgeoning interest in facilitating digital research skills throughout the research community (European Commission & EOSC Executive Board, 2021; OECD, 2020). For example, the first objective of the Strategic Research and Innovation Agenda of EOSC is as follows: "Ensure that open science practices and skills are rewarded and taught, becoming the 'new normal'" (EOSC, 2021b). This emphasis applies to skills for both researchers and for specialist professional roles such as RSE, and there are a range of initiatives that support this:

- Better Scientific Software (BSSw): includes resources for developer productivity and software sustainability.
- CodeRefinery: acts as a hub for Findable, Accessible, Interoperable, and Reusable (FAIR) software practices.
- INnovative Training Enabled by a RSE Community of Trainers (INTERSECT): will create a bootcamp targeting intermediate and advanced students from US institutions.
- National RSE Associations: such as Nordic RSE, which provides a community and professional development network for RSEs.
- The Carpentries (including Software Carpentry and Data Carpentry): teaches basic lab skills for research computing, including coding and data science, to researchers worldwide.

UK profile

A number of UK reports highlight the need for software skills and training. For example, the Talent Commission notes that "[t]he need for computational skills – including enhanced digital design, software engineering and skills in AI and machine learning (ML) – is widespread" (MI Talent et al., 2022). Refining the focus, the ExCALIBUR RSE Knowledge Integration Landscape Review provides a comprehensive review of the skills required by RSEs in HPC and their future training needs. This report argues that investment in software means investment in people, and identifies where major training and skills gaps exist in order to address issues of long-term career development for RSEs in HPC (Parsons et al., 2021). Recent reviews commissioned by individual research funders also have this focus. The Biotechnology and Biological Sciences Research Council (BBSRC) Review of Data Intensive Bioscience recommends that BBSRC should take specific actions to increase the UK capacity in mathematical and computational skills within the biosciences (BBSRC, 2020). One of the five areas of focus for the Economic and Social Research Council (ESRC) Data Infrastructure Strategy is skills and capacity for data use, with the goal of enabling researchers to effectively utilise data in their research for public benefit (ESRC, 2022).

The UK also supports research software skills and training initiatives in UK, which include:

- Engineering and Physical Sciences Research Council (EPSRC): investments include training and activities that support a thriving community of computational scientists who are recognised internationally. Key programmes include the RSE Fellows.
- ESRC National Centre for Research Methods: provides comprehensive training in research methods, including digital methods such as statistical modelling, ML, and data wrangling.
- Natural Environment Research Council (NERC): funds software skills training through its funding call to run advanced training for early career environmental scientists.
- Society of RSE: RSECon is the annual conference for RSEs from all around the world, and includes training sessions in its multi-day programme.
- SSI: skills and training programmes include Research Software Camps, and SSI coordinates the activities of The Carpentries in the UK (SSI, 2021).
- UK Research and Innovation (UKRI): Innovation Scholarships are increasing health and bioscience skills, and the UKRI's large-scale data training programme funding includes a focus on data analysis modelling and coding (UKRI, 2021).

2.1.3 Diversity of personnel

There are many initiatives in the research software community designed to promote greater diversity in terms of age, background, ability, ethnicity, and gender. They recognise the benefits of diversity to both society and research, with studies of EDI in the open source software community (and occasionally with a research software focus) also identifying the following benefits:

- Increasing innovation (Campbell et al., 2013; Liang et al., 2007).
- Increasing sustainability (Benjamin, 2019; Dunbar-Hester, 2020; Vasilescu et al., 2015).
- Decreasing duplication and increasing reuse (Ferreira da Silva et al., 2021).

International landscape

There is a wide range of strategic and operational approaches to addressing diversity and inclusion issues in the research software community. While there is limited data on this topic, available evidence clearly identifies issues. For example, analysis of 2018 international RSE demographics revealed that 73–92% of RSEs were male, and most commonly ranged in age from 25–44 (Philippe, 2018). A survey on research software in the US research community found that only about a third of the respondents felt that their projects did an "excellent" or "good" job of recruiting, retaining, and including in governance participants from under-represented groups, although respondents also indicated that they needed the most help with recruiting, retaining, and promoting diverse participants (Carver et al., 2022).

The ReSA People Roadmap identifies major issues related to people (or personnel) challenges faced by the research software community, finding that most of the 28 research software initiatives studied identified broadening the range of contributors as important to improving scientific and social outcomes (Barker, 2021). International efforts include Vive la différence - RSEs, a hybrid workshop bringing together participants from across the globe to consider how RSE could be reframed to make EDI a central organising principle. A complementary series of public webinars, DiveRSE, aims to support and raise awareness of EDI in the research software community internationally.

Efforts at the national level can be seen in both strategy and implementation. For example, the US-RSE's mission includes the following statement: "We will actively promote, encourage, and improve diversity throughout the broader US RSE Community consistent with our full diversity, equity, and inclusion mission statement. We will ensure we provide an inclusive environment with equitable treatment for all and we will prioritise a programme of diversity, equity, and inclusion activities for our organisation, led by a dedicated team of active community members" (US-RSE, 2021).

UK profile

UK organisations are also increasingly emphasising EDI, with useful resources for the research sector being provided by organisations including Advance HE and the UK Royal Society. The UK Research and Development People and Culture Strategy acknowledges the need for a positive, inclusive, and respectful culture (Department for Business, Energy & Industrial Strategy, 2021), and the first of the principles guiding the UKRI vision for a national digital research infrastructure is to "be driven by the ambition of UKRI's diverse communities" (UKRI, 2022d).

This focus can also be seen in the UK Research Councils. For example, the document EPSRC expectations for equality, diversity and inclusion (EDI) addresses six EDI themes, providing expectations and available resources for each:

1. Develop an approach to embedding EDI in the research lifecycle.
2. Implement good practices in recruitment and/or selection processes to ensure diverse teams.
3. Ensure diversity and inclusivity in all activities such as events, sandpits, and networking.
4. Create an inclusive and accessible environment.
5. Ensure career progression and training for all members of the team.
6. Inclusive research (EPSRC, 2022a).

EPSRC also funded 11 Inclusion Matters projects in 2018 to accelerate culture change with respect to equality, diversity, and inclusion. Nevertheless, research shows that the UK RSE community is currently less diverse than comparable professional communities, and there is a potential benefit to hiring from a wider range of backgrounds as the need for skilled software practitioners in research continues to increase (Chue Hong et al., 2021).

2.2 INFRASTRUCTURE

Global investments in digital research infrastructure and research software continue to increase. The International Research Infrastructure Landscape 2019 notes that almost all large-scale research activities include, or wholly consist of, digital components (Asmi et al., 2019). While research software infrastructure is an element of digital infrastructure, it has only recently come to be recognised as equally important in its own right. For example, the Australian government now stipulates that, "[d]ue to its importance, research software must be considered as research infrastructure itself" (Department of Education, Skills and Employment, 2022).

The UKRI Digital Research Infrastructure strategy is forward-thinking in characterising software as one of the building blocks of the digital research infrastructure system, which is outlined here:

- Large scale computer facilities, including HTC, HPC, and cloud computing.
- Data storage facilities, repositories, stewardship, and security.
- Software and shared code libraries.
- Mechanisms for access, such as networks and user authentication systems.
- People: the users, and the experts who develop and maintain these powerful resources (UKRI, 2022d).

2.2.1 Software usage

Surveys have found that 90% of UK researchers acknowledge software as important for their research (Hettrick et al., 2014), and that 95% of US postdoctoral students use research software (Nangia & Katz, 2017). Yet analysis of software usage is a limited field, partly due to challenges in software sharing and visibility.

International landscape

The role that research software plays in research outcomes is not widely understood, and the software itself is also often hard to pinpoint. For example, Canada's Research Software Current State Assessment identifies as one of its key challenges and opportunities that research software "is not widely disseminated or shared and not readily discoverable, inhibiting research transparency, reproducibility, and verification" (Digital Research Alliance of Canada, 2021). Initiatives such as the EOSC Task Force on Infrastructure for Quality Research Software seek to address this type of issue, with one of EOSC's three objectives being to "[f]oster the development and deployment of tools and services that allow researchers to properly archive, reference, describe with proper metadata, share and reuse research software" (EOSC, 2021a).

The Australian National Research Infrastructure Roadmap acknowledges that "[r]esearch software plays an essential but often invisible and undervalued role in generating, processing and analysing data" (Department of Education, Skills and Employment, 2022). The Research Software Capability in Australia report included data on respondents' (which included nearly 50% of Australian universities) familiarity with the research software generated by their staff, concluding that "[w]hile it is valuable to understand that at least some research organisations have some awareness of their research software outputs (or those that they contribute to), this is not evidence that the Australian research sector audits research software assets in the same way it now does for research data assets" (Barker & Buchhorn, 2022). The Visible Research Software Interest Group is an online forum for people within Australasian research institutions to discuss the visibility of research software in terms of citation, publication, and FAIRification (Martinez, Gustafsson, et al., 2022).

Increasing software visibility is also a priority of the Second French Plan for Open Science. This report's emphasis on promoting source code produced by research resulted in the Open Science Awards for Open Source Research Software, where 10 software applications developed by French teams were rewarded for their contributions to the advancement of scientific knowledge (Ouvrir la Science, 2021). In the US and Europe, Open Source Program Offices (OSPOs) are being introduced by a few universities to increase awareness and improve the management of open source scientific software. This can involve the use of a GitHub enterprise account to improve tracking, metrics, and organisational software assets (Choudhury, 2021).

UK profile

The UK's research portfolio has a strong emphasis on developing, supporting, and sharing research software, with organisations like the SSI and programmes like the Collaborative Computational Projects functioning as international exemplars. Large-scale computing: the case for greater UK coordination recognises that "high-quality software is fundamental to realising the benefits of investments in computing" and recommends that "software development must keep pace with advances in hardware" (Government Office for Science, 2021). This includes the need to help improve software quality and ensure that it is well maintained. Similarly, the UK Data Research Infrastructure Landscape has repeatedly called for new digital infrastructure investments not to 'reinvent the wheel' (DARE UK Consortium, 2021). Enabling software reuse is identified as an important part of this strategy.

BBSRC's Review of Data Intensive Bioscience recommends that BBSRC should take specific actions to "significantly increase its investment in provision of high-quality software and data resources for the research community". In the questionnaire, responses on the biggest issues that need to be addressed over the next five years identified software (new tools) as third most important. Responses to the follow-up question, on what the priority areas of action for BBSRC should be, identified software development as the third most important priority (BBSRC, 2020). A further issue identified in the survey was that the structure of BBSRC funding "does not support a sufficiently broad range of activities required in this area ... a limitation of the current model is that software and resources can often be under-developed from the perspective of the end user" (BBSRC, 2020).

2.2.2 Computational infrastructure

Many countries operate significant digital research infrastructure investment programmes to accelerate innovation and advance researchers' ability to solve societal issues. These include investments in different types of research computing, such as HPC and computational clouds.

International landscape

Significant technology advances such as exascale and quantum computing will enable a higher level of performance in computing that has the potential to rapidly advance research impact. Investments in the physical infrastructure required for exascale computing must be complemented by programmes to develop the code needed to utilise exascale computers. For example, the Centre of Excellence for Exascale in Solid Earth (ChEESE) is a European Commission initiative that has succeeded in preparing simulation codes and applications as geohazard mitigation services that can be used to support the private and public sectors in decision-making (ChEESE, 2022, p. 22).

Quantum computing will also rapidly advance research, and a number of countries already have significant investments. In the US, one of the NSF's three ways to commit to fostering quantum-based research is workforce development (NSF, 2022). China has investments on a similar scale, along with higher education reform that includes the development of a new double first-class network of universities, advanced postgraduate education, training of rare high-value skills, improvement in scientific research, construction of cutting edge science centres, and participation in national laboratory construction (Graps, 2022).

Mapping eResearch Ecosystems: the international situation is intensifying! provides one way to understand infrastructure needs through identification of four types of research projects, as shown in Figure 1:

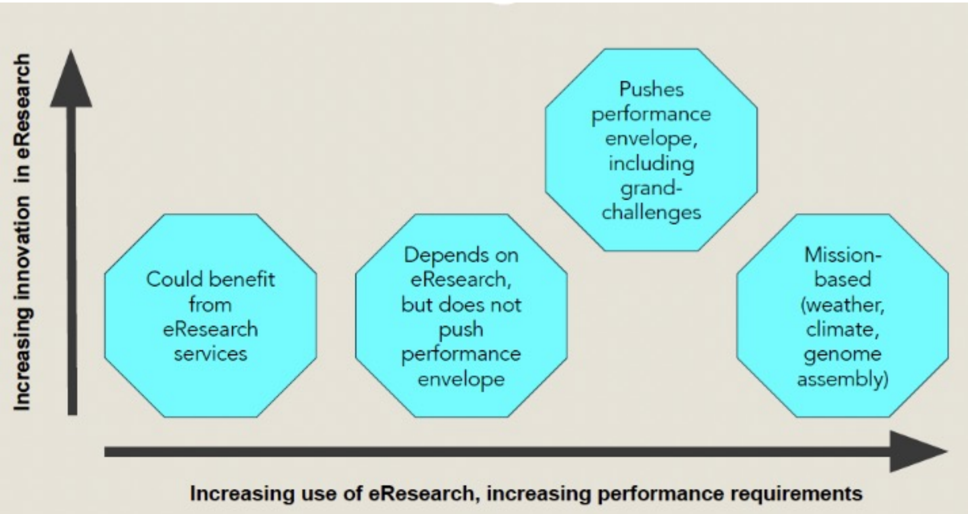


Figure 1: eResearch project spectrum (Dietrich, 2019).

The four project types in Figure 1 are ranked according to their use of eResearch (or digital research infrastructure) along the x-axis, which also identifies those projects that require the highest level of service (in terms of reliability). The same project types are ranked by increasing innovation on the y-axis, with those higher up the axis requiring more adaptability. The eResearch services of 11 countries were analysed, including the Netherlands, New Zealand, Singapore, Sweden, Switzerland, UK, and the US. This revealed that eResearch services are usually provided by multiple organisations, with the result that coordination and complementarity are common issues. This means that researchers must often take responsibility for finding and integrating the services that they need (Dietrich, 2019).

UK profile

UKRI is developing a Digital Research Infrastructure (DRI) strategy that recognises the substantial opportunities that will arise from exascale systems, and the need to "provide appropriate and ambitious compute capabilities reaching out to exascale for UKRI's diverse research and innovation communities" (UKRI, 2022c).

The UKRI Supercomputing Science Case presents cases showing that supercomputing resources will underpin the future success of UKRI-supported research. This report also emphasises that "in an era of new computing architectures, we need to nurture and develop the key skill sets that underpin successful supercomputing. [...] What is more, development of these skill sets as an activity in and of itself also helps develop the skill sets that lead to inventive technology" (Wilkinson et al., 2021). Initiatives such as the UK National Quantum Technologies Programme and the government's National AI Strategy also support the development of digital infrastructure capabilities and personnel.

Reports such as UK Data Research Infrastructure Landscape identify the digital research infrastructure required to support specific needs—in this case, to support the Data and Analytics Research Environments (DARE UK) programme in understanding the needs of those using current research environments, including trusted research environments (TREs). TREs are highly secure digital environments that provide access to sensitive data for approved researchers. They are designed to support the development of a coordinated vision for digital research infrastructure in the UK, with a particular focus on those managing sensitive data (DARE UK Consortium, 2021).

2.3 POLICY

Policy plays an important role in generating recognition and support for research software and personnel, covering research environments, funding, and coordination. A range of stakeholders provide different policy elements, including governments, funders, digital object repositories, publishers, and employers.

2.3.1 Enabling environments

There is increasing international recognition of the crucial role that research software and its personnel play in research outcomes. The Organisation for Economic Co-operation and Development (OECD) and the United Nations Educational, Scientific and Cultural Organisation (UNESCO) have both contributed high-level strategic support in this area. The OECD Council has revised its Recommendation of the Council concerning Access to Research Data from Public Funding to include software, with countries potentially facilitating policy to implement the recommendation. It provides policy guidance in areas including bespoke algorithms, workflows, models, and software (including code). Recommendations include fostering the adoption of good practice for research software management, training a cadre of RSEs, and properly recognising and rewarding software development skills (OECD, 2021).

Similarly, the UNESCO Recommendation on Open Science includes software in its definition of open scientific knowledge, such as source code that is available in the public domain or under copyright and licensed under an open licence (UNESCO, 2021). Analysis highlights that "in particular, the recommendation argues for users to gain free access to open source software and source code in a timely and user-friendly manner, in human- and machine-readable and modifiable format, under an open licence. The source code must be included in the software release and made available on openly accessible repositories, and the chosen licence must allow modifications, derivative works and sharing under equal or compatible open terms and conditions" (Barker et al., 2021). The Research Software Alliance (ReSA), which brings together the international research software community to campaign for the recognition of research software as a vital component of research worldwide, also emphasises three recommendations that comprise an international shared vision:

7. Research software must be recognised as a key element of research.
8. The development and maintenance of research software must be supported.
9. Research software must be as open and/or as FAIR as other components, so that the research it enables can be trusted and replicated (Barker et al., 2021).

International landscape

A number of governments now recognise the importance of research software and its personnel. The 2021 National Roadmap for Large-scale Research Infrastructure, published by the Dutch Research Council (NWO), has made FAIR, sustainable software, and a software management plan conditions for receiving funding (NWO, 2021). Software is central to the funding the NWO has made available for developing 'digital competence centres' as part of the national roadmap, and the NWO has set up an open science team to push this agenda. The Netherlands eScience Center (NLeSC) plays an important role in raising awareness about research software in the Netherlands, as a national research organisation specifically mandated to pioneer software in direct collaboration with researchers.

Many countries now support the recognition of open software as a key pillar of their open science strategies. These strategies are a major driver of research software advancements, encouraging open access to scientific publications, research data, metadata, open educational resources, software, source code, and hardware (UNESCO, 2021). Countries that have open science strategies include Canada (Office of the Chief Science Advisor of Canada, 2020), 14 European nations (although in some cases these are focused only on open data and/or open access) (SPARC Europe & Digital Curation Centre, 2019), and Japan (Japan Science and Technology Agency, 2017).

Other countries are beginning to emphasise open software practices. A reflection on open science practices and research software in Saudi Arabia provides an interesting perspective: "Saudi Arabia introduced coding as an important STEM-based skill and supported initiatives that focus on best practices in coding and building research software [...] The research system is relatively new in the Arabic-speaking countries, which makes it easier to incorporate open science principles rather than shifting and shaping established systems" (Almarzouq, 2022).

Recognition of software is also increasing in other parts of the research ecosystem. The Software Citation Policies Index provides details of publishers who have software citation policies at either the publisher or journal level, with many requiring authors to cite the software central to their findings in line with published guidance from the FORCE11 Software Citation Implementation Working Group (CHORUS, 2022). The development of the FAIR Principles for Research Software is similarly improving the sharing and reuse of research software, by making it FAIR (Chue Hong et al., 2022). A range of adoption guidelines and examples of implementation in a number of organisations are available to support implementation (Martinez, Barker, et al., 2022; Martinez-Ortiz et al., 2022).

UK profile

Software is seen as a key strength of the UK's research portfolio, and there are various policies in place to support this. UKRI supports open research, which focuses on open access, open data, and responsible research assessment. The UK government includes a policy on open source software in its Technology Code of Practice. This set of criteria, formulated to help government design, build, and buy technology, includes the following stipulation: "Be open and use open source: Publish your code and use open source software to improve transparency, flexibility and accountability" (GOV.UK, 2021).

There are some research software policy requirements within the work of the various UKRI Research Councils. For example, the EPSRC policy framework on research data defines data as "recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; consequently this can include research software" (UKRI, 2022b). Arts and Humanities Research Council (AHRC) is currently undertaking work that may lead to inclusion of software in a data strategy, and the work of the Particle Physics Technology Advisory Panel (PPTAP) includes recommendations around supporting research software in the Science and Technology Facilities Council (STFC).

2.3.2 Funding

One of the most commonly identified challenges to sustaining research software and its personnel is funding, with works such as the 10 Simple Rules for Funding Scientific Open Source Software providing clear guidelines on ways forward (Strasser et al., 2022). There are also overlaps between policy and funding elements, as policy mechanisms can be used to enable funding.

International landscape

It's possible to differentiate types of funding between the two stages of the research software lifecycle: 1) research and development, and 2) maintenance and support. Analysis using this framework has demonstrated that there is a significant level of investment in the research and development of research software—although the vast majority of this relates to research software developed for a particular research project rather than for general use. This funding comes from a range of sources, including government, philanthropy, and industry. By contrast, funding for the maintenance and support of research software is in far shorter supply, with the consequence that much of the research software initiated using research and development funds faces sustainability issues in the longer term. Additionally, where funding for maintenance and support does exist, it is still usually funded for periods that do not match the timescales of the software's use (Barker & Katz, 2022).

Canada identifies research software funding as a key challenge and opportunity in similar terms to the framework described above: "There is insufficient support in targeted and sustainable research software funding. Traditional research funding is fundamentally innovation-focused, leaving the need to sustain research software for the long-term up to the ability of principal investigators to describe their needs in an innovation context.[...] A new approach to research software funding needs to recognise different types and phases of research software (e.g., experimental; emergent but production-level; established/enterprise), and devise appropriate evaluative mechanisms, metrics, and funding streams for each." This applies equally to the personnel that support research software: "With the significant demand for RSEs across domains of practice, and the potential for research software developed within one domain to have applications in other disciplines, there is a need to develop a comprehensive RSE funding model" (Digital Research Alliance of Canada, 2021).

Examples of funding programmes that support the maintenance and support of research software include:

- CANARIE (Canada): Funded Research Software Platforms
- Chan Zuckerberg Initiative: Essential Open Source Software for Science
- NASA (US): Support for Open Source Tools, Frameworks, and Libraries

Examples of countries that provide programmes specifically for software across both parts of its lifecycle also include:

- German Research Foundation (DFG): Qualitätssicherung von Forschungssoftware durch ihre nachhaltige Nutzbarmachung and Research Software Sustainability
- US NSF: Cyberinfrastructure for Sustained Scientific Innovation (CSSI) includes a new project class, Transition to Sustainability.

UK profile

The UK currently provides funding for both stages of the research software lifecycle. For example, BBSRC and EPSRC's Transformative Research Technologies funding programme supported research and development projects where research software was intentionally developed as a product for general use in research by one or more projects; and EPSRC's Software for research communities programme provided funding for the maintenance and support of research software.

Funding has also been available through UK Research Councils for both research software and its personnel. For example, Embed digital skills in arts and humanities research is an AHRC funding programme to support the design and piloting of digital skills training on the use of digital tools and methods. Successful project leads will be invited to develop a scalable pilot for a regional or national training centre for digital skills in arts and humanities (UKRI, 2022a).

In another example, the EPSRC software infrastructure strategy 2018 sets out the investment strategy for supporting the development of reliable and reproducible research software by providing funding, training, and appropriate policy and best practice frameworks (EPSRC, 2018). Key elements of the EPSRC's research software investment trajectory are shown in Figure 2.

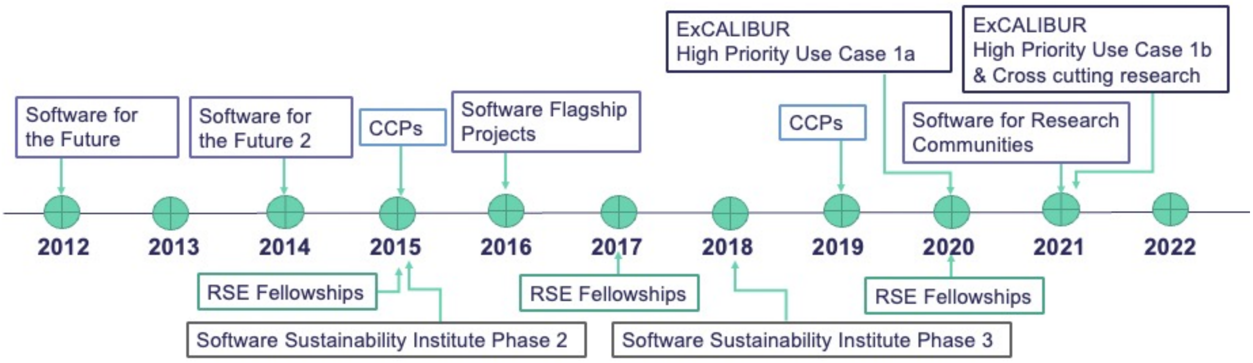


Figure 2: A snapshot of EPSRC funding programmes for research software (Janacek, 2022). Permission granted for reproduction.

3. METHODOLOGY

This section explains the methodological framework for this study, and outlines its strengths and limitations.

3.1 Study design

This study involved a mixed methods approach with the following elements:

- An online survey with 405 responses.
- Eight online group interviews (total of 23 participants) and one pilot group.
- Four online stakeholder interviews (total of six participants).

3.1.1 Survey configuration and distribution

This project included a large-scale survey of current software and skills requirements in the research community. **The survey was aimed at people who had been identified, or who self-identified, as users or supporters of research computing, e.g., a user, researcher, RSE, infrastructure provider, or trainer/educator.** The primary audience for the survey was principal investigators and co-investigators who had received funding from one of the UKRI research councils for their research within the last three years, excluding AHRC and ESRC as they had recently been contacted for complementary surveys. This identified 12,071 email addresses, of which 7,285 emails could be delivered. 2,263 (31.71%) of the emails were opened and 342 (4.7%) recipients clicked on the link provided to the survey. Secondary target audiences included other research software-related communities detailed later in this section. Out of the total of 7,285 delivered emails, 6,022 (83%) were worded incorrectly, which might have biassed the results of the survey as it was originally misworded to capture only EPSRC-funded researchers. However, evidence showed that researchers from other funding bodies still reacted and filled in the survey.

The survey (contained within Appendix B) aimed to gauge the extent of immediate needs and barriers. It was narrow in scope in order to maximise responses across a broad range of constituencies, including different career stages, disciplines, and institutions. Cognitive interviews, where a member of the study team observed pilot reviewers completing a survey, were undertaken to ascertain the comprehensibility and flow of the survey through a test run before its commencement. Invitations to participate in the survey were sent via email and shopping vouchers were offered as an incentive to compensate participants for their time. Participants were also invited to opt in to being contacted about follow-up research in the form of group interviews, the findings of which also inform this report.

It should be noted that the survey focused on, but was not limited to, the EPSRC remit. AHRC and ESRC researchers were not directly invited to contribute to the survey as related surveys were underway in different projects (SSI, 2022a, 2022b). While this report is limited to the data collected in its own survey, it will be possible in the future to compare this data with that of the AHRC and ESRC survey to cover the broader community. This survey also deliberately omitted RSEs, as the 2022 international RSE survey targeted this demographic.

The survey was advertised through email campaigns, Twitter, LinkedIn, Slack channels, and websites, including:

- SSI Fellows via email and Slack (180 members). These members consist of a mix of both EPSRC and non-EPSRC funded fellows.
- SSI's Collaborations Workshop Slack channels for 2020 (88 members), 2021 (114 members), and 2022 (108 members).
- UK-RSE Slack channel.
- EPSRC Twitter and LinkedIn.
- EPSRC Digital Research Infrastructure Committee and the ExCALIBUR Strategic Priorities Fund mailing lists.
- ARCHER2 weekly newsletter.
- SSI and Edinburgh Parallel Computing Centre (EPCC) Twitter and websites.
- Collaborative Computational Projects and Higher Education Commission distribution channels.
- N8 Centre of Excellence in Computationally Intensive Research RSEs and Bede Support Group (a subgroup of the N8 CIR Project Working Group) mailing lists.

To encourage participation from under-represented groups, survey distribution was also provided from groups including Women in HPC, University and College Union disabled members' standing committee, Centres for Doctoral Training, NERC doctoral training partnerships, and Medical Research Council (MRC) doctoral training partnerships.

3.1.2 The composition of group interviews

After the survey stage was completed, eight qualitative group interviews were conducted online with 23 participants. These participants had completed the survey and opted in to further participation in the study. Group interviews were designed to facilitate discussion of individual experiences, opinions, and attitudes, and to cover topics relevant to software, roles, and skills required for large-scale research computing. They aimed to capture the diversity of research practices and cultures within various software communities, and they asked for opinions on funding and policies required to future-proof access to, and use of, digital research infrastructures. The topic guide utilised for the group interviews is provided in Appendix C.

In order to ensure that the structure and questions of the group interviews had been suitably crafted, these were pilot tested amongst a group of five software researchers affiliated to the SSI. Group interviews were organised based upon the availability of participants. Where possible, efforts were made to stratify the group sessions by research institution, research background, and career stage based on the demographic information that was returned within the surveys. Qualitative data analysis was undertaken through the use of NVivo. Abductive/ iterative qualitative coding was employed: interesting or surprising findings from the community survey analysis were used as a funnel for group interview data analysis (see Tracy, 2020) in order to draw out explanatory top-level codes. These categorisations informed a more focused thematic analysis within the objectives of this study.

3.1.3 Stakeholder interview approach

The final input for this study was derived from four semi-structured interviews undertaken with key stakeholders. The specific structure and questions for these interviews were informed by the key findings of the community survey and the group interviews. Interview participants were drawn from digital research infrastructure providers, users, funders, and policy makers. These participants had not taken part in the group interviews. The stakeholder interviews involved one or two interviewers and interviewees. Speaking to key stakeholders independently allowed for in-depth and confidential accounts to be obtained. The interview questions, contained in Appendix D, were designed to capture views on the digital skills, software needs, opportunities, and barriers present in the digital research infrastructure landscape.

3.2 Methodological triangulation and limitations

This section describes the rationale behind the sequential ordering of this study's data collection methods and explains how multiple perspectives were analysed in order to obtain a more comprehensive understanding of the software and skills required for large-scale computing. Limitations of this approach are then discussed. A community survey provided participants with an opportunity to document the research software that they use, how they use it, the digital skills required for its use, and immediate needs and barriers within software research. Follow-up group interviews garnered further insight into survey responses, as well as incomplete or non-responses, since predetermined survey questions may not have entirely fitted the expectations of respondents (Pawar, 2004). Additional knowledge acquired during the follow-up qualitative research was used to contextualise (Flick, 2018) the results from the survey in stage one and to evaluate to what extent the findings from each of these processes converged, diverged, or supplemented each other (Kelle & Erzeberge, 2004). Figure 3 summarises this study's data collection process and the triangulation of different research methods.

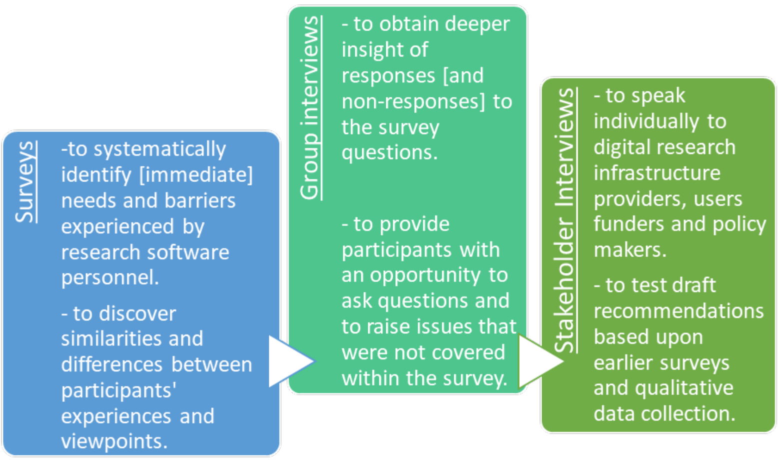


Figure 3: Data collection process and triangulation of study research methods.

Methodological triangulation and the use of diverse data collection techniques can increase the quality of research and extend the range of insights produced using a single methodology (Flick, 2018). However, there are limitations to its implementation. Firstly, the recruitment to each data collection stage was non-random, creating the necessity to report self-selection bias. An advertisement for this study, including a link to the survey, was disseminated via institutional mailing lists and within online fora used by UK-based RSEs. This may have resulted in self-selection bias, whereby individuals who were reached out to, and who participated in the survey, were not completely representative of the target population.

Secondly, the completion of a survey and participation in follow-up qualitative work resulted in attrition bias (Miller & Hollist, 2007). Attrition bias can arise when participants who drop out of a study, or who proceed to take part in follow-up research, may be systematically different to those who remain in the study (Miller & Hollist, 2007). Within this study, this may have occurred both in terms of those who returned a completed survey (1,380 people started a survey; 410 completed one) and those who took part in a survey and a follow-up qualitative session (124 participants were contacted and 23 stated that they would be willing to participate in a follow-up). Only one of the five organisations invited to take part in the stakeholder interviews did not respond. Finally, while focusing on a small sample allowed detailed insight to be obtained within a short timeframe, these contributions may not be representative of research software personnel in the UK. It should be noted that throughout this report the analysis of data by gender is restricted to participants who described their gender as male (including trans men) or women (including trans women). Survey and group interview participants were also able to choose other options in relation to gender, but response numbers were too low to enable meaningful analysis. The number of respondents who stated that they were in the Junior career stage was also too low to enable meaningful analysis of this cohort.

Finally, to quantify uncertainty in the observed proportions for survey data, approximate 95% confidence intervals for binomial proportions, based on asymptotic normality, were used. Data were analysed using two-sided approximate hypothesis tests for difference between two proportions at a 5% significance level. Where there is evidence for rejecting the null hypothesis at a 5% significance level, the information below the figures includes the corresponding probability value (a number describing how likely it is that the data would have occurred by random chance). Due to the nature of the confidence intervals (shown in Appendix E for all survey questions) and the hypothesis tests, interpretations should be conservative.

4. ANALYSIS

This section presents the study's results across people, infrastructure, and policy areas. It begins with participant demographics, then presents data from the survey questions, group interviews, and stakeholder interviews. A question by question (or univariate) breakdown of all survey responses is contained in Appendix E, and some analyses where answers to more than one question are combined (or multivariate) are contained in Appendix G. Any names of people that appear are pseudonyms in order to protect the identities of study participants.

4.1 PARTICIPANT DEMOGRAPHICS

Figure 4: Mapping of location of survey participants.

Analysis of the ethnicity, gender, and reported disability of participants in the survey, and group interviews, shows the following breakdown in comparison with a range of broader UK demographics in Table 1. Figure 4 also shows their location within the UK.

Percentage who are:	Survey	Group interviews	UK RSEs	UK academics	UK software developers	All UK workers
Ethnicity: BAME/mixed	7%	22%	6%	15%	21%	12%
Women	23%	17%	20%	47%	14%	48%
Report disability	6%	9%	6%	5%	10%	13%

Table 1: Under-represented minorities participating in this study in comparison with other UK studies. Green indicates that the value is more than 25% above the percentage of all UK workers. Orange indicates that the value is more than 25% below the value for all UK workers.

Sources: RSE data for ethnicity and women (RSE International Survey 2022, 2022), RSE data for reporting disability (softwaresaved/international-survey, 2019); academics (HESA, 2021b); software developers and all workers (BCS, 2022).

This shows that the breakdown of the 405 survey responses by these categories were at least 33% smaller than for the UK workforce as a whole, and 50% smaller for women (see Appendix E, Q29, Q30, and Q31). Of participants who opted in to and completed a group interview, the 21.6% rate of participation by those who identify as Black, Asian, and minority ethnic (BAME)/ mixed was almost 50% higher than that of the UK workforce as a whole; figures for women (17.4%) were more than 50% lower than for the UK as a whole; and for those who reported a disability the figures were more than 25% lower than the UK workforce as a whole (see Appendix F). This aggregation of ethnic minorities is problematic as it emphasises certain minority groups at the exclusion of others, and masks disparities between ethnic groups that can create misleading interpretations of the data (GOV.UK, 2021). This terminology is being used in this report to allow comparison with previous studies.

The survey asked participants which organisations they worked for. The vast majority of participants were employed by universities or other types of research institutions (see Appendix D). Of participants who took part in a group interview, 87% reported that their host institution was a university, 8.7% reported Research Centre, and 4.3% reported Research Council.

Figure 5 shows the breakdown of "Research" and "Research and Teaching" among survey respondents, and all participants in the group interviews, by the discipline in which they are employed, with the option to choose more than one answer. This is also compared in Figure 5 with HESA data combining the higher education staff involved in "Research" and "Research and Teaching" (HESA, 2021a). Details of which subjects are categorised in each of these areas are provided in the Higher Education Statistics Agency (HESA) principal subject codes (HESA, 2013).

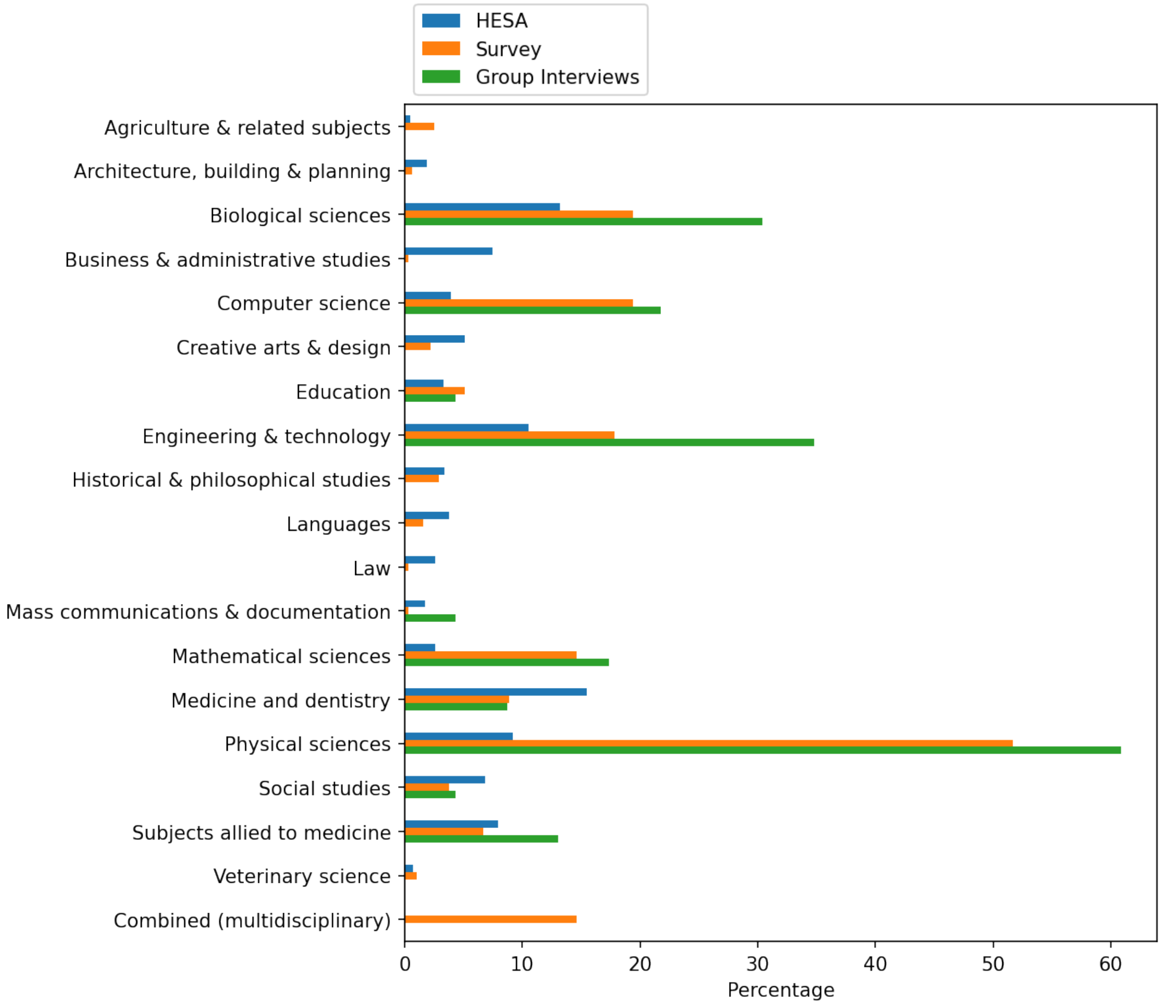


Figure 5: Survey and group interview participants by disciplines they work in, compared to the number of researchers in the UK research sector (HESA data).

This shows that the participants described in this report do not reflect the UK research sector as a whole—the report includes a greater proportion of respondents from disciplines like physical sciences, mathematical sciences, engineering and technology, and computer science. An uneven distribution of disciplines is to be expected, since this study had a focus on researchers within the EPSRC community, and did not target communities such as those of the AHRC and ESRC due to the existence of related surveys. It should also be noted that there are some differences between how this report and HESA data links researchers and disciplines. This report enabled participants to self-identify and choose more than one discipline, whereas HESA maps staff disciplines based on the cost centre that their employment is linked to. There is also no exact alignment between the principal subject codes used in this survey and the HESA cost centres (see Appendix H).

A number of survey questions investigated issues related to funding, with survey participants being asked which organisations they had applied to for funding. Results are shown in Figure 6.

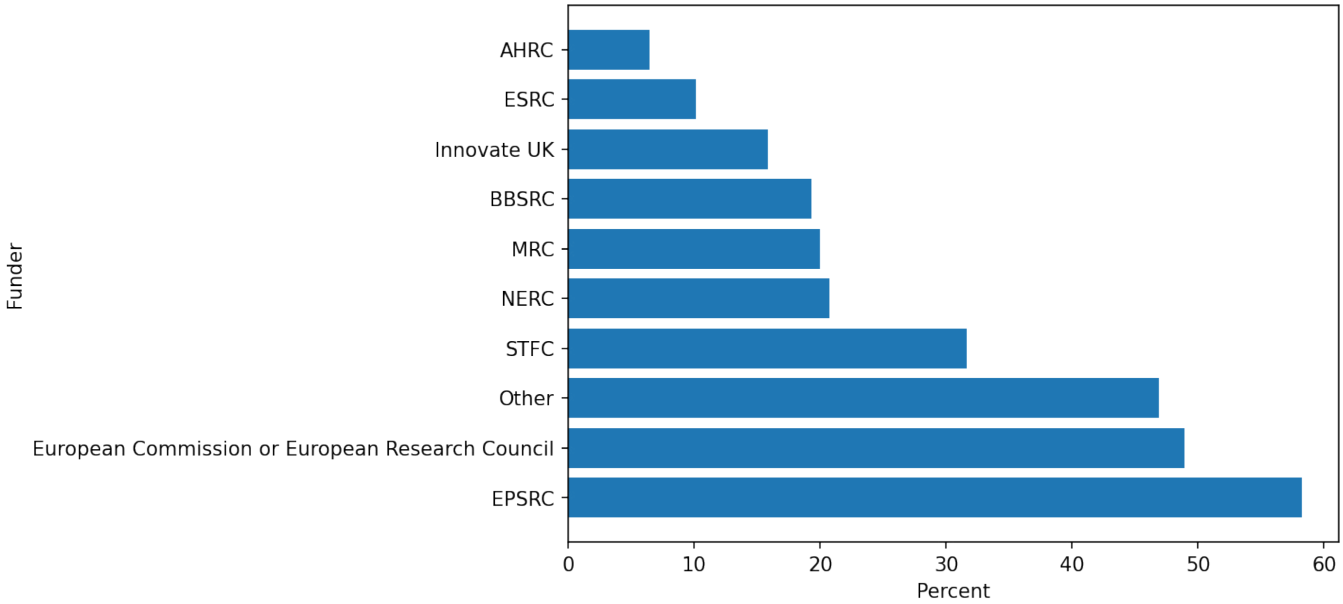


Figure 6: Where survey participants have applied for funding from.

52% of respondents had applied for funding from EPSRC, only 10% for ESRC, and 6% for AHRC, reflecting the uneven distribution of disciplines described above. The 152 survey participants who chose "Other" as their response gave rise to 100 unique responses. The majority of these were charities such as Leverhulme Trust (42 responses), Royal Society (34), and Wellcome Trust (31); government agencies such as National Institutes of Health (NIH) and National Institute for Health and Care Research (NIHR) (12); and industry/company funders such as Dunhill, Google, and NVIDIA (two responses each).

4.2 PEOPLE

This section aims to identify opportunities to ensure that there are appropriately recognised and skilled personnel in the research community to maximise the use of research software. This is examined across three subsections:

10. Career paths: what roles are required to incorporate research software expertise into research outcomes, and how do career stages affect this?
11. Skills and training: what digitals skills are needed to facilitate/enable access to software and infrastructures, and what else will be needed in the future? What digitals skills exist and how do they differ across disciplines and career stages?
12. Diversity: where do people have divergent experiences?

Key findings:

> RSEs are seen as an essential role in the research computing ecosystem that must be supported with appropriate career paths.

> Some research computing adjacent roles, such as Research Librarian and Data Steward, are less visible, and therefore less often utilised by researchers engaged in scientific computing.

> Challenges in accessing adequate training may result in barriers to professional development and learning, particularly for those earlier in their careers. These challenges relate to areas including embedding skills in curricula and recognising that much learning is on the job/self-taught.

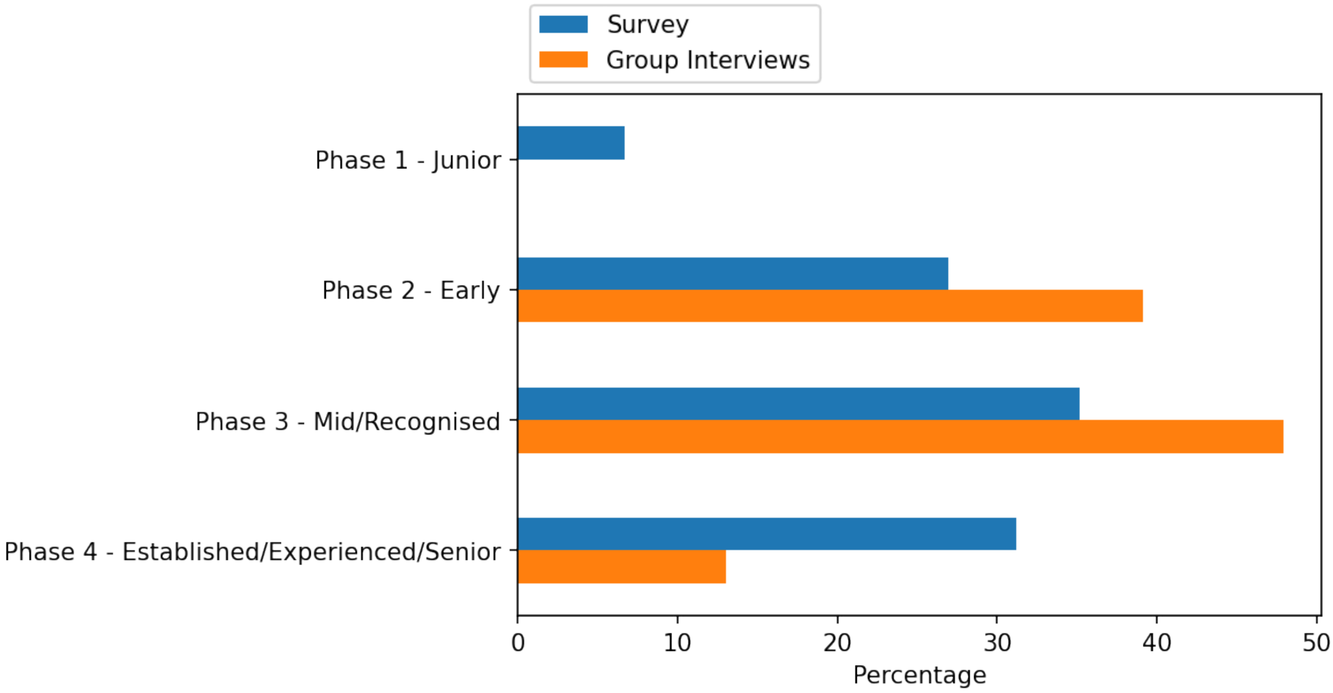
> There are still gender inequalities in scientific computing, particularly with regard to skills acquisition. More research is needed to obtain a more nuanced understanding of opportunities and challenges, and to enable similar analysis of other under-represented groups.

4.2.1 Career paths

The study aimed to provide information on career stages and primary roles in order to explore issues related to career paths. To achieve the aim of understanding the current workforce, the study asked participants to identify their primary roles, as shown in Figure 7. The survey used the four phases of career stage identified by the League of European Research Universities (LERU), to enable broad mapping of equivalencies between various national research systems (LERU, 2022):

13. Phase 1: Junior (e.g., PhD candidate, Junior RSE).
14. Phase 2: Early (e.g., Research Assistant/Associate/Fellow, first grant holder, Lecturer, RSE).
15. Phase 3: Mid/Recognised (e.g., Senior Lecturer, Reader, Senior Researcher, Senior Research Fellow, Senior RSE, Research Software Group Leader, Senior Data Scientist).
16. Phase 4: Established/Experienced/Senior (e.g., Professor, Director of Research Computing, Distinguished Engineer, Chief Data Scientist).

Figure 7: Survey and group interview participants by career stage.



It should be noted that the number of participants from the Junior career stage was relatively low in the survey (7%, or 27 participants), and none of this cohort volunteered to participate in a group interview. Consequently, **analysis within the report of data related to the Junior career stage is limited.**

In order to investigate different roles required to incorporate software into research, the survey asked participants to identify their current primary roles. These results are displayed in Figure 8.

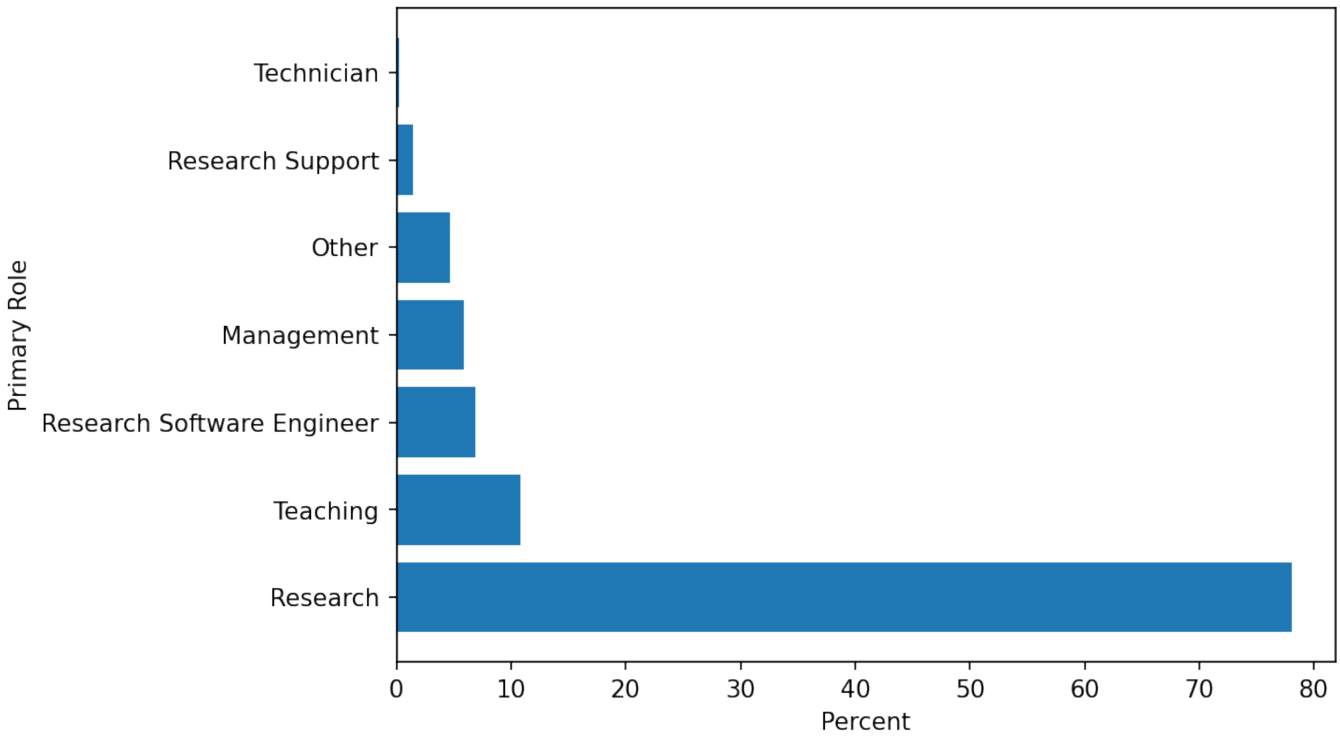


Figure 8: Survey participants by primary role.

This shows that while 74% of survey respondents were researchers, a range of other roles engage with research software and its personnel. Further breakdown of this data by participants' career stages showed a fairly even spread across the primary roles (see Appendix G, Q5 x Q6), with the exception that Established/Experienced/Senior career stage staff were more engaged with management and less with teaching, as would be expected.

7% of participants identified RSE as their primary role (see Appendix E, Q5). While a growing number of researchers code (and it is the norm in some areas like bioinformatics), they do not necessarily identify as RSEs. An Australian report commissioned by the Australian Research Data Commons (ARDC) identified 80 different job titles commonly used for research software staff, with 39% of these being titles commonly utilised for traditional academic roles such as lecturer, professor, etc. (Barker & Buchhorn, 2022). How many RSE in the world? calculated the number of RSEs in the UK based on job advertisements to suggest that 14% of total UK academics (31,000 people out of 220,000) could be considered RSEs (Hettrick, 2020a). The Australian study also analysed and contextualised the survey results against national and international studies to consider the extent to which this RSE capability may be meeting Australia's research needs. It found that 56% of respondents felt there was insufficient research software capability in their area, and 43% indicated an intention to recruit more personnel in the next one to three years (Barker & Buchhorn, 2022).

Comparison of those survey participants who identified their primary role as RSE with their career stage shows that 3% were from the Mid/Recognised career stage, and 3% from the Early career stage, out of a total of 7% of participants (see Appendix G, Q5 x Q6). RSE has only been recognised as a job function and/or title within the last five to 10 years, which explains why Established/Experienced career stage RSEs are still rare.

Two of the four stakeholders interviewed claimed that their respective organisations had played a role in securing acceptance for RSEs in the research sector: "I think we've supported the growth of RSEs in the UK and, and internationally, and shown that it is a group of people worth having, and who can come in a sense pay for themselves." However, interviewees also emphasised that there was much more to be done. In the same vein, 10 Simple Rules for Funding Scientific Open Source Software identifies as rule number five: "Promote RSE as an academic career" (Strasser et al., 2022).

Interviewees consistently identified career paths as a the primary issue for RSEs, with other commonly referenced topics including low salaries in comparison to industry, lack of permanent career paths, and inability to move to and from industry: "One of the things we're trying to do is look at how we can provide sensible career progression for RSEs. We have models in place at the moment that I think we are drastically lagging behind where we need to be to compete with industry." These views are common in international literature. In an Australian report, 61% of respondents stated that there was no process for their research organisations to recognise software development or maintenance effort towards academic progression (Barker & Buchhorn, 2022).

Mobility is also commonly cited as an issue in international literature. Canada has a strong focus on industry engagement in this regard: "Support for research programs with a research software requirement within Canada has often come through employment programs for student [sic] and recently graduated Highly Qualified Personnel from computer science, software engineering, and disciplinary fields. There is a strong history of industry-institution collaborations via co-op programs, strengthened by MITACS and other government incentives" (Digital Research Alliance of Canada, 2021).

Some UK research organisations have programmes that facilitate more engagement between research staff and industry. For example, Distributed Research using Advanced Computing (DiRAC) Innovation Placements provide an opportunity for businesses, startups, UKRI partners, and technology companies to work with DiRAC on a joint project. DiRAC also provides hackathons, proactively engaging with manufacturers and suppliers to help their researchers assess the impact of technology on their particular research codes.

Given the focus of this study on research software, the survey sought to understand how many respondents develop software (i.e. write their own code) for their research. 80% of participants reported that they write their own code (see Appendix E, Q12) and this is in line with previous literature. A 2014 UK survey of researchers from Russell Group universities (albeit among a slightly different sample) found that 56% developed their own software (Hettrick, 2014).

To establish whether any trends could be observed in the research disciplines of survey participants who reported creating software, Figure 9 shows the results for five disciplinary areas. **In this analysis, and all discipline analysis throughout this report, the disciplines of computer science and mathematical sciences are combined—as are medicine and dentistry, and subjects allied to medicine.**

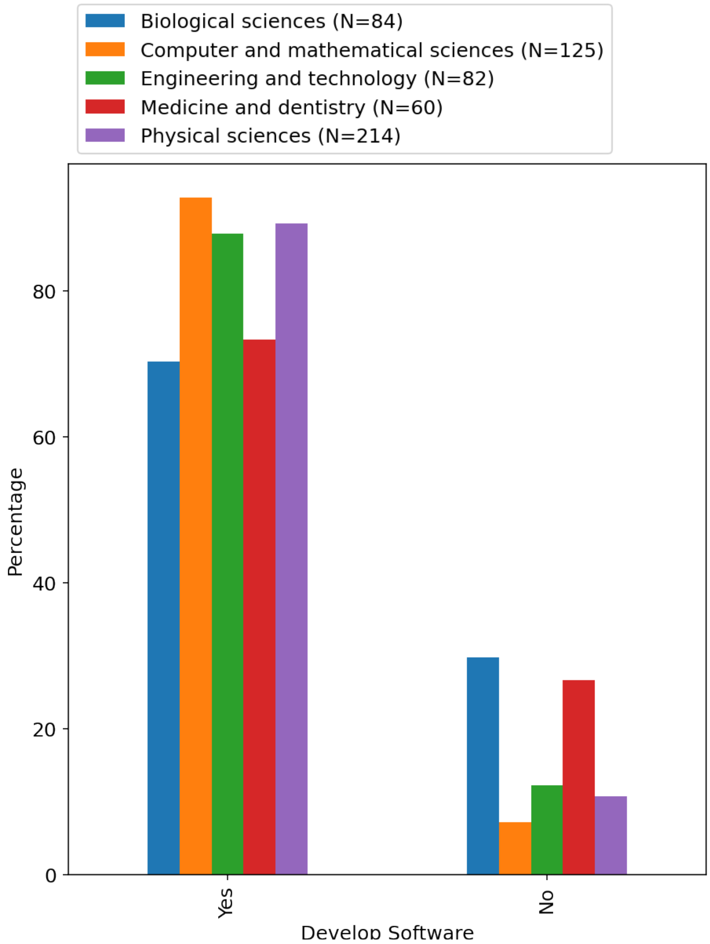


Figure 9: Survey participants by disciplines they work in, and whether they develop code.

This shows that participants from two disciplinary areas (biological sciences, and medicine and dentistry), are slightly less likely to develop code (biological sciences 70.24%, medicine and dentistry 73.33%) than the other three disciplines considered here (computer and mathematical sciences 92.80%, engineering 87.80%, physical sciences 89.25%). This is to be expected because the disciplines of computer and mathematical sciences, engineering and technology, and physical sciences are usually regarded as computationally intensive. Analysis of the UK RSE survey responses provides a point of comparison. While RSEs are a slightly different demographic to the sample of participants for this study, these results assist in establishing computation-intensive disciplines. The top three UK RSE responses to discipline of study and work were: physics and astronomy (38%), computer science (37%), and biological sciences (28%) (Philippe, 2018). Interestingly, within the group interviews individuals at a Mid/Recognised career stage described a widespread lack of computational skills in their field, inhibiting software development at a Junior or Early career research stage (these are described in greater detail within section 4.2.2).

These participants also described facilitators and inhibitors that they had either experienced or noticed in relation to software development and career progression:

> *I've never recruited anybody either as a postdoc or as a PhD student who has any skills at all in this [computational science], they are almost always coming from an experimental biology background and need to more or less learn from scratch. There are various resources you can give them to learn how to do basic Python or R programming but I've rarely found any useful resources beyond the basics … a lot of the computational biology stuff doesn't go beyond very basic stuff.*

> *Two of my team are going to get to the top of their grade very soon and what do I do about those [colleagues] because we are not traditionally researchers? … our research areas [are] environment modelling, atmosphere, global climate, weather forecasting, anything to do with that … we're like an embedded RSE group … it's difficult for people to move up in any career form because they are not leading any publications, they are not doing research, they are, let's say, facilitating research.*

> *I'm really lucky compared to most RSEs because we have proper career progression and things like that available. I started at grade seven … and I've gradually progressed up and I'm now .. the most senior technical role you can have at [anonymised].*

> *When I was a student … I don't think there was that sort of training [software engineering: use of version control] available, I'm not entirely sure. I didn't seek it out … NERC doesn't really recognise software engineering as a needful part of its science, so there's no career path as a software engineer in NERC. So, all the scientists end up trying to do that as well and doing it badly, or not as well as could be done.*

To further investigate the prevalence of individuals developing code in different disciplines, Figure 10 shows the funders of survey participants who reported developing software. **The results for analysis by funders throughout this document are shown only with regard to the five UKRI Councils for whom the participant responses were larger: BBSRC, EPSRC, MRC, NERC, and STFC.**

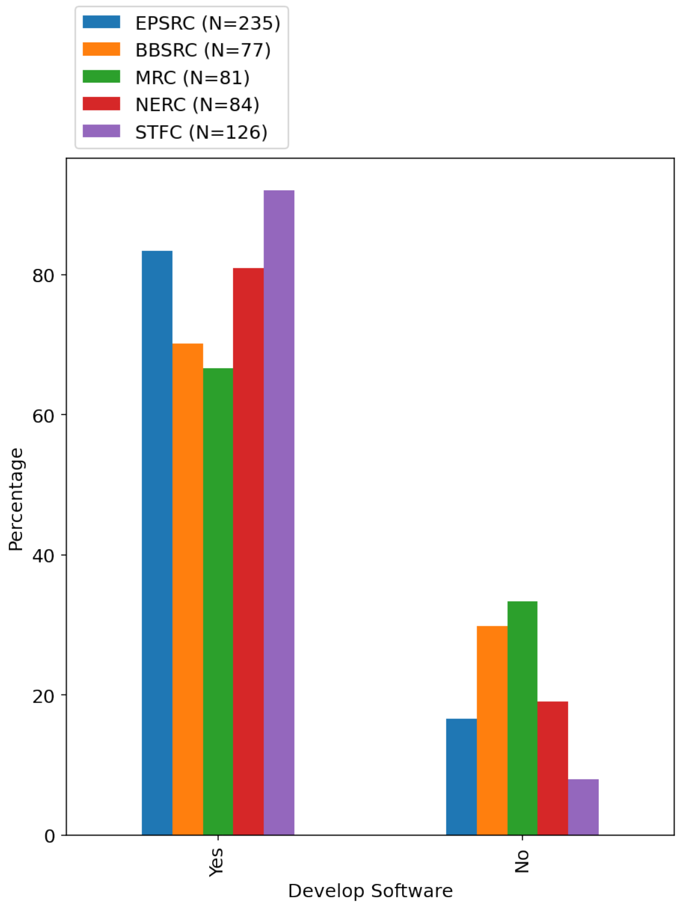
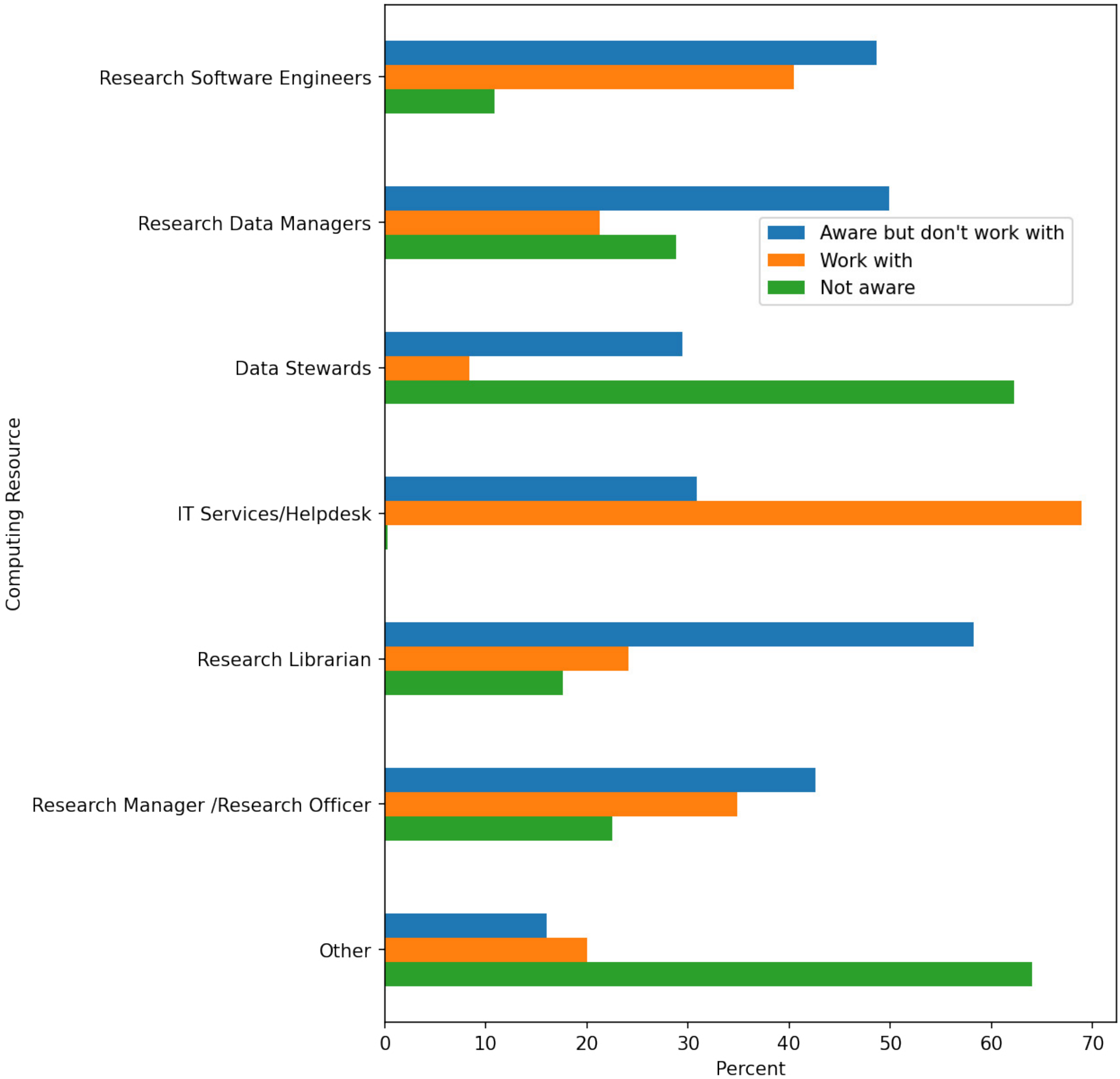


Figure 10: Funders that survey participants have applied to or who fund their work, and whether the participants develop code.

Survey participants engaging with BBSRC and MRC are slightly less likely to develop their own software. Software is developed by 83% of participants engaging with the EPSRC, 81% of participants engaging with the NERC, and 92% of participants engaging with the STFC—but participants engaging with the BBSRC and MRC are less likely to develop their own software, with only 70% and 67% developing software respectively. This trend corresponds with the figure above, which shows that participants who work in biological sciences and/or medicine and dentistry are slightly less likely to code. A further breakdown of this data by career stage demonstrated that coding steadily decreases as personnel become more senior, as would be expected due to increasing responsibilities in other areas.

Figure 11 shows the results of the final survey question related to career paths: "Which of the following supporting roles are you aware of, and which do you currently work with?"

Figure 11: Survey participants' awareness of supporting roles.



For many of these roles, a large proportion of survey participants are aware of them but do not work with them. This is particularly true of the Research Librarian role: 58% of survey participants were aware of the Research Librarian role but only 24% reported having worked with Research Librarians (see Appendix G, Q22.5 x Q6). The role of Data Steward may also not be as well understood or communicated as other roles, particularly for researchers in the early stage of their career, as 62% of survey participants were not aware of Data Stewards, rising to 71% among those in the Early stage of their career (see Appendix G, Q22.3 x Q6). Further research might be able to identify why this is the case, whether changes are required to make these roles more relevant to researchers, and whether researchers should engage organisations such as Research Libraries UK (RLUK) and the Chartered Institute of Library and Information Professionals (CILIP).

One interviewee also suggested that additional roles are required, highlighting a need to go beyond recognising and valuing RSEs to include a focus on developer relations. Developer relations roles are common in industry, and enable developers to maximise their abilities within an organisation in a range of ways, from engaging with users to improving project management.

There are already recommendations on how to improve career paths for research software personnel in the UK. Large-scale computing: the case for greater UK coordination provides explicit suggestions on how job security, salary structures, and progression opportunities should be improved to retain talent and respond to career path and progression challenges. This report concludes with steps that should be taken across academia and the public sector to retain large-scale computing professionals. It recommends that the UK public sector and academia should consider new career frameworks, fellowships, baselines, and salary structures to retain large-scale computing administrators and RSEs. The report also suggests that establishing RSE groups within universities offers a model for broadening access to RSEs and improving job security (Government Office for Science, 2021).

Similarly, the ExCALIBUR RSE Knowledge Integration Landscape Review makes a number of recommendations:

- UKRI should continue to invest in the development of RSE in the UK.
- Clear career paths for RSEs, along with funding opportunities that allow them to apply and develop skills, are crucially important to ensure that the knowledge they gain stays in the research sector and grows over time. The contribution of software engineering needs to be recognised in university recruitment and promotion procedures.
- Greater collaboration and transfer of skills by RSEs in both directions between the academic and industrial research sectors should be encouraged, particularly from industry to academia (Parsons et al., 2021).

The TALENT Commission also concludes that "[a]s many of the skills will be in high demand within industry, to compete, higher education institutions will need to support recognition, competitive pay structures and provide opportunities for progression to recruit and retain technical staff" (MI Talent et al., 2022).

Concrete solutions are beginning to be implemented. Some organisations, such as the UK Met Office and EPCC, have created career-based progression structures that recognise the skill sets and growth of software-focused staff (Parsons et al., 2021). The ExCALIBUR RSE Knowledge Integration Landscape Review notes that both duration of contracts and career progression need to be addressed, and also states that "the Society of RSE is currently reviewing plans to implement their own accreditation system", highlighting accreditation examples such as Chartered Engineer (CEng), Chartered Scientist (CSci), Institute of Physics (IoP), Institute of Mathematics and its Applications (IMA), and British Computer Society (BCS) (Parsons et al., 2021).

International solutions have emerged that could form the basis of further action. The National Center for Supercomputing Applications in the US provides five levels in its career path for research programmers: Assistant Research Programmer, Research Programmer, Senior Research Programmer, Lead Research Programmer, and Principal Research Programmer. The University of Manchester RSE group has three levels of RSE, from Junior to Senior (Katz et al., 2019). These can also be found in Australia (Barker & Buchhorn, 2022). Senior level RSE career paths (with an s) also suggest career paths that progress into different specialisations (Katz et al., 2021). Similar initiatives exist for other emerging research roles, such as the Academic Data Science Alliance and the RDA Professionaling Data Stewardship Interest Group. The latter is tackling challenges for data stewardship roles including identification of a business case, terminology, integration across an organisation, job profiles, training, career tracks, networking and knowledge exchange, and certification (Professionalising Data Stewardship IG, 2020).

There have also been community-driven efforts in the UK to increase the range of research outputs that are formally recognised (which should include research software). Analysis of the 2014 UK Research Excellence Framework (REF) showed that while 70% of research relies on research software, only 38 of the 191,000 outputs submitted included software in the 2014 REF (Hettrick, 2020b). In response to this, the Hidden REF crowd-sourced submission categories that included software, research datasets and databases, performance, physical artefact, training materials, and others. One article suggested that changes to the REF could positively change academic culture: "It is important that this issue is developed in the implementation of the Research and Development People and Culture Strategy" (Cleaver et al., 2022).

Major findings:

- RSEs need appropriate career paths and improved mobility with industry.
- EPSRC, NERC, and STFC-funded personnel are slightly more likely than BBSRC and MRC-funded personnel to develop their own software.
- There may be opportunities to increase engagement with Research Librarians, and to increase both awareness of, and engagement with, Data Stewards, particularly for researchers in the early stage of their careers.

4.2.2 Skills and training

The project focused on skills that participants had either built or sought to build, and how training was provided, to understand how best to enable a research workforce that can meet current and future research needs.

Figure 12 shows responses to the following question: "Over the last five years, have you learned new skills or knowledge in relation to any of the following? Select all that apply."

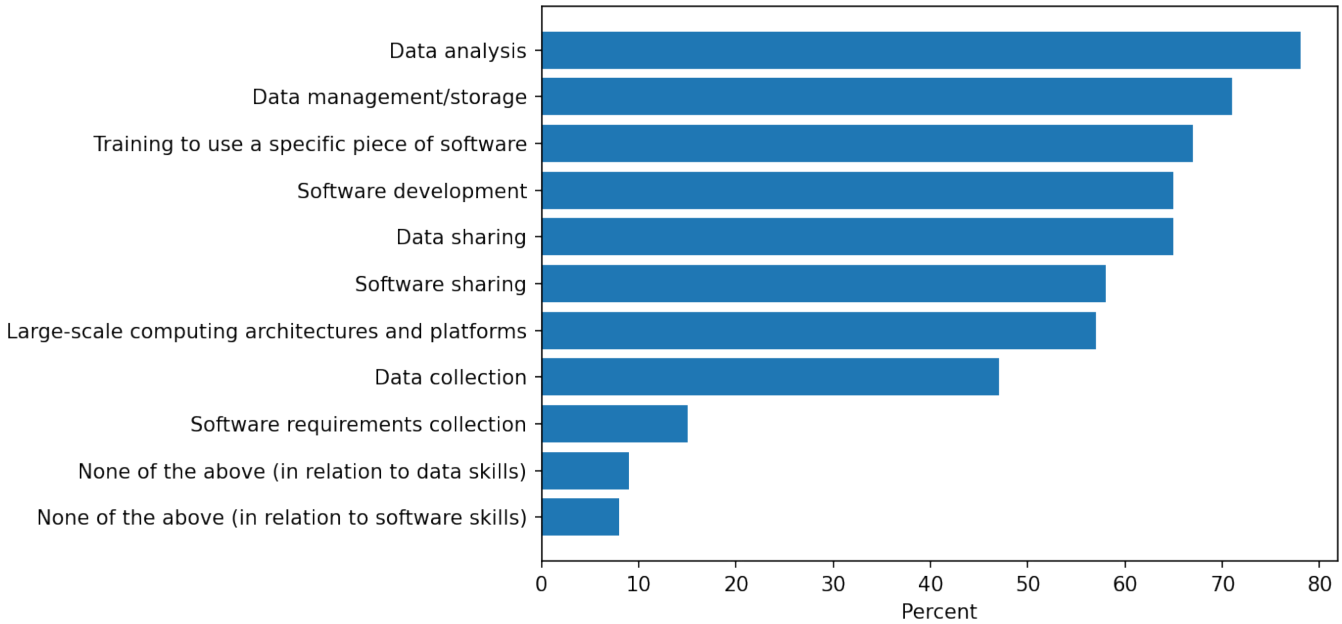


Figure 12: New skills or knowledge learned by survey participants over the last five years.

This shows that, on average, 68% of the survey participants are learning about data skills, compared with 60% for software skills. This is to be expected as almost all participants work with research data, while only 80% reported writing code.

65% of participants reported sharing data, which is higher than in some studies. For example, an analysis of carried out by BioMed Central of all articles from 333 open-access journals published during January 2019 found that of the 42% of articles (1,792 manuscripts) whose Data Availability Statement indicated that the datasets are available on reasonable request, 93% of authors either did not respond or declined to share their data (Gabelica et al., 2022). Increasing skills in research data management was identified as a key challenge in one stakeholder interview: "We need to change the culture so that people are putting enough metadata on their research data to make them FAIR. We can provide the tools and the hardware to do that, and some people to support it, but the community has to engage and essentially make their data more easily available".

These results demonstrate that the majority of participants are upskilling with regard to software. However, the question does not assess the quality of this training. Research on the software development training that people who develop research software require often shows that more skills are needed. In 2014, a UK survey reported that 21% of researchers who develop their own software (56% of respondents) had no training in software development (Hettrick, 2014). In comparison, in the Stanford Software Survey, where individuals assessed their own needs, 75% of participants felt that they had not received sufficient training for software engineering best practice. However, the majority did consider themselves to be at least proficient in terms of their software development expertise (Stanford Software Survey, 2020).

Further analysis of these survey responses to consider how the software skills being learned differ across career stages shows that training tends to decrease as participants' career stage advances (see Appendix G, Q23 & Q24 x Q6). This is a reasonable progression as senior staff often do less hands-on software development.

The survey responses on new skills or knowledge learned in the last five years were analysed further by focusing on responses that selected large-scale computing architecture and platforms, and segmented using data on funders from which participants had applied for and/or received funding. This is shown in Figure 13.

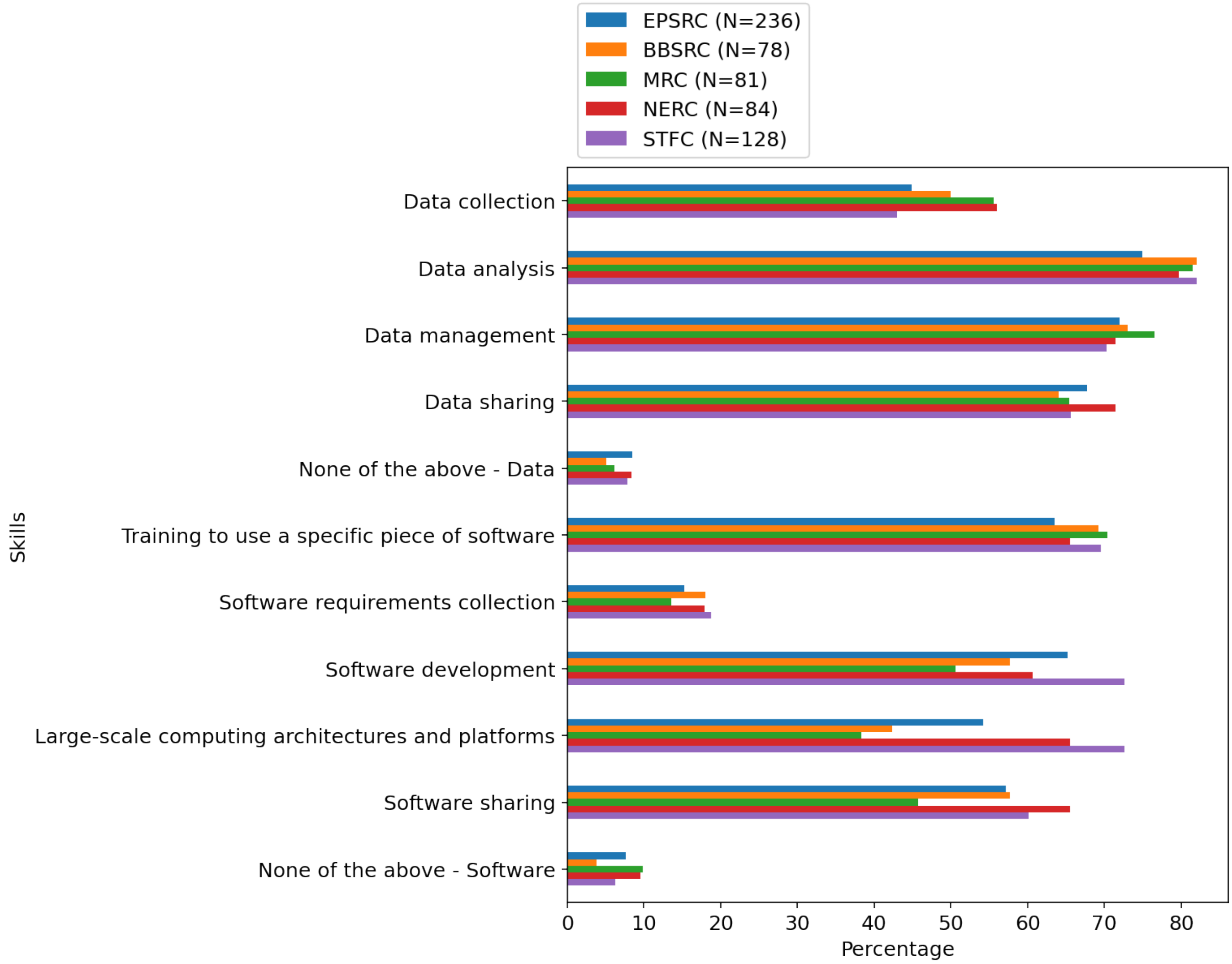


Figure 13: Survey participants identification of new skills or knowledge learned in the last few years, and funders from which participants had applied for and/or received funding.

This shows that survey participants who engage with some funders, particularly NERC, EPSRC, and STFC, have upskilled more in relation to computing in the last five years than those engaging with BBSRC and MRC. Training in large-scale computing architectures and platforms had been completed by 73% of participants engaging with STFC, as compared to 65% from NERC, 54% from EPSRC, 42% from BBSRC, and 38% from NERC. Similarly, training in software development had been completed by 73% of participants engaging with STFC, 65% from EPSRC, 61% from NERC, 57% from BBSRC, and 51% from MRC. The percentages of participants having completed data skills training is more uniform between funders than the percentages of participants having completed software skills training. This result is consistent with the analysis of Figure 10 showing that survey participants engaging with BBSRC and MRC, or who work in biological sciences and/or medicine and dentistry, are slightly less likely to develop their own software.

Survey data focusing on the programming languages used by participants to develop code also provided deeper understanding of some of the technical skills utilised. The results (see Appendix E) showed that the top three programming languages were Python 59%, Fortran 25%, and C++ 23%. These results are similar to the breakdown of languages utilised by RSEs in the UK: Python 87%, C++ 39%, C 37% (Philippe, 2018). While RSEs are a slightly different demographic to this study, it is notable that there is a strong overlap between the top languages identified by this survey and those of UK RSEs, as this means that it is likely that RSEs will be able to support researchers.

The survey also investigated where participants had acquired their skills and knowledge about data and/or software, with results shown in Figure 14.

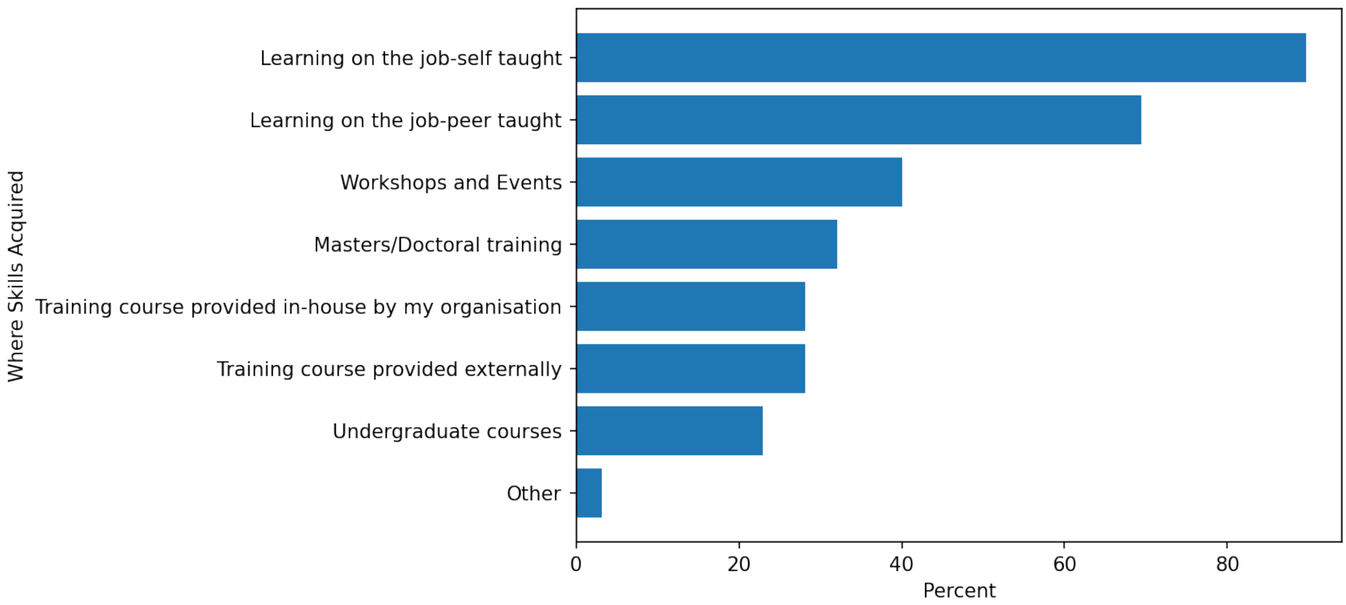


Figure 14: Where survey participants had acquired their skills and knowledge about data and/or software.

The majority of survey participants stated that learning on the job/self-teaching (e.g., websites, embedded help/training) was key, with learning on the job/peer taught (e.g., learning from project teams, members, colleagues, forums) also scoring highly. This was supported by perspectives offered within the group interviews. Participants mentioned the importance of undergraduate and postgraduate degrees in developing skills necessary for research software. These skills were described as having been developed alongside personal work and through additional learning opportunities such as online courses. Some examples are provided below.

Learning Python was mostly done not only through my undergraduate and masters, but also through my PhD. I managed to improve my skills to a great extent. In terms of C and C++ ... [I] learnt most of them through my undergraduate ... and my projects that I have been working on. (Early career stage, Computer science)

I started learning Fortran as a first year undergraduate student. The university in its physics degree taught Fortran as standard for all its research students ... but I didn't feel I started to get that good at it until I was doing a PhD in computational physics. (Mid/Recognised career stage, Physical sciences)

I first learnt MATLAB at university. I studied engineering in undergraduate; we used MATLAB in our course ... then I think I learnt Fortran alone. I followed some courses online ... a bit alone and a bit with paid courses or summer schools when I was a PhD ... I still attend now. (Early career stage, Physical sciences)

There were recommendations among the group interview participants for more embedded learning for students to ensure software training is provided in a research context rather than taught as an abstract skill. Increased learning about software within an applied context was described as essential prior to, or very early in, one's career:

At an institutional level, there's beginning to be ... optional enrichment courses that students can take in coding, but the point is they are bolt-ons ... they could be seen as rather tokenistic: you take the course, you tick the box, you put it on your CV, but then that's really the end of it. It's not embedded. (Early career stage, Multidisciplinary)

We've been doing Python in year one. All students do an intensive Python course. The problem is, by the time they get to year three or four, they've forgotten it all [...] And what we're trying to do is think of ways of embedding Python in the entire course ... so that they keep those skills up to date. (Mid/ Recognised career stage, Physical sciences)

If you're having to train every single PhD student from scratch ... with minimal computational skills, it's a lot of time and effort, whereas really these skills are going to be generally useful for all undergraduates in the STEM field [...] I think that in general we can do much better across all subjects, not just physics. (Mid/Recognised career stage, Physical sciences)

Further analysis of the survey considered how participants acquire their skills and knowledge in each career stage. In general, there is a slight decrease in training as career stage progresses (see Appendix G, Q23 & Q24 x Q6 and Q25 x Q6). Over the past five years, training courses provided in-house had been attended by 31% of Early career researchers as compared to 23% of Senior researchers, and similar figures are shown for external training (30% vs 24%). Workshops and events had been attended by 48% of Early career researchers but only 35% of Senior researchers. The exception is access to Masters and doctoral training, which declines more rapidly for participants who are more advanced in their careers. Data analysis training had been completed in the last five years by 88% of Early career researchers but only 67% of Senior researchers. Similarly, training in software development had been completed by 71% of Early career researchers but only 53% of Senior researchers.

Within the group interviews, participants explained how various facets of their careers allowed them to learn and develop skills related to research software. They mentioned the importance of computer science conversion courses and technical qualifications, informal self-teaching and peer learning, work placements and work experience, and formal (online) training courses. Participants' reflections upon skills and training are provided below:

I signed on for a part-time Business and Technology Education Council [course] in computer studies, where we learnt programming with Pascal, that's when I decided to do a conversion and masters course into computer science. (Female, White British, Mid/Recognised career stage, Multidisciplinary)	**Conversion courses and technical qualifications**	*I was offered the opportunity to join an MRC scheme. It was a three year training programme to be a computational biologist ... the idea was to take postdoctoral level wet lab scientists and train them to be principal investigators, computational scientists. (Male, White British, Mid/Recognised career stage, Biological sciences)*
I've had very little formal training. Most of the stuff I've done has been picked up on a job or by speaking to people or by working with people and just learning ... certainly in terms of programming or technical skills. (Male, White British, Mid/Recognised career stage, Multidisciplinary)	**Informal self-teaching and peer learning**	*I first started coding when I was doing my masters and I inherited some MATLAB code that needed altering. I just kind of picked up what I needed to do to get the job done. (Male, White British, Early career stage, Multidisciplinary)*
When I was working at [the commercial organisation] there was very little [training] and the only thing that happened there was ... we got sent on a residential course ... and we heard about Python and we thought we might use it. (Male, White British, Mid/Recognised career stage, Physical sciences)	**Work placements and work experience**	*I had a summer job working in the technology centre [of a manufacturer]. They had various in-house simulation models. It was getting involved in that where I actually learnt how to ... practise ... understand ... construct ... programming rather than just learning the syntax, which is what I learnt as an undergraduate. (Male, White British, Established/Experienced/Senior career stage, Physical sciences)*
When I started my postdoc I started learning Python, mostly on my own. I did end up attending a few courses which were a great help because when you learn on your own you're only looking at some specific problems that you're trying to solve; whereas sometimes these courses give a very wide outlook. (Female, Asian/British Indian, Early career stage, Physical sciences)	**Formal (online) training courses**	*I did take up a course online ... on Coursera on HPC ... when I was doing my PhD. At that time we didn't have actual courses in my institute on HPC ... so that's where I learnt about version control, Python and so on. (Male, Asian/ British Indian, Early career stage, Engineering and technology)*

One early career researcher in a group interview described how their desire for more applied learning led to considerations of how to factor new learning opportunities into personal career plans. This was described by the participant as "self-investment", the need to use personal resources perhaps reflecting the inadequacy of training provision in fixed term research contracts:

I can certainly benefit from more data science expertise ... it is becoming such an important need that I'm considering ... at the end of my latest fixed-term contract rather than moving onto another contract, I might have to self-invest some of my own time and money into getting training.

Other studies have noted the importance of ensuring that researchers can access a range of training types. Canada's Research Software Current State Assessment states that "[t]here is a shortage of training in research software, with a particular need for research software training in those domains that are newer to the use of digital research infrastructure in support of their research. Community-, domain-based, and regional training beyond degree programmes thus remains of paramount importance" (Digital Research Alliance of Canada, 2021).

A further survey question investigating what other skills are needed went as follows: "Thinking about your current skill set, what is the main data, software or computing skill that you would like to obtain to improve your research/work?" This was a free text question and clustering of the 286 responses with data (see Appendix E, Q26) identified the most common responses to be:

- ML, deep learning, AI and/or data mining (20%)
- Coding, programming improvement, good practices and/or documentation (18%)
- Python (12%)
- Data processing, management, access and/or sharing (12%)
- GPUs and/or CUDA (7%)

Within the group interviews, participants were given an opportunity to mention and describe activities in relation to software, skills, and computing that they would soon be undertaking and/ or that they would like to undertake. The most common answers were:

- Deep learning
- Geographic Information System (GIS) software
- Project management training
- Software Carpentry training
- Software Carpentry instructor training

Three of the stakeholders interviewed emphasised that new types of training will also be needed, particularly for future technologies: "As these technologies become available, such as quantum simulators, a user base will be needed that is already a little ahead of the curve - that already know that their algorithms will suit that kind of technology." Interviewees also suggested that more feedback was needed on what training users want, and the importance of feedback surveys at regular intervals to understand the impact in terms of changed practices. It was highlighted that a gradation of training opportunities is needed, from beginner to intermediate to advanced, all with a practical focus. One interviewee also highlighted the need for training on more basic digital skills: "There is a very large number of researchers who just need a Jupyter notebook and a well provisioned virtual machine and sensible practices for keeping their data long term."

Two of the interviewees commented broadly on how skills needs may change: "There is this huge technology convergence going on at the moment between supercomputing and data science. What hasn't happened yet is this convergence in software; that's where people come in. Data science staff are going to work more closely with the HPC staff and each will learn from the other. So it is not just the technology convergence we'll end up with, actually it's a research convergence of people that can span both worlds. And so people that are really, really good at dealing with data science projects, when they start needing more compute or larger datasets they will need to also feel comfortable moving over into the larger scale computing world, and vice versa."

The second stakeholder made similar comments: "We will need more people with different skills at different scales. Because that's where our researchers are now, at different scales. And the other thing we're going to need is more people with interdisciplinary domain skills. Because one of the things that the exascale era is going to enable is perhaps some of these codes, which would be run separately in the petascale era, will now need to be run together in a coupled sense. We're going to have different scales of computation running at the same time and people having experienced knowledge of how to make those things." This was summarised thus: "So as a community for the next five years, we need to decide to stretch ourselves over an even larger domain."

Disciplinary differences were also highlighted in the stakeholder interviews: "AHRC's community are doing more large-scale computing, as they get encouraged to do more AI and ML projects. And this is a community that doesn't live in the large scale computing world at all … so we have to think really carefully about how we start to move that community up the pyramid. But at the same time, we can't just present that community with where we are today, we need to think about how our world is going to change to look more like their world. Because their world is extremely rich from a research and software point of view, whereas our world is very focused on making best use of the resource, and somehow these two worlds are going to have to meet. And, and that means we're going to have to engage a lot with these new user communities."

Large-scale culture change is also needed in order to equip researchers with skills in other essential areas, such as data curation. Data curation maximises the usability of data from creation through to use, and this process can also enhance data quality (Freitas & Curry, 2016). One interviewee highlighted that while some infrastructures have personnel with the skill sets to assist with data curation, "we need to change the culture so that people are putting enough metadata on their research data to make it findable, to make it FAIR. We can provide the tools and the hardware to do that, and some people to support it, but the community has to engage and essentially make their data more easily available. And some communities are very good at this already. So we're building on their experience and expertise, and developing something which will be not too onerous, and will help other communities to do the same thing."

Further views on research systems, processes, and data stewardship were given within the group interviews in relation to the General Data Protection Regulation (GDPR). An early career researcher in biological sciences, medicine, and dentistry explained that "[w]e use a lot of data that's potentially identifiable and it has to be all GDPR secure. The infrastructure that I think is missing, mainly for us, is open data stores that meet those requirements. It's very hard for me to share data, which is obviously a problem for reproducibility." A similar issue was raised by an early career researcher in engineering and technology when describing the complexity of working with data for physiological modelling: "On the data side we aim to separate the sensitive data, the clinical GDPR-protected data, but [then] accessing that is a problem. These are curated by hospitals… getting a usable data transfer agreement took us more than a year, which was striking for me … it's easier to become a visiting member of an institute than it is to get a usable data transfer agreement." Others described the tensions created by the expectation that they provide access for secondary use of their data/software without adequate instructions on the appropriate infrastructure.

Barriers to training were also considered in the survey. Figure 15 shows the barriers for survey participants to gaining skills that they would like to acquire to improve their research/work.

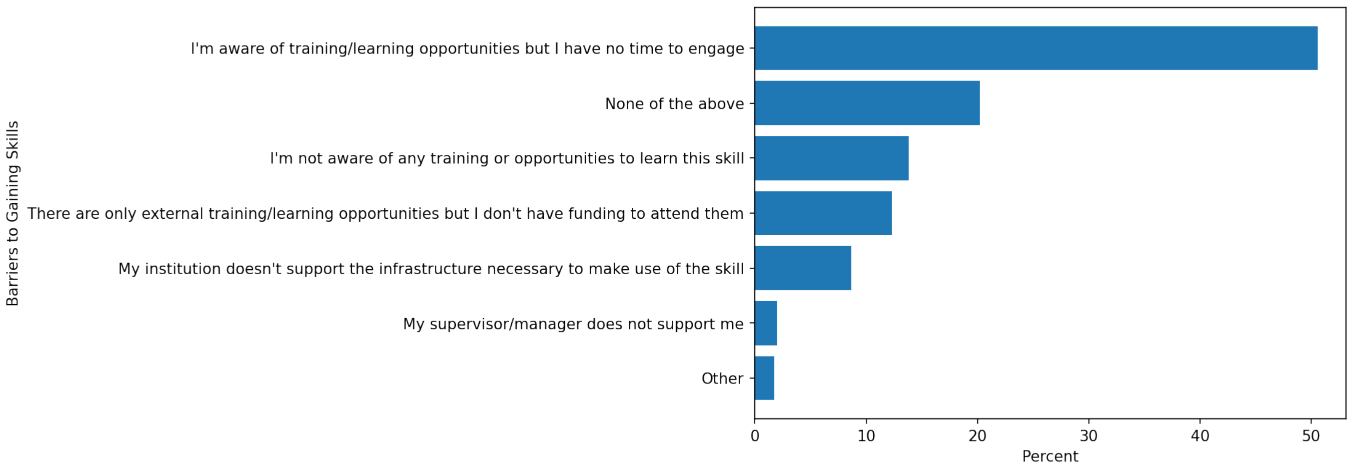


Figure 15: Barriers for survey participants to gaining skills that they would like to acquire.

Figure 15 shows that "time to engage" is the primary barrier (with 25 of the 31 "Other" responses also relating to lack of time). The remaining six responses to "Other" included clarification that training is often either too generic or not easily applicable to specific user requirements. Often, what is required is one to one mentoring to achieve a minimum working example.

Some of the reasons that time for learning is not prioritised came to light in the group interviews. There were recollections that it was unusual to have time to become sufficiently familiar with new programming languages/version control. Furthermore, although some participants were aware of ongoing training, they had limited time to participate, or at least for the training to be of optimal value. Within the context of academic software training, it was stated that this can be limited by the availability of teachers who can lead the training, as well as insufficient funding.

Sometimes we just don't have the time because of other duties. I've been wanting to do a [free and online] machine learning course, but I can't seem to find the time. (Female, Early career stage, Physical sciences)

I wouldn't say that it's not available [training for ARCHER2], it's just that it's not focused on what I would require. It would be a bit more diverse and I don't have time to invest in a full workshop.(Male, Early career stage, Engineering & Technology)

There is hardly ever a time to learn [changes from Python 3 to Python 4] properly. We just pick it up and try to do what we were doing before but it doesn't lead to optimal results. (Male, Mid/Recognised career stage, Engineering & Technology)

[Our software] lacks things that we would like it to be able to do but it would take considerable time and money to develop so it's not something I can do when I feel like it. (Male, Mid/Recognised career stage, Multidisciplinary)

Figure 16 further considers the barriers to gaining skills that survey participants identified, by analysing this against their career stage.

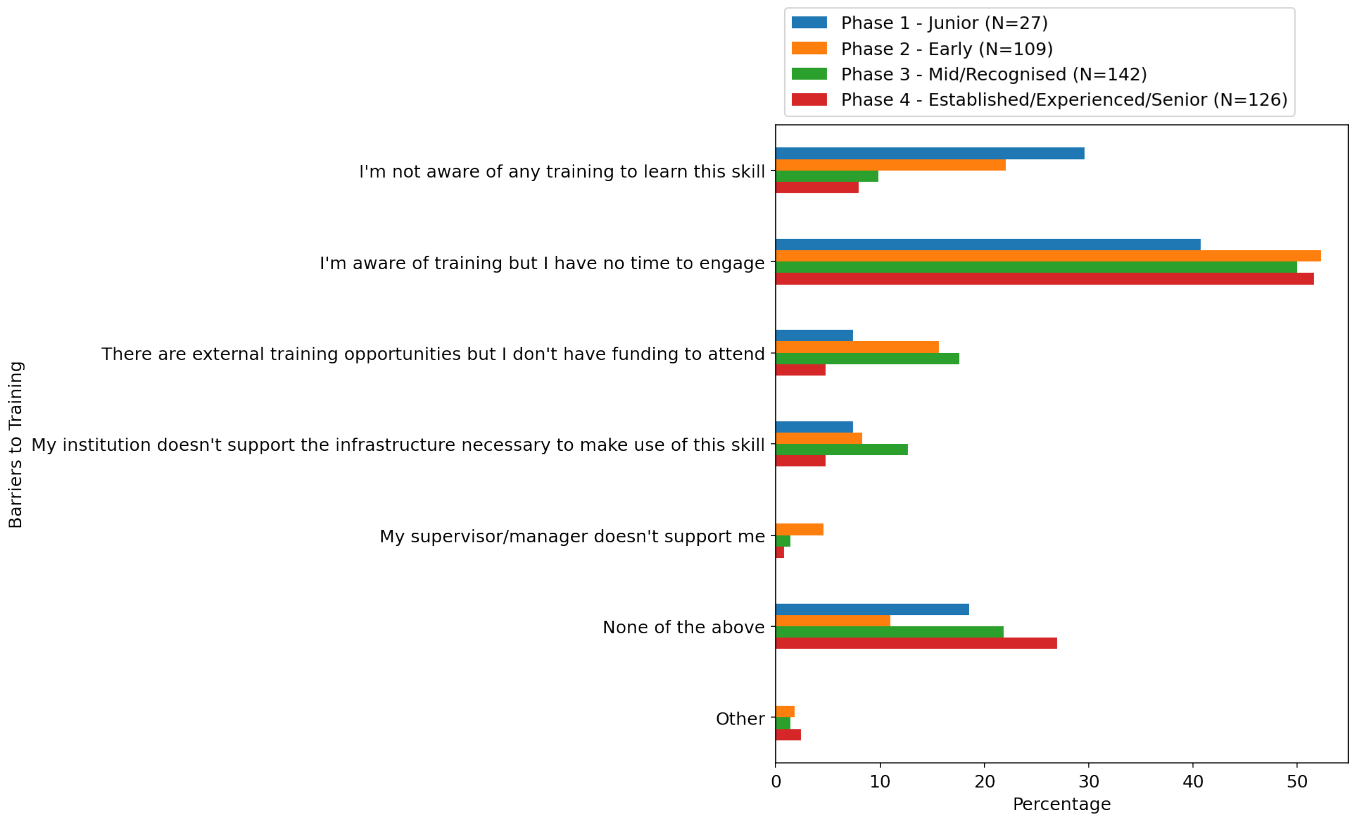


Figure 16: Barriers to gaining skills that survey participants identified, by career stage.

Lack of awareness of training opportunities is a more common barrier for Junior (30%) and Early career (22%) researchers than for Mid (10%) or Senior (8%) researchers. The most common answer, "I'm aware of training but I have no time to engage", is not significantly affected by career stage, suggesting that researchers at all levels feel that they do not have time for training. This articulation of time pressures echoes studies of the intensification of academic labour (Houston et al., 2006) and the prioritisation of some tasks (publishing) over others (skills development) created by infrastructure such as the REF. The result is challenges for professional development and training initiatives (Harman & Darab, 2012; Maisuria & Helmes, 2019). This data suggests that there may need to be a more radical rethink around how professional development opportunities are both offered and incentivised. Training/upskilling for later career researchers could perhaps be more compact, targeted, and incorporated into work streams.

Another frequent barrier to upskilling the workforce is limited training capacity. One stakeholder interview highlighted the value of the 'train the trainers' model that organisations such as the The Carpentries and SSI utilise, but notes: "The challenge, though, with that is as that knowledge and training is distributed, somehow there needs to be an income stream to the people that are doing the training." An OECD study adduced this as one of five key areas that must be considered in creating digital workforce capability for data-intensive science. Common issues cited included the voluntary nature of the workforce (resulting in difficulties in scaling up to meet the ever-increasing demand); that trainers are usually not certified; and that trainers often do not receive recognition for their expertise and contributions in training (OECD, 2020). The group interviewees raised the same points:

> *The graduate school would love to do much more software training than they do but their issue is funding and/or academics' time, or getting people to teach it.*

> *[The Faculty] get constant requests to provide Python training, but they don't have the funds or can't find teachers who will do it ... their teaching is for undergraduates. They don't have the time or it's not in their remit to be able to do the training for researchers.*

A 2021 Australian report estimated that in 2021 there was one staff member per 118 researchers in Australian research organisations providing advice and training on relevant tools and applications. Though insufficient, this marked an improvement on the 2019 figure of one per 200 researchers (Barker & Buchhorn, 2022). An interviewee suggested that RSEs can assist with this challenge in scaling training by providing some instruction in situ: "When we work with people, we always try and leave a bit of ourselves with them ... we'll give them training as to why we've done that, and how we've done that, and how they can continue to interact with it ... then we'll hand them back the stuff that we've done in a position where they can continue to develop it if needs be." It was also suggested that some RSEs could focus more on providing guidance to others, particularly for less computationally intensive research, perhaps by mentoring postdoctoral and PhD students. Better engagement with industry was also discussed as a mutually beneficial opportunity for training, since it would ensure that researchers are familiar with industrial hardware.

In one stakeholder interview, staffing was identified as a potential bottleneck: "There are many large communities who have lots of excellent researchers doing world-leading research. But in terms of the person who could move the code from one generation of hardware to the next, there might be as few as one or two people who can do that. And that's a huge risk for the UK science programme, because these are often areas where the UK is world-leading, the software is one of the key differentiators of the UK community. And it would be much better if more of the community could support it."

Major findings:

- Participants from biological sciences, and medicine and dentistry (and/or who engage with BBSRC and MRC) have a slightly different skills profile to those from computer and mathematical sciences, engineering and technology, and physical sciences (and/or who engage with EPSRC, NERC, and STFC):
    - Are slightly less likely to develop code.
    - Have upskilled less in relation to computing in the last five years.
    - Are slightly less likely to develop their own software.
- Degrees are important in developing the skills necessary for research software. However, these skills need to be more embedded in curricula.
- Skills were developed alongside personal work projects and through additional learning opportunities such as online courses. A significant amount of learning is on the job/self taught, with learning on the job/peer taught also important.
- 20% of participants would like to improve their skills in ML, DL, AI, and/or data mining, and 18% in coding, programming improvement, good practices, and/or documentation.
- Time is the biggest barrier to upskilling across all career stages, and is most challenging for Junior and Early career stage staff. This is affected by pressures from teaching, research, administration, management, etc.
- Training needs to be developed in sustainable and scalable ways to ensure an adequate supply of trainers.

4.2.3 Diversity of personnel

The UK academic, software development, and RSE workforces are less diverse than the UK workforce in general, as explored in section 4.1 in relation to Table 1. This subsection explores how different personnel may have different experiences. As noted in the section on this study's limitations, throughout this report the analysis of data by gender is restricted to participants who described their gender as male (including trans men) or women (including trans women), due to the low number of respondents who identify as non-binary. While non-binary had also been an option on the survey, response numbers were too low to enable meaningful analysis—a common problem when trying to identify challenges faced by under-represented groups.

Survey participants' responses to the question on whether they write their own code, analysed by gender, reinforces the lack of diversity indicated in Figure 17.

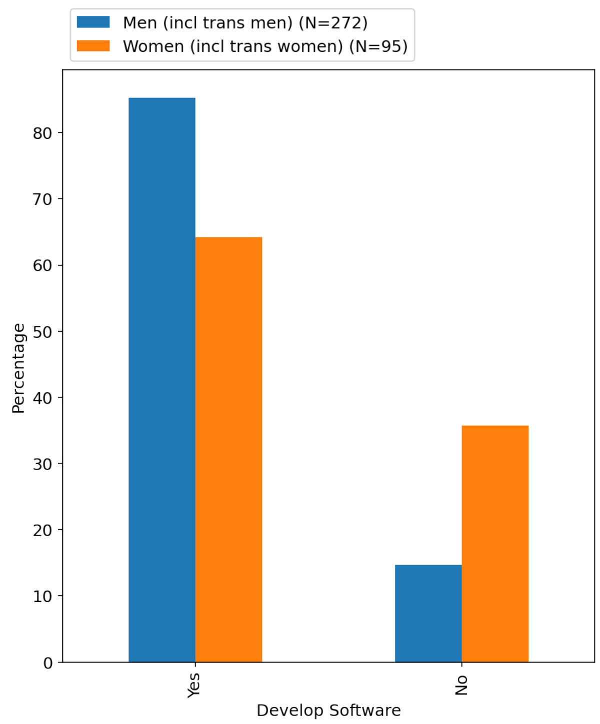


Figure 17: Survey participants who write their own code, by gender.

Figure 17 shows that writing code is much more common for men (85%, against 64% for women). Large-scale computing: the case for greater UK coordination notes that very few initiatives exist to address diversity issues in the large-scale computing workforce specifically, and highlights Women in HPC as a global network. Established by EPCC, this organisation seeks to improve gender diversity by supporting both women in large-scale computing and the organisations who employ them (Government Office for Science, 2021).

While there was little difference in some areas, such as the breakdown of participants' primary roles by gender (see Appendix G, Q5 x Q29), variation was evident in certain areas. As shown in Figure 18, there are differences in responses by gender to the question: "Over the last five years, have you learned new skills or knowledge in relation to any of the following?"

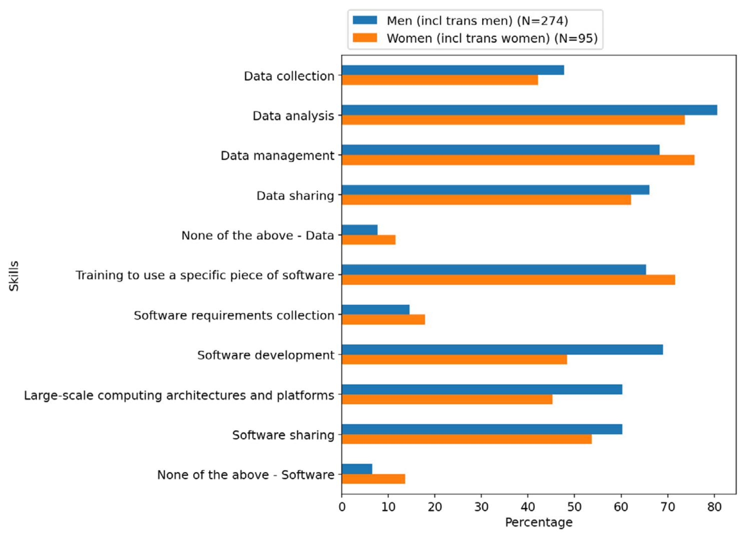


Figure 18: New skills or knowledge learned in the last five years by survey participants, by gender.

Women (including trans women) appear to receive less training than men (including trans men). Women are 21% less likely to receive training in software development and 15% less likely to receive training in large-scale architectures and platforms. Two other studies show similar results. Comparison of US results and a survey of US postdoctoral students (Nangia & Katz, 2017) showed a gap between training for men (63% in the UK and 63% in the US) and training for women (39% in the UK and 32% in US) (Carver et al., 2022). However, the picture is nuanced, since women receive slightly more training in data management, the use of specific pieces of software, and software requirements collection. Canada specifically identifies the lack of diversity in the research community as one of its key challenges and opportunities at the national level, and suggests responses such as "partnering with technical training and networking initiatives for under-represented groups, as well as supporting EDI-oriented training programmes related to research software use and development within particular research communities will help expand the talent pool" (Digital Research Alliance of Canada, 2021).

There are initiatives in the UK research community that consider this issue. Trialling approaches for increasing diversity and inclusion at events, including workshops and conferences, has been suggested as a useful way forward, alongside evaluation of the relative success of these approaches and their contribution to improving diversity within the computational science developer community (Chue Hong et al., 2021). A range of UK events are already embracing this. For example, the 2021 Collaborations Workshop organised by SSI highlighted diversity and inclusion as one of its three main themes (Laird, 2021), and sought to improve access through measures such as live transcription and financial assistance for members of under-represented groups, students, early career stage researchers, and others who would not otherwise have been able to attend or fully participate in the event (Ainsworth, 2021).

Slight differences in responses by gender can also be observed in the answers shown in Figure 19 on issues preventing survey participants from gaining the main data, software, or computing skill that they would like to obtain to improve their research/work.

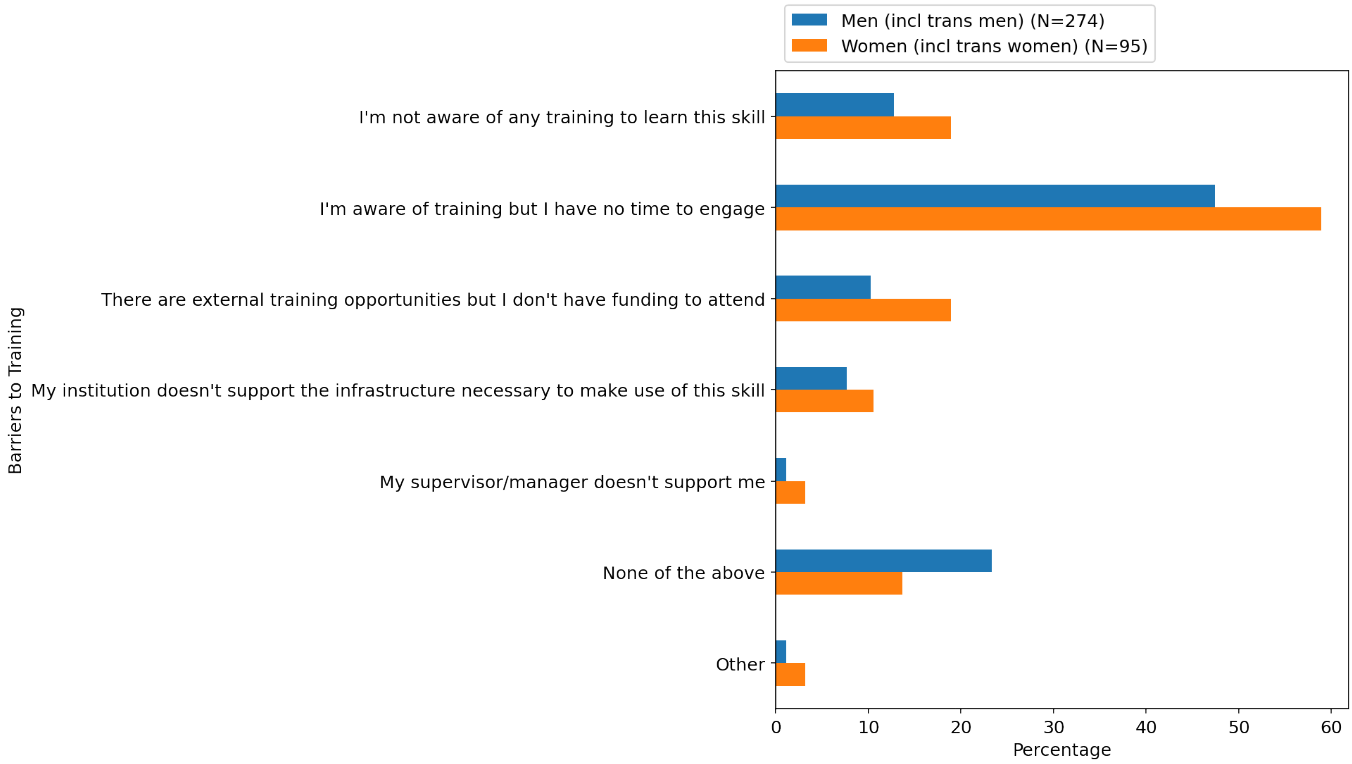


Figure 19: Issues preventing survey participants from gaining skills, by gender.

This shows that women (including trans women) more commonly experience obstacles due to all of these issues. This is consistent with the conclusion from Figure 18 that women are learning fewer new skills, and gaining less knowledge: 18% of women reported that they are aware of external training but do not have funding to attend, as compared to only 10% of men.

There are also some differences in practices related to sharing of research software, as indicated in Figure 20, which shows responses to the following question: "Over the last five years, which of the following practices have been part of your standard research process? Select all that apply."

Figure 20: Practices that have been part of respondents' standard research processes over the last five years, by gender.

Figure 20 shows that women (including trans women) are less likely than men (including trans men) to report sharing their data. The greatest difference between genders was for the response "I share my research software publicly", which 67% of men but only 41% of women indicated was the case. The results of confining responses to those who have applied for funding that explicitly includes costs for software development is shown in Figure 21.

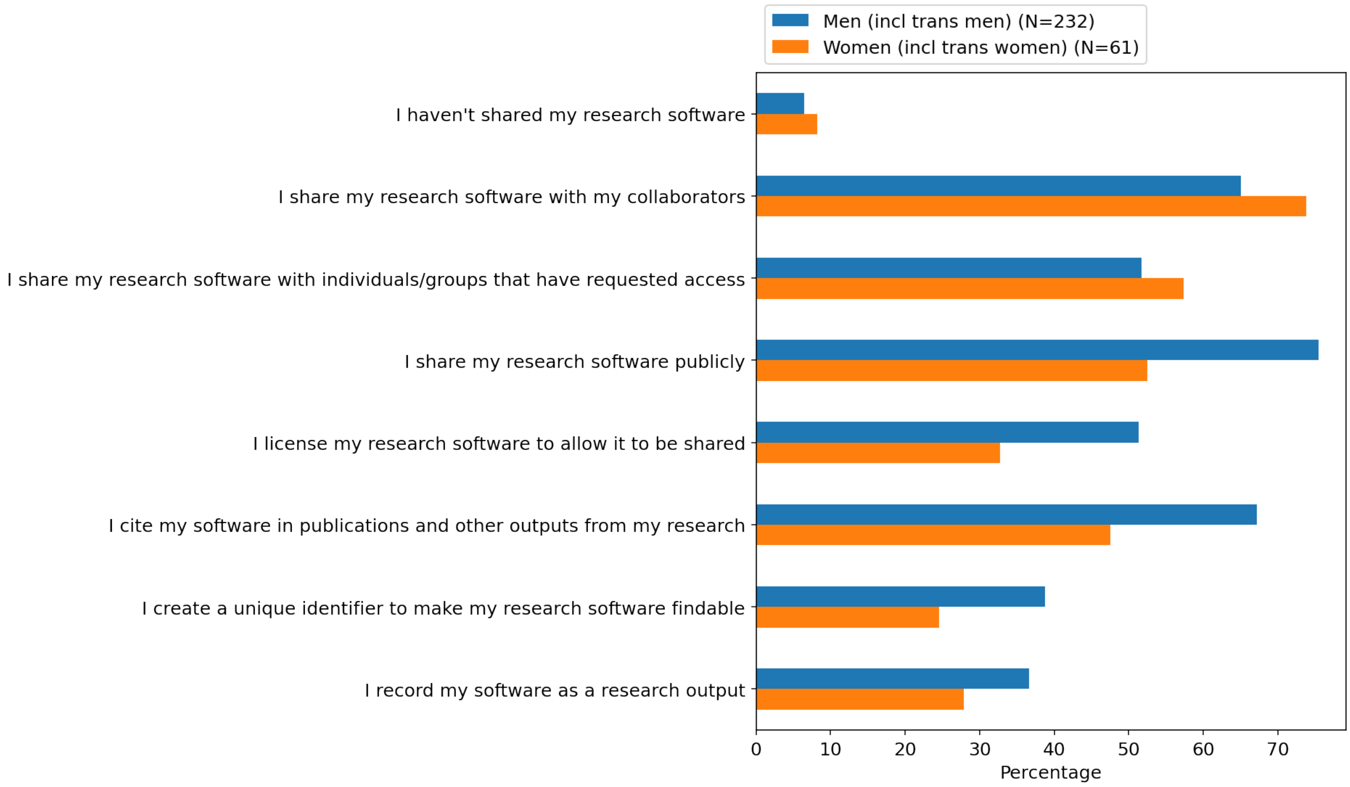


Figure 21: Practices that have been part of standard research process over the last five years for survey participants, by gender, limited to only those who have written their own software.

When including only survey participants who develop software, women were less likely than men to report sharing their data in areas such as "I share my research software publicly" and "I license my research software to allow it to be shared" (Figure 21). The greatest difference between genders was shown in responses to the statement "I share my research software publicly". 75% of men but only 52% of women indicated that this was the case (compared with 67% of all men and 41% of all women who reported sharing their research software publicly).

Further analysis attempted to establish whether there were gender differences in applying for funding. This showed that 47% of men (including trans men), and only 35% of women (including trans women) who write their own code have applied for funding that explicitly includes costs for software development (see Appendix G, Q12.b x Q29). However, there is too little evidence to support this difference at the 5% significance level.

Further breakdown of applications for this type of funding by career stage, examining only those who have applied for funding that explicitly includes costs for software development, is shown in Figure 22.

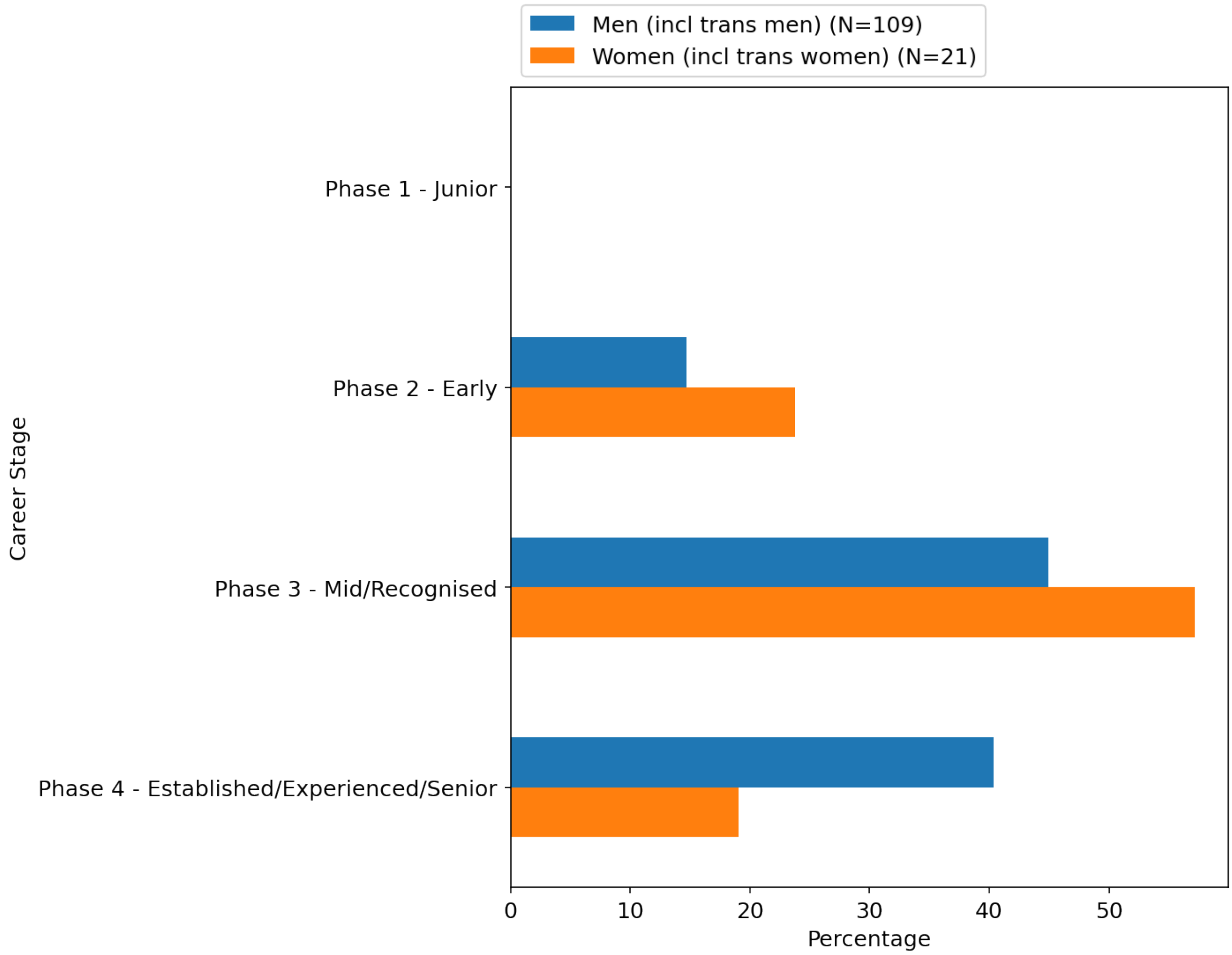


Figure 22: Survey participants that have applied for funding that explicitly includes costs for software development, by career stage and gender.

This shows a greater percentage of women than men applying in Early and Mid/Recognised career stages, and a greater percentage of men than women applying at Established/Experienced/Senior career stages. Multiple studies have demonstrated that women are more likely to receive lower research funding than their male colleagues (Nogrady, 2019; Oliveira et al., 2019; Wijnen et al., 2021). In the UK, gender disparities have been highlighted in an EPSRC report, Understanding our portfolio: a gender perspective, across a 12-year period. It revealed that:

- EPSRC receives consistently low numbers of applications from women across their portfolio.
- Women are under-represented in EPSRC's principal investigator applicant pool.
- Application numbers from women for large grants are particularly low at just 6% (EPSRC, 2020).

Gender differences were also considered in relation to future infrastructure needs. Analysis of whether survey participants felt that their research would be limited by access to computational and/or data research infrastructure, further broken down by gender, showed little difference (see Appendix G, Q19 x Q29).

Further survey questions sought to understand how decisions on infrastructure usage are made. Figure 23 shows the responses to the following question, broken down by gender: "Who do you ask for advice on choosing the computational and/or data research infrastructure (computing, storage and network) you use?"

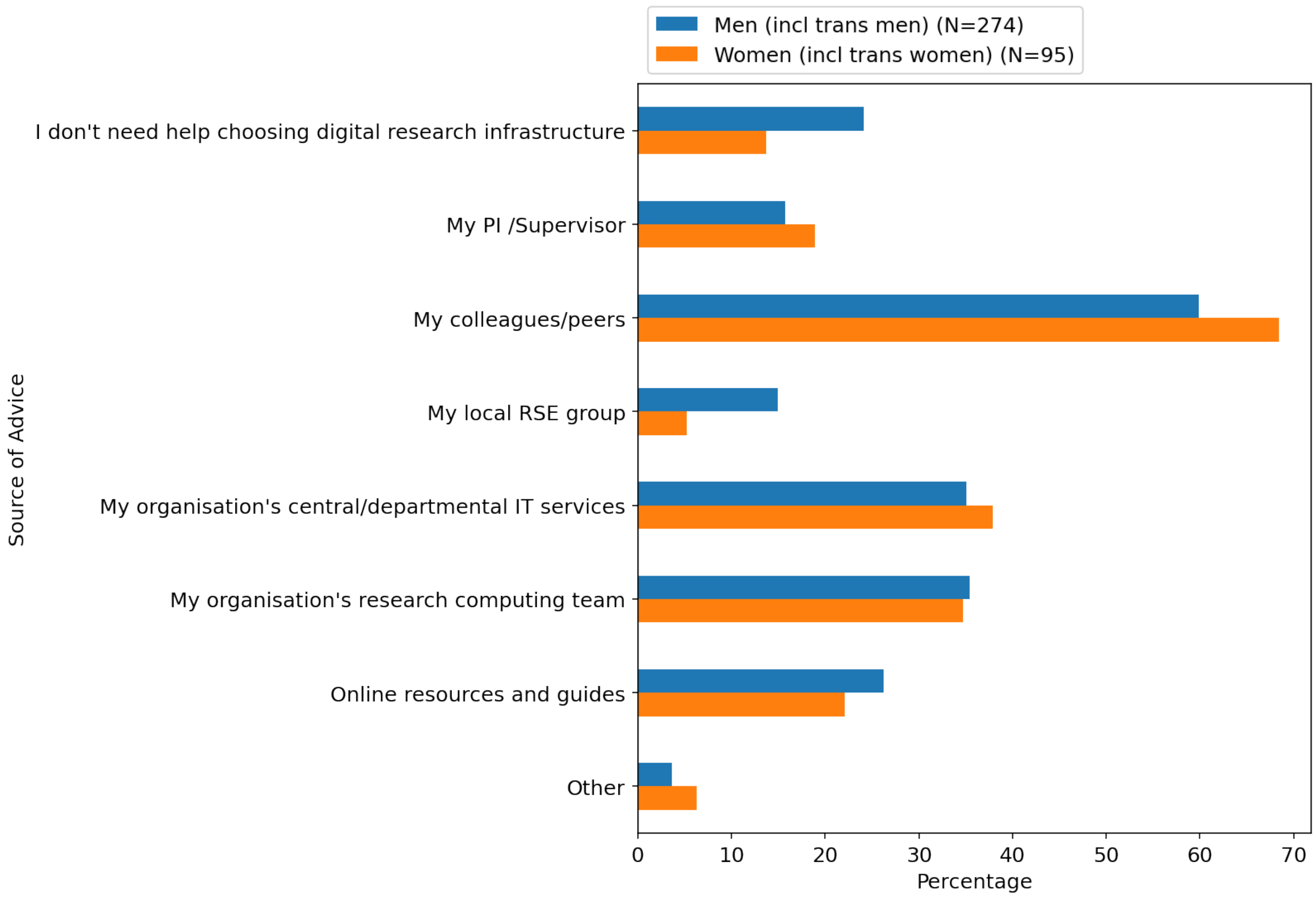


Figure 23: Who survey participants ask for advice on choosing the computational and/or data research infrastructure they use, by gender.

This shows that women (including trans women) are less likely to engage with their local RSE group (women 5%, men 15%), although this is a small sample size. Further investigation may be warranted to ensure that all genders are well supported in their use of research software.

Lack of data related to EDI in the research software community is a commonly cited issue. The US Department of Energy circulated a Request for Information on Stewardship of Software for Scientific and HPC. The potential scope of the stewardship activities included training on software development and use, workforce support, infrastructure, and project support. A review of 37 independent responses identified challenges in building a diverse workforce and maintaining an inclusive professional environment that encourages improvements in both recruiting practices and pipeline challenges (Finkel, 2022; responses to the Request for Information on Stewardship of Software for Scientific and High-Performance Computing, 2021).

Suggested areas for focus in the UK include understanding entry paths and "[e]xploring why levels of diversity among RSEs are lower than in many of the areas of study and research that already feed into RSE careers, such as Computer Science and Physics and Astronomy. What are the levels of diversity in other career paths for these subjects? What influences the career choices that individuals make and are there specific aspects that steer them away from an RSE career?" (Chue Hong et al., 2021)

Within the UK, there are also recommendations on improving the pipeline that can provide more research software personnel. Large-scale computing: the case for greater UK coordination suggests that "initiatives which look to increase diversity and foster inclusion in Science, Technology, Engineering and Maths (STEM) fields, or in the IT sector will likely impact on the large-scale computing workforce". These suggestions include increasing uptake in primary, secondary, and tertiary education; alternative entry pathways including traineeships and apprenticeships; and improving job retention, pay rates, and career progression for employees from under-represented groups. This report also highlights the Athena SWAN and Race Equality charters as ways to achieve change in the higher education and research sectors (Government Office for Science, 2021). A BCS landscape review recommends that a task force should be established to understand, examine, and report on access and participation in Computer Science qualifications for learners across key demographics (BCS, 2022).

Major findings:

- Women (including trans women) are learning fewer new skills and/or gaining less knowledge in some areas than men (including trans men).
- Women and men (including trans) seek advice on choosing computational and/or data research infrastructure slightly differently, although this is a small sample size and requires further study.
- More detailed analysis on EDI in the research software community is needed in many areas.

4.3 INFRASTRUCTURE

This section explores evidence for the support required for software across two subsections:

> Software usage: what software is being used, how it is being found and stored?

> Computational infrastructure: what are large-scale computing usage and future needs? How are infrastructure choices made? Are there differences in the infrastructure requirements of researchers across different disciplines and career stages?

Key findings:

- 97% of survey participants see software as important to their own research, with 85% citing it as essential.
- The large drop-off in the number of researchers using larger scale computing infrastructure compared to university-level resources needs to be addressed.
- 27% of participants believe their research will be limited by access to computational/data infrastructure in the next year, rising to 35% of participants at the most senior career stage.

4.3.1 Software usage

This subsection considers software usage. The definition of software provided in the survey (as contained in Appendix B) is any software or digital tool used in the course of research that has assisted in that research or helped to produce a research output (e.g., a publication). This can be anything from a short script to a fully-fledged software suite or specialised toolset. This is similar to an international community-derived definition of research software from the FAIR Principles for Research Software: "Research software includes source code files, algorithms, scripts, computational workflows and executables that were created during the research process or for a research purpose. Software components (e.g., operating systems, libraries, dependencies, packages, scripts, etc.) that are used for research but were not created during or with a clear research intent should be considered software in research and not Research Software. This differentiation may vary between disciplines" (Chue Hong et al., 2022; Gruenpeter et al., 2021).

To assess the role of research software in the UK research landscape, participants were asked how important software was to their research. Responses are summarised in Figure 24.

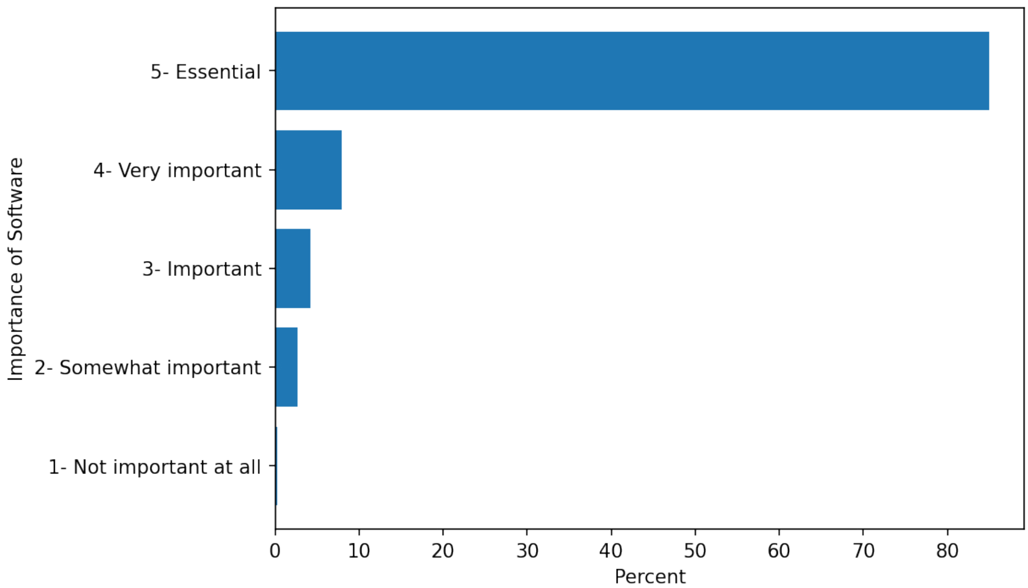


Figure 24: How important software was to the research of survey participants.

This evidence that 97% of survey participants see software as important to their own research (with 85% considering it Essential, 8% Very Important 8%, and 4% Important) is a valuable addition to the limited data on research software usage in the UK. In 2014, a survey of researchers at 15 Russell Group universities on their software use found that 92% of academics used research software, and 69% said that their research would not be practical without it (Hettrick, 2014). It is to be expected that our own more recent survey showed a greater reliance on research software, since research has grown increasingly data-intensive. However, this could also be a reflection of the high number of participants from physical sciences, mathematical sciences, engineering and technology, and computer science being examined in this report. Additionally, the 2014 survey focused on researchers, while this survey targeted participants who use, or support the use of, research computing—e.g., a user, researcher, RSE, infrastructure provider, or trainer/educator. Multivariate analysis of the same data in relation to funders that those participants had applied for and/or received funding from showed little difference (see Appendix G, Q8 x Q7).

Research software citation is an emerging method for improving data on software usage, and the work of initiatives such as the FORCE11 Software Citation Implementation Working Group to promote the software citation principles (Smith et al., 2016) is helping to improve citation standards in research software (Park & Wolfram, 2019), which are presently suboptimal. For instance, only 27% of ecology articles that conduct some type of statistical analysis and/or simulations are accompanied by the underlying analytical code (Culina et al., 2020). In addition, there are many research software repositories and/or registries that could theoretically provide some information on software usage, including disciplinary examples such as bio.tools for the life sciences (Ison et al., 2019) and the Astrophysics Source Code Library (ASCL).

The next stage of the analysis sought to understand how survey participants discover the software they use in their research, with participants invited to choose more than one option. The most common responses were "Recommended by colleagues" (87%), "Heard about it in a paper / talk / event" (75%), "Search engine" (51%), and "Recommended / provided by my organisation" (42%) (see Appendix E, Q20). Only 3% responded that their research did not need new software, further underlining the importance of research software for modern researchers. These results partly echo those of a study of experienced researchers, of whom the majority were involved in software development, and who worked primarily in the physical, computing, mathematical, and biological sciences (Hucka & Graham, 2018). The participants in that study identified the following as their top answers on how they searched for ready-to-run software and/or source code:

17. Search the Web with general-purpose search engines.

18. Ask colleagues.

19. Look in the scientific literature.

Many initiatives exist to make research software more findable, including the development of FAIR Principles for Research Software (Chue Hong et al., 2022), which a number of organisations are beginning to implement (Martinez-Ortiz et al., 2022). There is continuing development of a range of research software repositories and registries, including those that are focused on a particular geographic area, discipline, organisation, and/or programming language (NLeSC, n.d.). The fourth and fifth ranked responses on how to find source code from the Hucka and Graham survey were to search in public software project repository sites such as GitHub and to look in social help sites such as Stack Overflow (Hucka & Graham, 2018).Figure 25 shows the same data further broken down by funder.

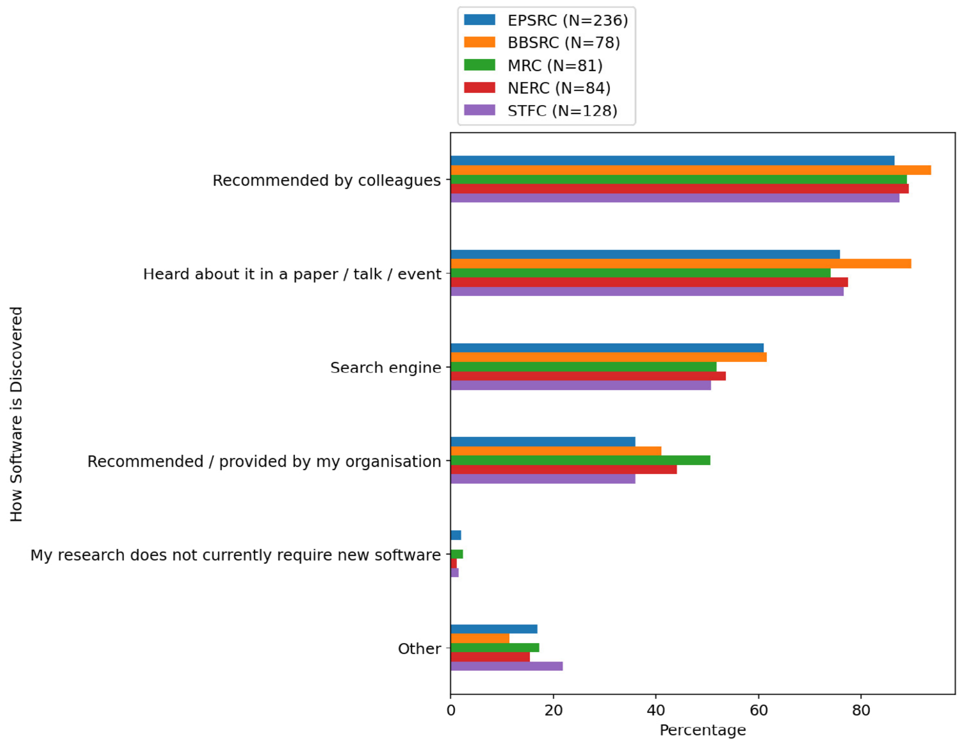


Figure 25: How survey participants discover the code they use in their research, by discipline.

This shows that participants involved with the BBSRC were more likely than participants funded by other organisations to discover software by hearing about it in a paper/talk/event (BBSRC: 90%, EPSRC: 76%, MRC: 74%, NERC: 77%, STFC: 77%) It is unclear why this might be, and further investigation may be warranted to consider whether this reflects factors such as investment levels in training, infrastructure, and networks. In addition to these results, analysis of the 51 free text responses from participants who nominated "Other" revealed that many contained references to software written by themselves and/or collaborators. Another breakdown of how survey participants found software alongside the disciplines in which they work (focusing on disciplines associated with BBSRC, EPSRC, MRC, NERC, and STFC) showed only small differences (see Appendix G, Q9 x Q4).

The next survey question asked: "What are the names of (up to three) pieces of research software you've used to support your research in the last month?" The most common responses were Python, MATLAB, and R (see Appendix F), suggesting that participants answered mainly in terms of programming languages. The question had aimed to elicit the names of the software that survey participants believed was most important to their work, but did not achieve this. An Australian survey that asked research organisations to list research software developed and/or maintained by their staff resulted in a list that included 50 GitHub repositories, 10 bitbucket repositories, and R packages (Barker & Buchhorn, 2022).

Figure 25 shows the results of the next survey question: "Over the last five years, which of the following practices have been part of your standard research process? Select all that apply."

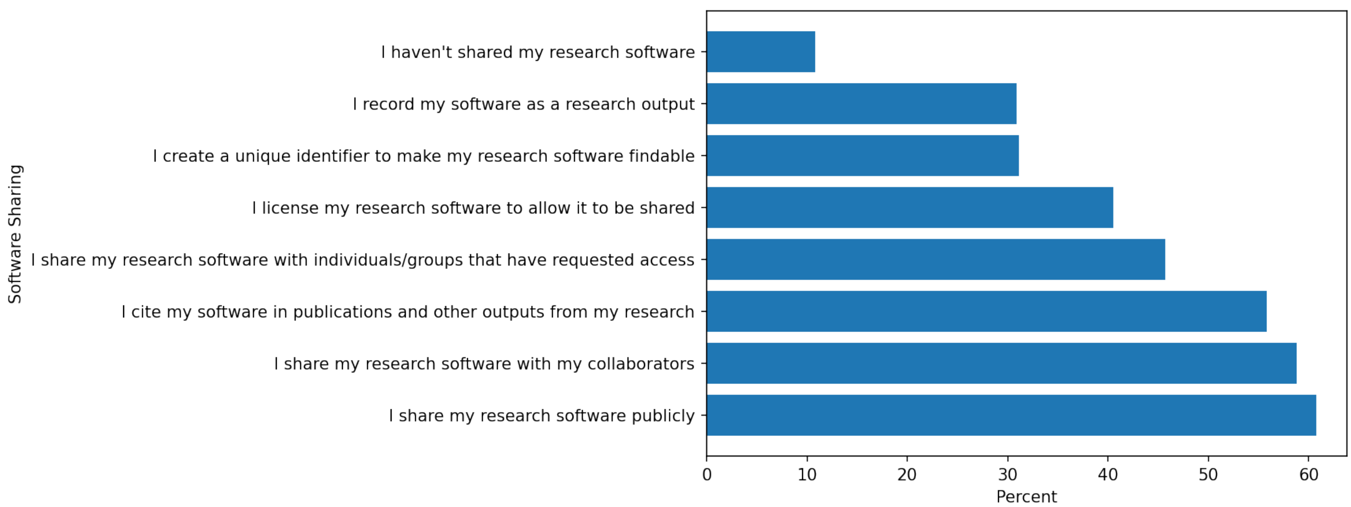


Figure 25: Survey participants' standard research process practices over the last five years.

This data demonstrates that good practice is being observed in the UK research sector, with 61% of respondents sharing their software publicly. This data is a valuable addition to our understanding of research software citation practices, enabling us to formulate further improvements. Analysis of 2022 UK RSE data found that only 38% of UK RSEs cite their software all the time (an increase of 4% from 2018), with just 12% using a digital object identifier (an increase of 2% from 2018). However, 39% use an open source licence (Hettrick et al., 2022).

Further analysis of the report data on practices that have been part of participants' standard research processes showed very little variation by career stage (see Appendix G, Q14 x Q6).

The survey asked a similar question with regard to research data practices: "Over the last five years, which of the following statements best apply to how you share your research data? Select all that apply." The answers are shown in Figure 26.

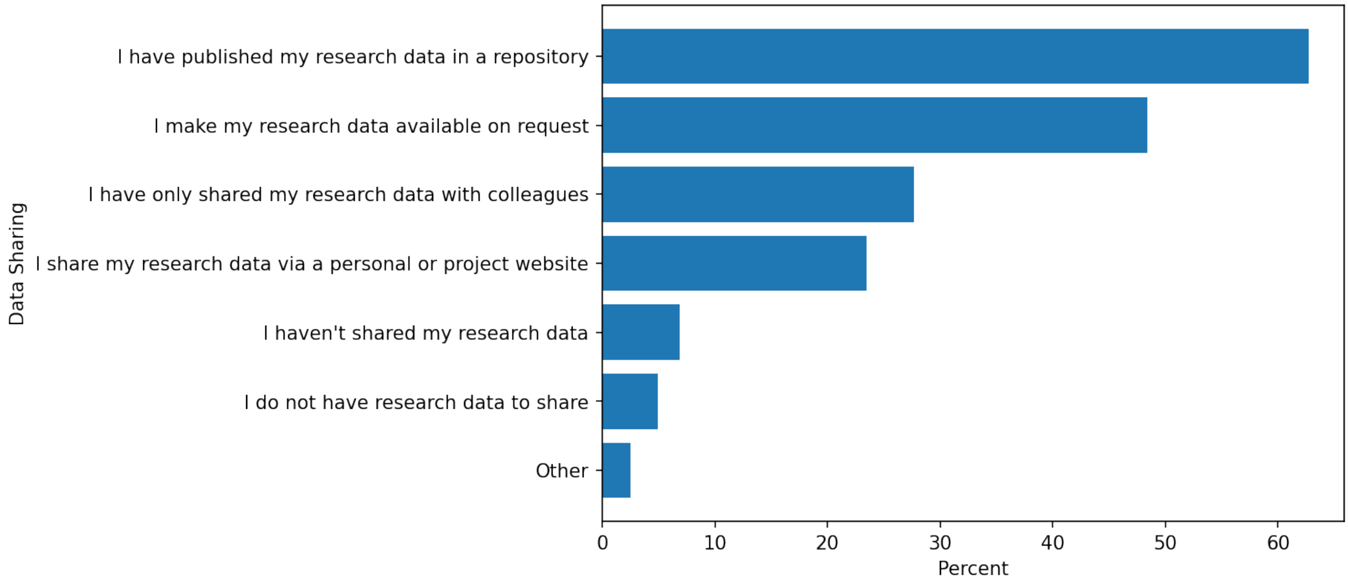


Figure 26: How survey participants have shared their research data over the last five years.

This shows that 63% of survey participants have published their research data in a repository, which is a positive result when compared to other studies. For example, the State of Open Data 2021 survey found that "[a]lmost half (46%) of 779 life sciences researchers responded that they share their research with the public using institutional repositories, followed by external repositories (e.g., Figshare, Zenodo) at 39%, cloud file sharing (e.g., Dropbox, Google Drive) at 20%, funder repositories at 19%, blogs/ websites at 14% and other at 13%" (Kipnis, 2021). The State of Open Data 2021 survey identified a range of barriers to data-sharing from 4,500 responses: 43% had concerns about misuse of data; 39% cited not receiving appropriate credit or acknowledgement; 35% were unsure about copyright and data licensing; and 25% were unsure whether they had permission from their funder or institute (Goodey & Hardeman, 2021).

Figure 27 shows responses to the final survey question: "Over the last five years, once you've completed a research project/ study, which of the following have you used to store the data that were generated after your research has been completed? Select all that apply." These responses are broken down by career stage.

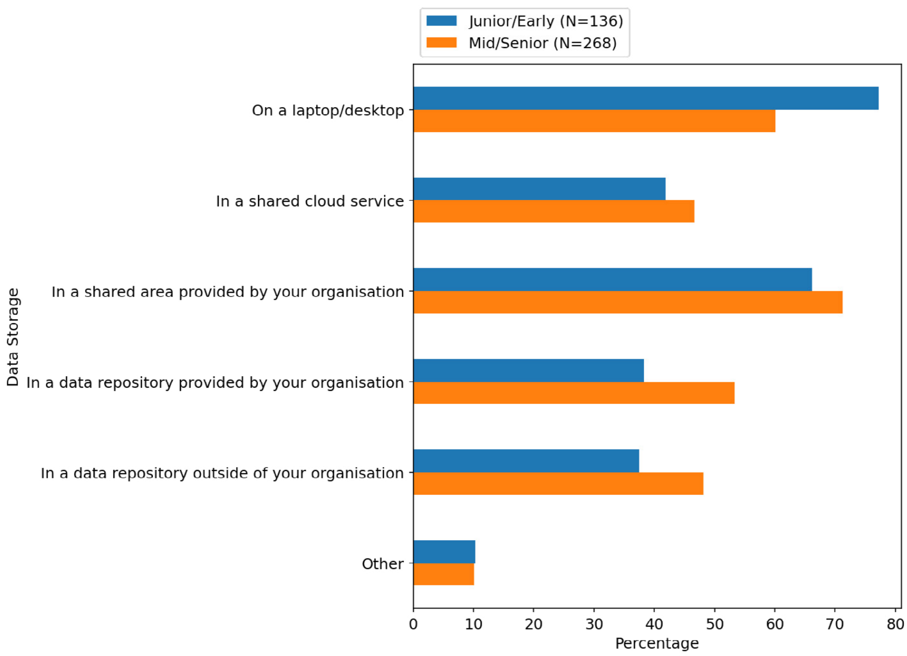


Figure 27: Where survey participants have stored data generated after a research project/study completed over the last five years, by career stage.

This shows that researchers at the Junior and Early stages of their careers are more likely to use a laptop (77%) than their Mid and Senior career counterparts (60%), and less likely (42% vs 47%) to use a shared cloud service, a shared area provided by their organisation (66% vs 71%), and/or a data repository provided inside (38% vs 53%) or outside their organisation (38% vs 48%). It is reassuring that only 13 of the 404 survey participants reported only storing data on a laptop/desktop. A similar breakdown of this data by funder is also provided in Figure 28.

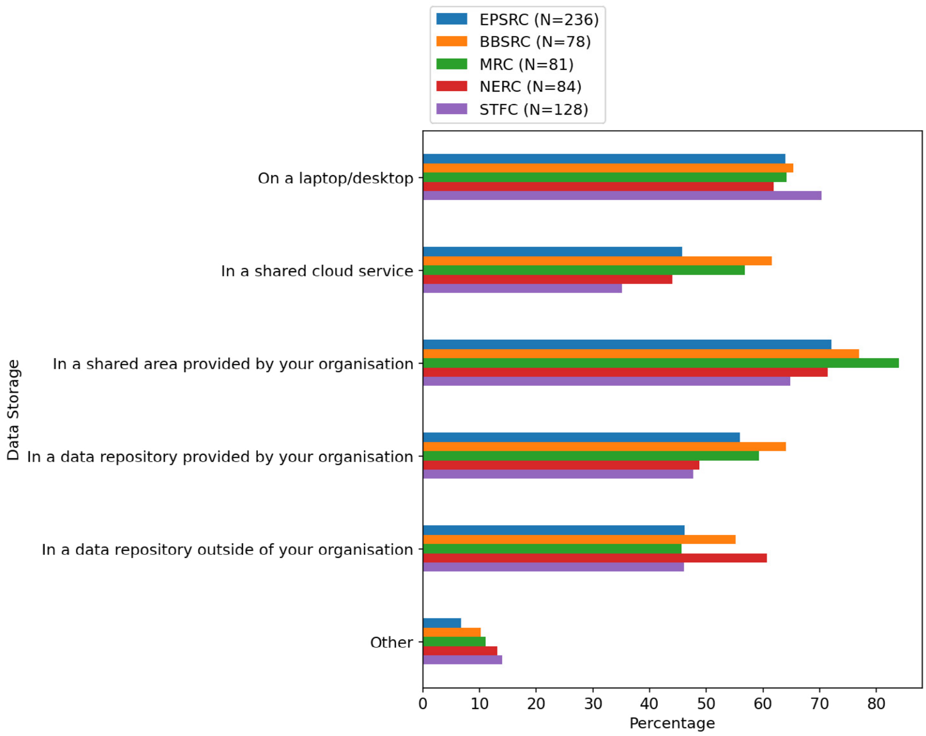


Figure 28: Where survey participants have stored data generated after a research project/study completed over the last five years, by funder.

Researchers who are funded by BBSRC are more likely to store data in a shared cloud service (62%) than those funded by STFC (35%). BBSRC-funded researchers are also more likely to store data in data repositories provided by their organisations (64%), as compared with researchers funded by NERC (49%) and STFC (48%).

One interviewee emphasised the need for more investments in both software and the RSEs to support it: "We want the codes to become more efficient. And that means more energy efficiency, as well as everything else. Which requires us to move to hardware, which is actually more performant, but harder to use. The typical researcher is not going to be able to get all the way to fully optimised code. A lot more investment in RSE teams is vital."

There are multiple stakeholders supporting research software and its personnel, and alignment of their strategies and programmes can assist in maximising impact. There is some coordination of research software initiatives with other digital infrastructures at national and regional levels. However, the landscape is complex: a report mapping international digital research infrastructures notes that while governments want to ensure that researchers have the eResearch tools and resources they need, they are unclear on the government's role (as compared to the role of research institutions) in meeting these needs, and how to enable benefits such as first move advantage/strategic opportunities, economies of scale, economies of scope/spillover effects, networking effects/spillover effects, and reputational impacts (Dietrich, 2019).

Some countries are now combining digital research infrastructures into single entities. For example, a new national body, the Digital Research Alliance of Canada, has been formed to integrate activities in advanced research computing, data management, and research software. However, there is still work to be done with regard to research software. The Canadian Research Software Current State Assessment suggests that "[f]ederal and provincial research organisations need to develop a strategy for working more collaboratively with higher education and across governments. [...] Although national coordinated leadership for research software is emerging, efforts to crystallise research software communities have been hampered by a lack of adequate funding and a formal mandate. Without coordination of investment in the context of research software, it is difficult to develop the shared policies, processes, protocols, best practices, and standards that are so essential" (Digital Research Alliance of Canada, 2021).

This centralised coordination is occurring at an even greater scale in some regions, including the pan-European EOSC and the African Open Science Platform, with the latter focused on ensuring that African scientists have access to computational infrastructure. And at the community level, ReSA provides coordination of research funders through the Research Software Funders Forum, as part of its mission to bring research software communities together to collaborate on the advancement of the research software ecosystem. Other community initiatives are seeking to address international challenges collaboratively, such as the FAIR Principles for Research Software (Chue Hong et al., 2022), and the FORCE11 Software Citation Principles (Smith et al., 2016).

In the UK, Large-scale computing: the case for greater UK coordination's recommendations include national coordination: "We recommend establishing a team within Government to provide policy leadership of large-scale computing. This team would be responsible for developing a rolling long-term roadmap for large-scale computing. This roadmap should cover the whole UK computing ecosystem, including software, skills, and user needs. This would help to improve resource sharing between organisations and provide a conduit for industry engagement" (Government Office for Science, 2021).

Some of the UKRI Councils strongly emphasise this need. The UK Data Research Infrastructure Landscape notes that members of the research and innovation community consider data analysis to be a key bottleneck for bioscience in the coming years, in light of the growing volume and diversity of data (DARE UK Consortium, 2021). NERC's Digital Strategy 2021-2030 states that, in order to meet its ambition for the NERC community to have access to sufficient computational resources to secure advances in environmental science, NERC will seek to "build and maintain a forward-looking view of the likely computational capacity needs for environmental science over the next decade, and support the necessary capability development" (NERC, 2021).

Major findings:

- 97% of survey participants see software as important to their own research, with 85% citing it as essential.
- Most people find software through colleagues, and/or from a paper/talk/event.
- Almost two-thirds of respondents share their software and data publicly.
- Junior and Early career researchers are more likely to use less advanced infrastructure, particularly if outside their organisation.
- Researchers funded by BBSRC are more likely to store data in a shared cloud service (62%) and store data in a data repository provided by their organisation.

4.3.2 Computational infrastructure

The project focused on understanding large-scale computing usage and future need, how infrastructure choices are made, and the different infrastructure requirements of researchers across different disciplines and career stages. The importance of combining improvements in computational infrastructure with improvements in personnel was demonstrated by one stakeholder: "Some of our RSE efforts have doubled the speed of our hardware, and improved the optimisation of the code by a factor of ten. These things are making much better use of our resources. It's undeniable that training is key to maximising output from a minimum input. So a minimum research funding input gives us a maximum output if we're able to coordinate these things."

To assist in understanding computational infrastructure needs, survey participants were asked the following: "Which of the following do you currently use or plan to use in the next five years for the computational parts of your research?" Results are shown in Figure 29.

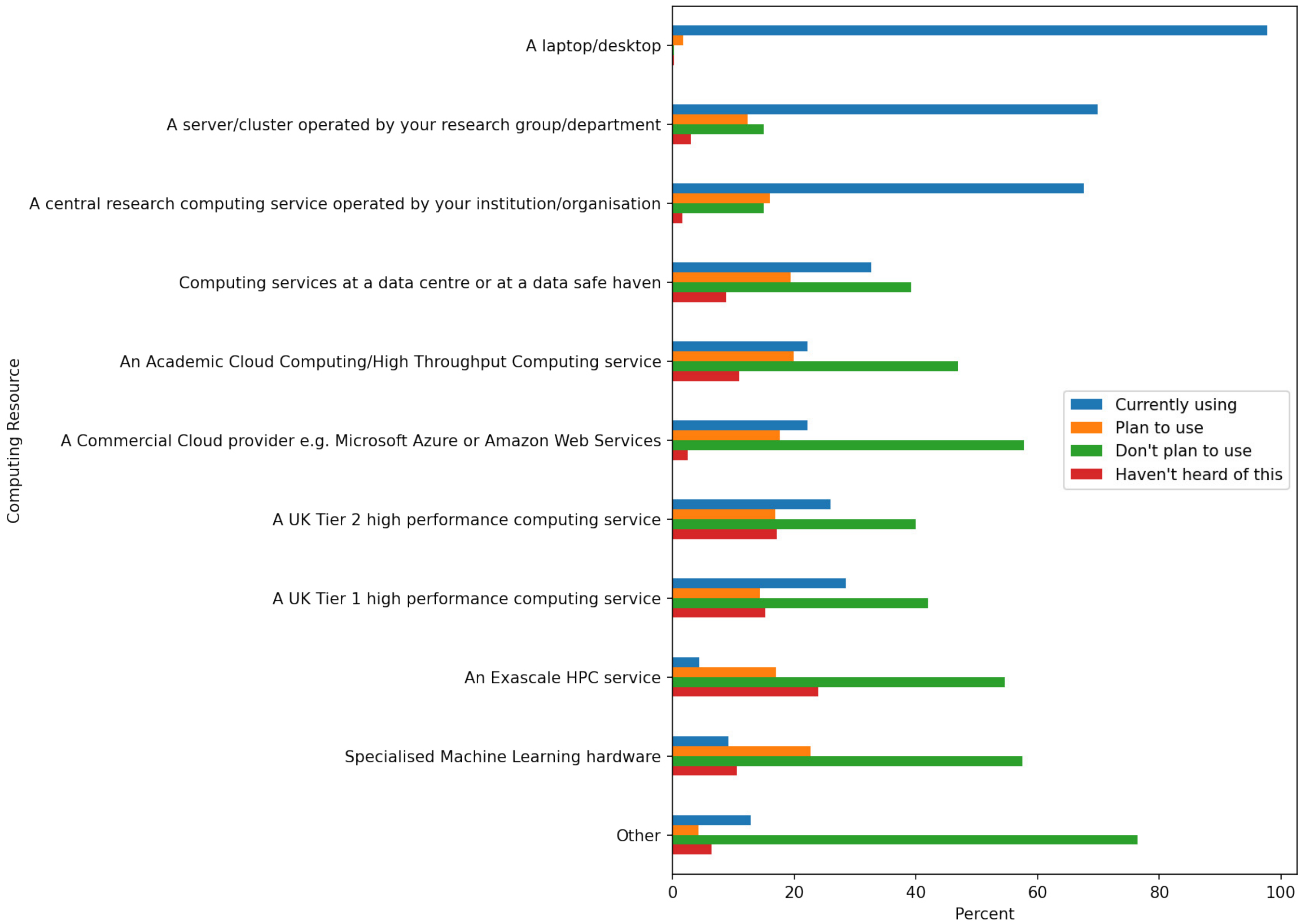


Figure 29: Infrastructure that survey participants currently use or plan to use in the next five years for the computational parts of their research.

Analysis of this data from survey participants reveals trends that include:

- A gradual decline in the type of infrastructure currently being used as the infrastructure becomes more advanced.
- 95% currently use or plan to use infrastructure beyond their own laptop/desktop.
- High levels of awareness of most infrastructure options. The least known are:
    - An exascale HPC service: 24% are not aware of this.
    - A UK Tier 2 HPC service: 17% are not aware, though it is possible that some researchers may not recognise Tier 2 HPC services as such, since many Tier 2 HPC services are hosted by institutions and might therefore be mistaken for central research computing services.
    - A UK Tier 1 HPC service: 15%.
    - An academic cloud computing/HTC service: 11%.

Most researchers appear to focus on local resources and are largely not planning to use larger scale resources. The commonest responses among the 21 that comprised the "Other" category were: specialist computing hardware (embedded controllers, field-programmable gate arrays), which accounted for seven responses, and non-UK HPC, which accounted for five.

A synthesis of group interview participants' perspectives on large scale computing is helpful for understanding these usage patterns. These viewpoints were offered during discussions in relation to the use of digital research infrastructure beyond researchers' own institutions. The quotes below provide insight into what participants found works well in research computing:

ARCHER2 … you can raise requests there and people have been prompt responding and helping me out. No complaints. (Early career stage, Engineering and technology)	*JASMIN … I know a lot of people complain it's too crowded … personally, it seems the best option.* (Early career stage, Physical sciences)
The national resources work well … they're not perfect because there's only enough money to buy certain supercomputers or JASMIN storage. (Mid/Recognised career stage, Multidisciplinary)	*Through collaborations, we've been using the Barcelona Supercomputing Centre tier zero facilities and XSEDE resources in the US.* (Early career stage, Multidisciplinary)
We have access to Supercomputing Wales, a large set of dedicated machines. This has worked well over the last 10 years. (Mid/Recognised career stage, Engineering and technology)	*We're happier with what we do between ARCHER and JASMIN, the Oxford-based NERC facility.* (Mid/Recognised career stage, Physical sciences)

In addition, participants identified a number of challenges for research computing:

A new trend seems to be to use cloud computing … not just a supercomputer … Amazon Web Services has been proposed to replace supercomputer usage - that might be interesting but I don't think there's training on this in academia. (Early career stage, Engineering and technology)	*Every day I see graduate students and postdocs struggling to get research done because they can't get the software to build … or the code they've written needs a week on a supercomputer.* (Mid/Recognised career stage, Multidisciplinary)
Our researchers use the Met Office supercomputer, ARCHER, JADE or BEDE. The problem is getting things in and out. We're talking terabytes of data so it's no mean feat. (Mid/Recognised career stage, Physical sciences)	*Supercomputers are critically important resources but they are badly aligned with what we need.* (Mid/Recognised career stage, Multidisciplinary)
A lot of [EPSRC] calls go out for … RSE positions [to speed up coding in research]… we haven't seen anything go through the STFC route at all and that's disappointing when STFC-funded research needs just as much as EPSRC-funded research. (Mid/Recognised career stage, Biological sciences)	

Finally, group interview participants identified a range of other issues deriving from their experiences of larger scale computational infrastructure and next level computing:

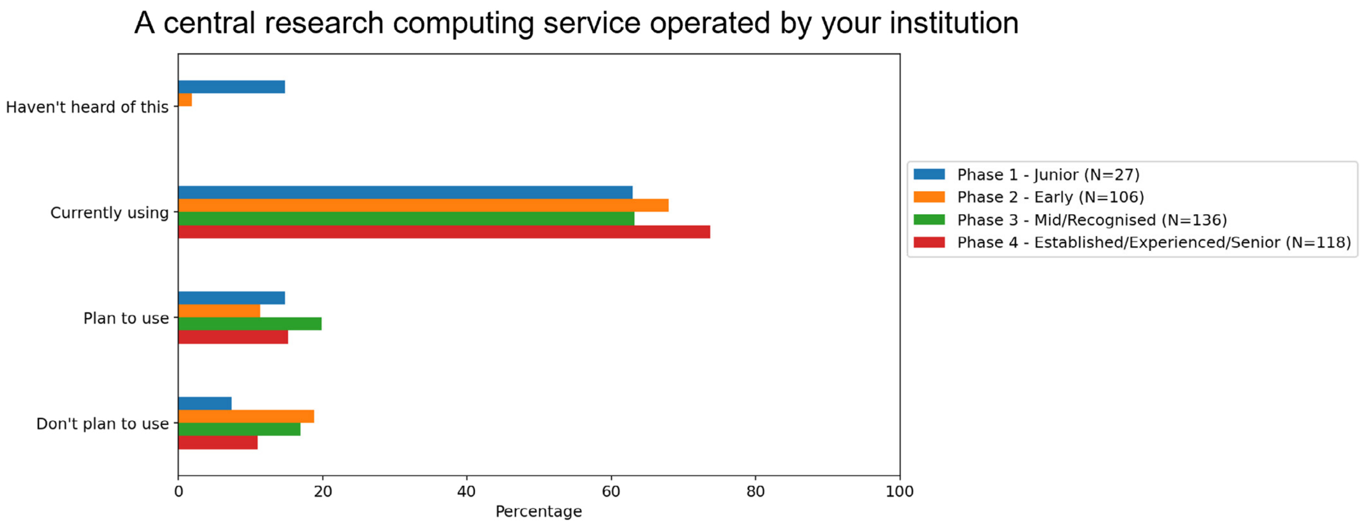
There are two competing things: capability and capacity. E.g., we recently went from ARCHER to ARCHER2 ... while there was a raw increase in capacity if you count it as core hours, on the order of six-fold or something like that. But also it delivers new capability to run coupled, closely parallel calculations, up to a quarter of a million cores which seemingly was impossible on ARCHER which only had 120,000 cores. You could have achieved the capacity of this probably more cheaply by lots of smaller systems but then you wouldn't have had the capability to do the potentially larger scale models.

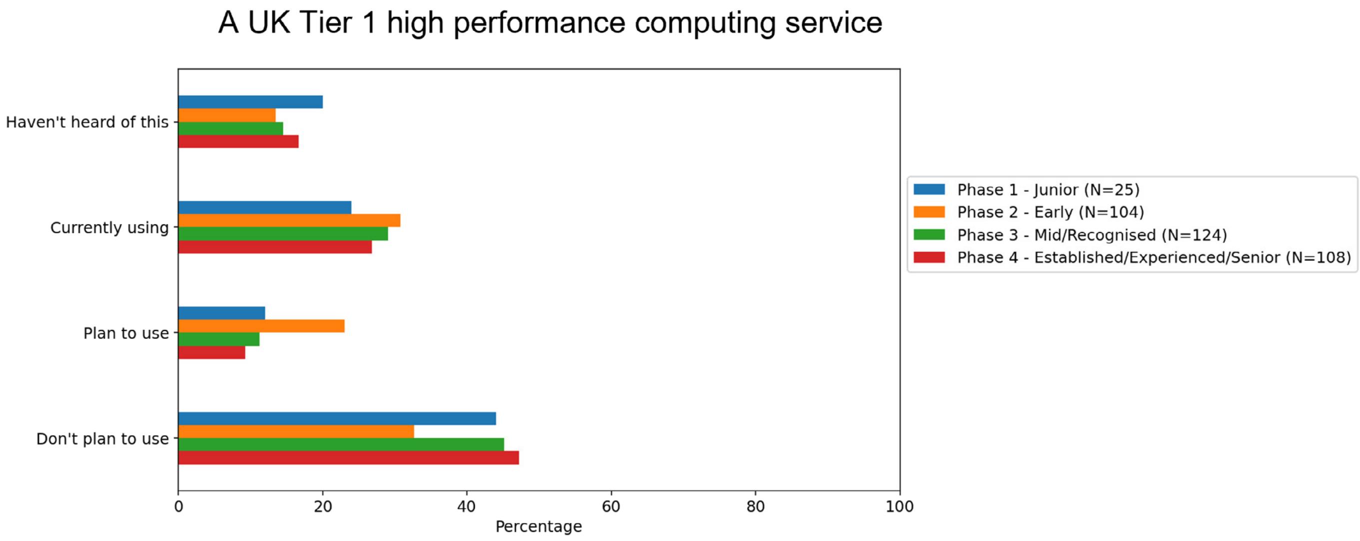
The supercomputing stuff works great when you have a small dataset that is going to solve a lot of big problems; so you have a set of parameters, you have some code, and it runs away and generates some data. And that's fantastic where it's computationally intensive. As soon as it becomes data intensive just the logistical cost of shifting stuff around is "nightmare-ish"; we can't do it. Which is why we have our own dedicated network which we're in the process of upgrading by an order of magnitude.

Something that spurred my interest recently was cloud computing. I think it's an interesting opportunity, not necessarily as a replacement for supercomputing, but just to do something else. Because of the problems I looked at recently I had to use different types of software. Some of them were Windows tools, others were Linux tools. This is a heterogeneity in the software, not just on the hardware. Maybe cloud computing is something that could exploit all the different types of resources optimised for different types of problems to then address something that we couldn't before.

For [our] research we have internationally funded, dedicated computer centres, UK-based at Supercomputing Wales ... [there's] a reasonably large set of dedicated machines there. This has generally worked well over the last 10 years and given us the computing where we needed it but a lot of what we do is HTC rather than large parallel jobs.

Further analysis was undertaken concerning three of the infrastructure types that survey participants currently use or plan to use, by career stage, as shown in Figure 30.





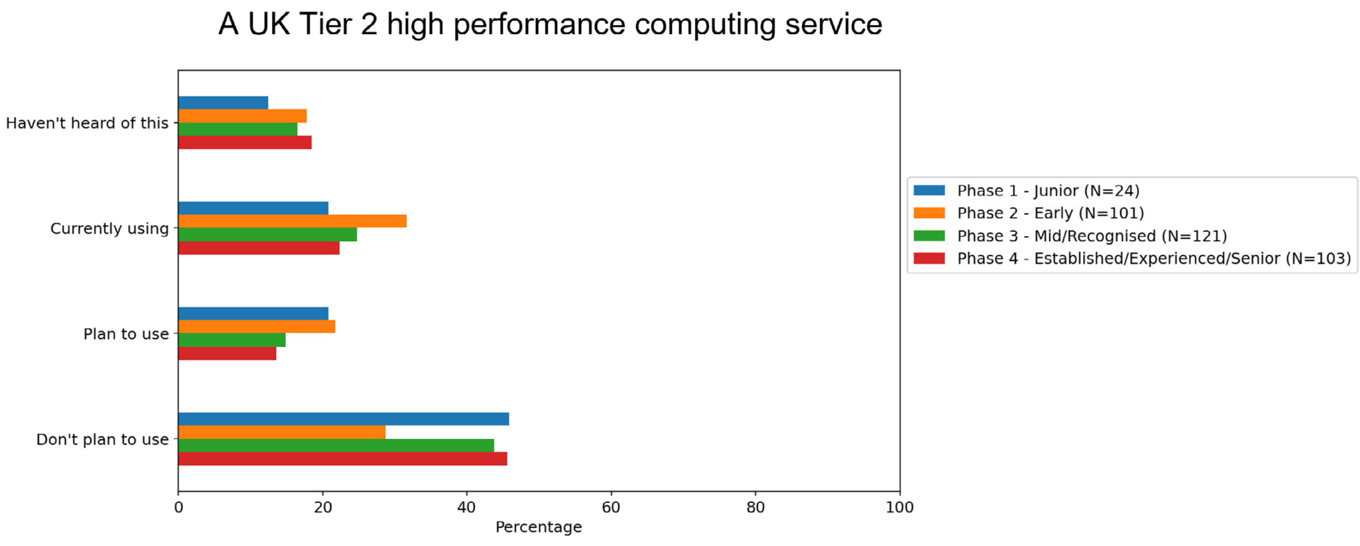


Figure 30: Three types of infrastructure that survey participants currently use or plan to use in the next five years for the computational parts of their research, by career stage.

This provides a range of interesting insights:

- 15% of Junior career researchers are unaware of research computing services operated by their organisations (noting that the number of Junior career stage respondents for the survey was only 7% (27 participants). Almost all other career stages are aware of this facility.
- Across career stages, similar percentages of survey participants report currently using a Tier 1 HPC service. However, a greater percentage of Early career researchers report planning to use this (23%) as compared to Mid/Recognised (11.29%) and Established/Experienced/Senior (9.26%) researchers.
- A greater percentage of Early career researchers report currently using a Tier 2 HPC service (32%) compared to Mid/Recognised (25%) and Established/Experienced/Senior researchers (22%). Early career researchers are also most likely to report that they plan to use a Tier 2 HPC service in future (21.78%), as compared to Mid/Recognised (15%) and Established/Experienced/Senior researchers (14%).

One interviewee called for more outreach to help researchers entering the field to understand the value of using more advanced technology: "We would benefit from more community workshops and conferences where people who are using more advanced simulation techniques in a particular situation can demonstrate that their codes have indeed been ported to an accelerated technology successfully, to demonstrate the success of that sort of software development to other users".

The next survey question asked: "Do you use any non-UKRI funded computational infrastructure in preference to a UKRI one? If so, what is the main reason?"

The most common responses were that resources are provided by collaborators (18%), and that UKRI-provided infrastructure is not sufficient (see Appendix E, Q16.a). The 10% of respondents who chose "Other" gave explanations such as that other infrastructure suited them better; that they were using their own organisation's infrastructure or that of other organisations by stipulation; or that they were using alternatives that provided easier and/or faster access to infrastructure and support mechanisms.

The survey questions also sought to understand future needs, with participants asked whether they felt that their research would be limited in the next year by access to computational and/or data research infrastructure. 45% responded "No", 28% chose "Yes", and 27% said they didn't know (see Appendix E, Q19). This data was then further broken down by career stage, as shown in Figure 31.

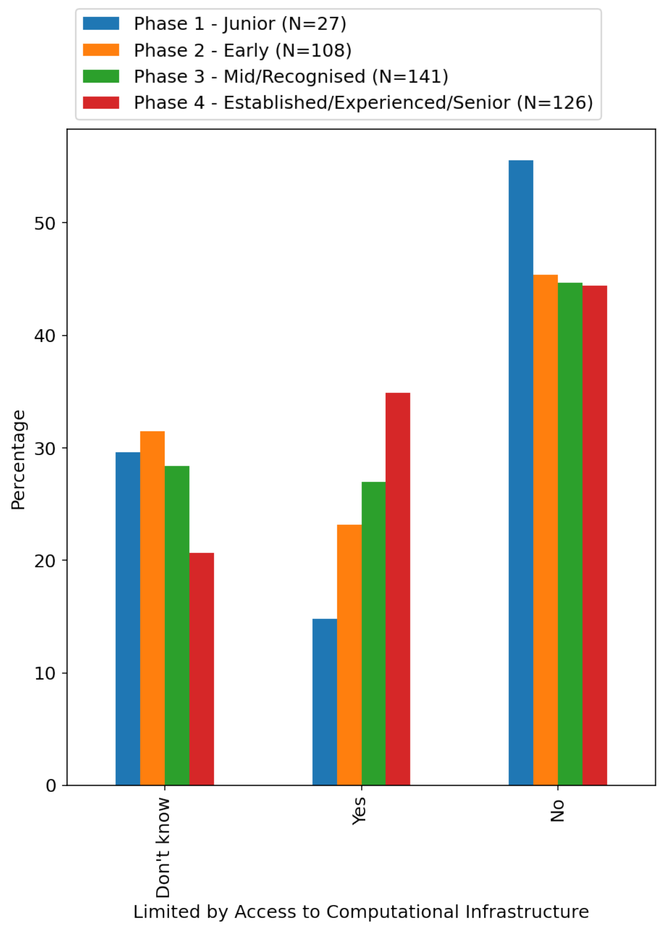


Figure 31: Whether survey respondents feel that in the next year that their research will be limited by access to computational and/or data research infrastructure, by career stage.

This shows that the more advanced the career stage, the greater the perception that research will be limited by access to computational and/or data research infrastructure. 35% of participants in the Established/Experienced/Senior career stage reported that their research in the next year will be limited by access to computational infrastructure, as compared to 27% of Mid/Recognised career stage researchers, 23% of Early career researchers, and 15% of Junior researchers.

The table below contains data from group interviews on access to computational infrastructure. These are grouped by career stage in order to establish whether a relationship exists between seniority and access to computational infrastructure. No clear trend could be discerned, as group interview participants at all levels experienced both challenges and successes with regard to access.

	Issues with access to computational infrastructure	No issues with access to computational infrastructure
Early career stage	*A problem we seem to be constantly running up against is security which is so much tighter than my previous job. We have installation problems all the time on Windows … licensing is [also] a problem because the institute will have a licence tied to one person in the department … it's a weird system to go through them to access that software.*	*I can't say we've ever come across a situation where we couldn't get a piece of software that we needed. In terms of access to specialist support … the university does have an RSE core team, but … although it's technically there, there's far more need for it than there is capacity.*
	At [an] early career stage it's very difficult to even get access sometimes to these computers unless you are in the lab of a principal investigator who is interested in that. If you are a PhD student and you feel that you need a supercomputer but you're not in a lab that has experience handling this kind of access to HPC applications … you will be in trouble.	
Mid/ Recognised career stage	*I have used outside supercomputing facilities, like Cirrus … but right now I don't have access to those. I think you have to cost for them on a grant proposal. We have looked at Microsoft Azure as well … although that hasn't gone further. Those are the resources I think we are aware of and if there is funding, we would like to access them.*	*If you can justify why you need [software] then our IT team will give you admin rights. I have a Windows computer with admin rights. I have no issues downloading what I need to download in order to do my job. There's a university software library, which is readily accessible … it's relatively straightforward to get licences and get what you need.*
	Recently our local HPC people, rather than just having the 6,000 core HPC system, have talked about using Azure. Some of the researchers that they have put onto Azure have discovered it's quite expensive. Once your funding stops you don't get access to it which is frustrating for them.	*I don't think I have any particular issues getting access to software. We all have admin access to laptops so we can install whatever we want. Plus with the advent of things like Docker, actually getting access to software is a lot easier through that route for various things that are a pain in the neck to install; you can just get the container and off you go.*
Established/ Experienced/ Senior career stage	*We're at a limit of how much spinning disk we can have fast access to already. How will we work in the future when we're storing things differently? [What are] the science implications of different storage solutions but also the software ones? How are we going to access it in a way that is quick and easy?*	*I can go to the research computing facility … I could see if [the software is] there already, and if not, we've got a guy … his job is to make sure that the software stack is up to date and respond to software requests.*
		The department is constantly developing that [HPC] … from some local HPC and then to clusters, etc. Our students and all staff can access a European HPC centre. If you have a problem you can just submit a proposal to use some of those.

This data was also analysed in relation to the funders that survey participants said they had applied to and/or received funding from. This showed fairly similar breakdowns across EPSRC, BBSRC, MRC, NERC, and STFC, which were the five funders considered (see Appendix G, Q19 x Q7).

Stakeholders revealed in interviews how the range of computational infrastructure being used is expanding: "There is a need to work with researchers at the high end of infrastructure usage, but also at the lower end too—and the gap is getting bigger. It is starting to be more obvious that the people at one end of the scale, who are pushing the absolute limits of what hardware can do, and into exascale, they need incredible amounts of knowledge, experience, and skill to eke out the best performance from those sorts of machines. And at the other end, we're still seeing people stepping up from doing no digital research to doing some digital research. And the scale is getting longer between the people pushing the limits and the scale of people just starting. So we've got more people to help with the exascale stuff and help upskill people who need to use it. But I think we still need to keep our feet firmly on the ground and also help other people onto the ladder."

Further survey questions sought to understand how decisions on infrastructure usage are made. Figure 32 shows responses to the following question: "Who do you ask for advice on choosing the computational and/or data research infrastructure (computing, storage and network) you use? Select all that apply."

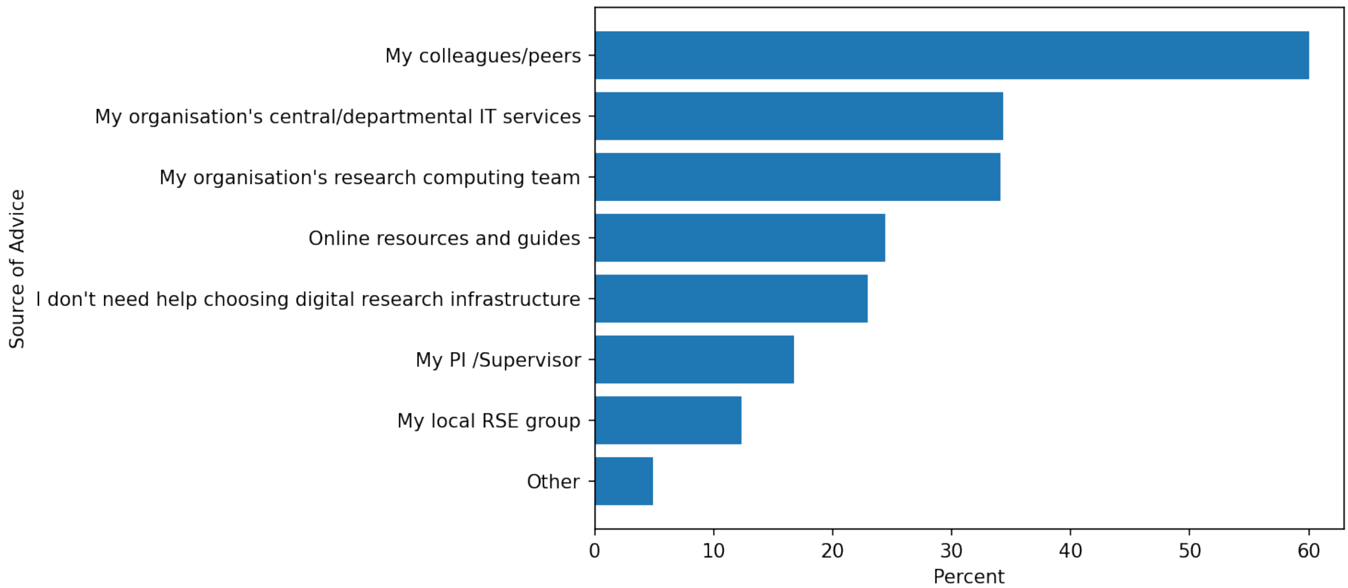


Figure 32: Who survey participants ask for advice on choosing computational and/or data research infrastructure.

This shows that colleagues and/or peers are consulted by 60% of survey participants, with 34% seeking advice from within their organisation. Further analysis of this data by career stage revealed that people in the earlier career stages more often use their principal investigators and/or supervisors. This tendency naturally diminishes as personnel become more senior (see Appendix G, Q20 x Q6).

Major findings:

- 15-20% plan to use infrastructure beyond their own laptop/desktop.
- There are high levels of awareness of most infrastructure options.
- Early career researchers are more likely to be using and/or planning to use Tier 1 and Tier 2 HPC than their mid-career and senior counterparts.
- 27% of participants believe their research will be limited by access to computational/data infrastructure in the next year, rising to 35% of participants at the senior career stages.
- 60% ask colleagues and/or peers for advice on choosing computational and/or data infrastructure.

4.4 POLICY

This section explores how policy and other enablers can assist in providing an environment that both directly supports, and is conducive to supporting, research software and its personnel, across two subsections:

> Enabling environments: analysis of the policies required to future-proof support and enable sustainability.
> Funding: what funding policies are required to enhance and sustain provided support?

Key findings:

- Organisational processes that support RSEs and research software initiatives, including funding, are beneficial at institutional levels, and would also be valuable at national level.
- Funding for research software and its personnel is limited, with the most common research software funding mechanism being standard research grants. There is some perception that inclusion of research software development costs in applications is detrimental.

4.4.1 Enabling environments

Enabling environments are needed to ensure that policies and processes support change. This attempt to formulate a conception of the enabling environments required to future-proof support and enable sustainability began with the subject of software choice. In response to a question on whether survey participants are free to use the software they want to use, 89% replied "Yes". This is an improvement on 10–20 years ago, when more barriers existed due to factors related to commercial software licensing and organisational security requirements for managed desktops.

The 45 survey participants who answered "No" to the question on freedom of choice in software usage were asked to choose from a list of factors preventing them from using the software in question. Multiple options could be selected. The responses are shown in Figure 33.

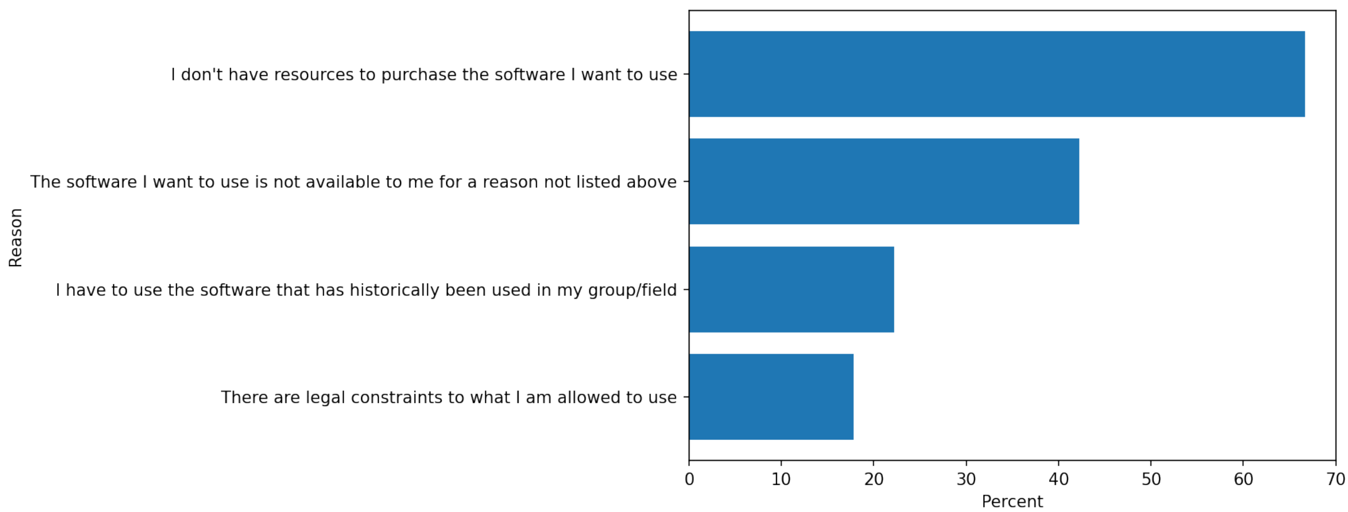


Figure 33: Barriers faced by survey participants who were not free to use the software they want.

To help create a more nuanced picture, participants contributed discussion points within the group interviews. Among the key talking points were "paywalls", leading to issues of reproducibility; "time", because there can be insufficient time to develop software beyond the conclusion of a funded project; and the "nature of funding", which relates to difficulties respondents encountered when asking MRC for project funds to meet costs related to software and computing. A review of funders that each cohort of respondents had applied for and/or received funding from found only small differences (see Appendix G, Q10.a x Q7). The reasons for this are not clear, and the subject merits further investigation.

Software usage can also be constrained by licensing issues. A report on challenges for research software sustainability in Germany and beyond includes analysis of legal aspects: "Many obstacles for research software pertain to legal issues, such as applicable licensing and compatibility of licenses, and decisions about license types" (Anzt et al., 2021). The report recommends the development of a national organisation akin to the SSI, and/or research software task forces within all German organisations that perform research, to assist research teams in the licensing of research software and related legal issues. This would facilitate implementation of the DFG Code of Conduct, which includes guidance on the provision of public access to research results, stipulating that "if self-developed research software is to be made available to third parties, an appropriate licence is provided" (DFG, 2022).

Examples of enabling environments also include organisational processes. For example, a Canadian report notes that: "Making real headway on EDI [for the research software community] will involve challenging entrenched aspects of research culture as well as risk to individuals, so it will be crucial to create policies, systems, and workflows in research communities, teams, and within university human resources (HR) departments that protect vulnerable populations and address systemic patriarchy, colonialism, and racism" (Digital Research Alliance of Canada, 2021).

For example, two of the four stakeholders interviewed for this research emphasised the need for both centralised teams of RSEs and RSEs who are embedded in research teams. Another interviewee cited promoting the awareness of RSE roles across the university community as a useful practice: "Targeting the research support offices has assisted greatly in improving RSE numbers as the research support officers can assist academics to build this into their grant. This is a very good way to scale awareness without having to raise awareness amongst all researchers." The international RSE surveys also provide substantial information on the employment conditions of this cohort in the UK, and globally for comparison.

Coordination levels can also be a major factor in creating enabling environments. An analysis of the US landscape concludes that it "is falling behind in the accessibility and connectedness of its research computing and data infrastructure, compromising competitiveness and leadership and limiting global science that could benefit from US contributions. The challenge is more cultural and institutional than technical and demands immediate and sustained leadership and support, starting with policy-makers and research funders" (Bourne et al., 2022). One interviewee for this study suggested that a multi-level research software strategy is needed. This could include some coordination at the UKRI level, with possible examples including:

- An RSE pool/college to avoid duplication of effort. But there will still be a need for some domain specific requirements as well.
- Mandating of policies such as ensuring that researchers are delivering open source code that is appropriately licensed.
- Use of accreditation measures such as Athena SWAN, to ensure that universities provide sensible career paths for RSEs so that all types of staff in an organisation are treated equally.

Two of the four interviewees highlighted the national advantage that coordination could provide: "If there was a national strategy around digital infrastructure, which really centres [on] people—whatever nation gets this done first would have a jump on others." It was also noted that one of the benefits of a national strategy is that it helps all universities to reach the same level. International collaboration was also cited as a way to accelerate the preparation of a wide range of codes for exascale computing, which would also then enable the UK to have its own complementary focus on those most relevant to the UK.

Large-scale computing: the case for greater UK coordination already makes recommendations related to this: "LSC forms part of the UK's national infrastructure and many of the key issues span multiple sectors. The Government therefore has a significant role to play in nurturing and supporting the UK ecosystem both as a consumer and a funder. Therefore, our primary recommendation is to establish a team within Government to take policy responsibility for large-scale computing and address the challenges that are identified in the report" (Government Office for Science, 2021). UKRI is also developing a Digital Research Infrastructure strategy that states: "we are currently entering a period of rapid change, with increasingly heterogeneous system designs, the emergence of novel architectures, and the blurring of the traditional distinction between central processing unit (CPU) and graphics processing unit (GPU)-based systems, enabling convergence of currently disparate workflows. Our overarching priority is to provide appropriate and ambitious compute capabilities reaching out to exascale for UKRI's diverse research and innovation communities. UKRI continues to pursue promising opportunities, both individually and in partnership with others both nationally and internationally."

Major findings:

- Organisational processes that support RSEs and research software initiatives are beneficial at institutional levels, and would also be valuable at the national level.
- Choice of software usage is constrained mostly by funding rather than licensing or organisational policies.

4.4.2 Funding

A number of survey questions investigated issues related to funding, beginning with the following: "Have you ever applied for funding that explicitly includes costs for software development?" In response, 45% of survey participants replied "Yes" (see Appendix F). This was further broken down by career stage, as shown in Figure 34.

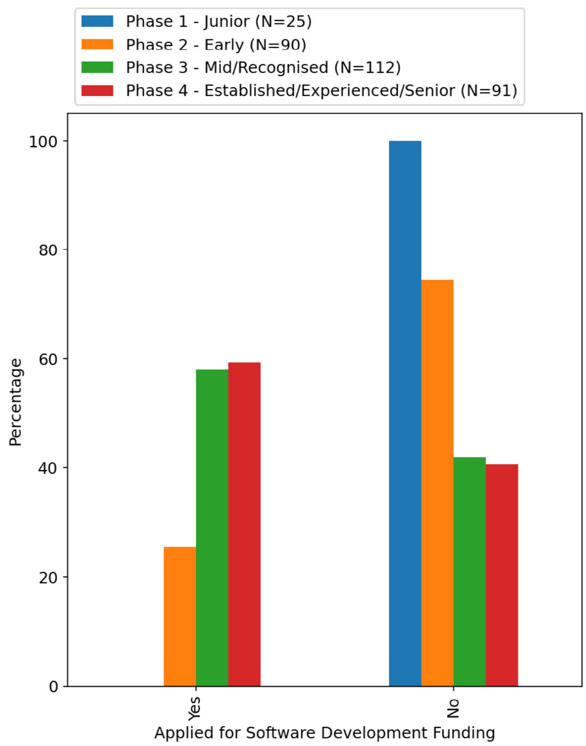


Figure 34: Survey participants who had applied for funding that explicitly includes costs for software development, by career stage.

The proportion of survey applicants who had applied for this type of funding increased with career stage (Early career stage: 26%, Mid/Recognised career stage: 58%, Established/Experienced/Senior career stage: 59%). This increase in applications is to be expected, since more senior staff often take more responsibility for obtaining resources. However, it may be beneficial to create an environment where personnel at earlier stages of their careers more actively seek this type of funding.

This data was compared with responses from another question, which asked applicants who had applied for funding that explicitly included costs for software development to identify the funder, as shown in Figure 35.

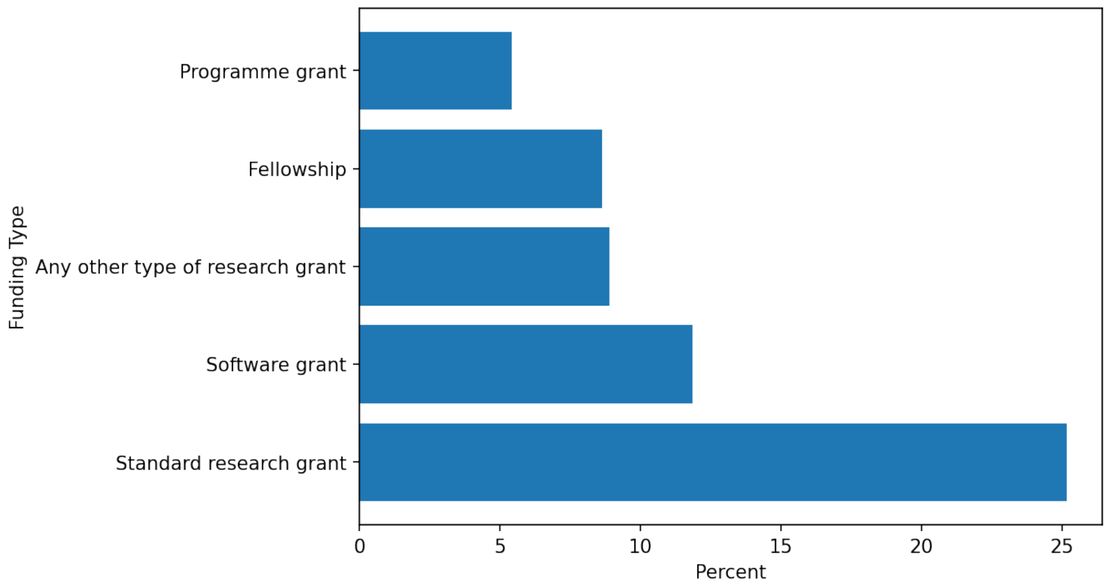


Figure 35: Funders that survey participants had applied to for funding that explicitly includes costs for software development.

These findings are consistent with other results from this study—i.e., that EPSRC, NERC, and STFC have somewhat higher rates than BBSRC and MRC. They also demonstrate that there are other major funders in this space, such as the European Commission/European Research Council. These responses were further compared with data on which bodies respondents were funded by generally. This analysis showed that survey participants were most likely to apply for software development funding from their overall funding body (see Appendix G, Q12.b x Q7). One exception is that participants funded by NERC were proportionally more likely to apply for software funding from the EC/ERC than participants funded by other research organisations. 32 respondents to this question selected "Other" and provided more details.

The most commonly cited bodies were Wellcome Trust (six responses), British Heart Foundation, NIHR, and US NSF (three responses each).

Figure 36 shows responses to the next survey question: "What type of funding did you apply for?" More than one option could be chosen, and this figure shows only responses from survey participants who had applied for funding explicitly including costs for software development.

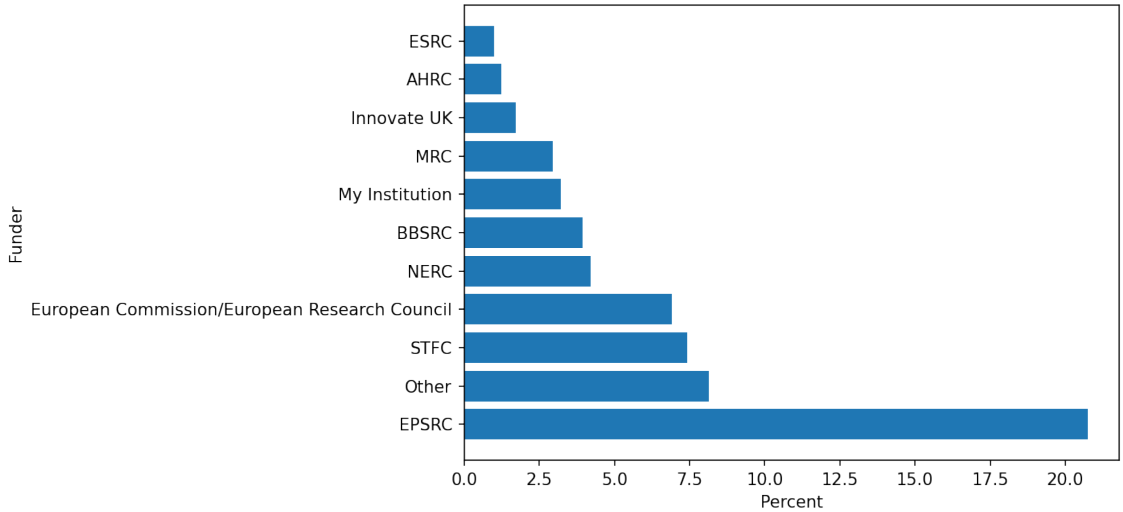


Figure 36: What type of funding survey participants applied for.

This shows that research grants are still the most common form of research software support. Data on software grants within EPSRC shows that there is only a 10% success rate—11 (EPSRC, 2022c) of 111 proposals (EPSRC, 2021a, 2021b, 2021c) were funded from the 2021 open call, whereas in 2021 the overall average for all EPSRC grants was 26% (EPSRC, 2022b). In comparison, the overall funding rate for EPSRC in 2020-21 was about 36%, and filtering by the award category of "fellowship" shows 27% of awards by number, or 33% awards by value were funded (EPSRC, 2022b). This illustrates that the success rate for software grants is much lower than for other types of funding. This could cause wider issues such as lack of maintenance funding for research software, "hiding" work in other types of grants, poorer recognition for software, and loss of personnel.

Another survey question solicited views on including software development costs in funding proposals, as shown in Figure 37.

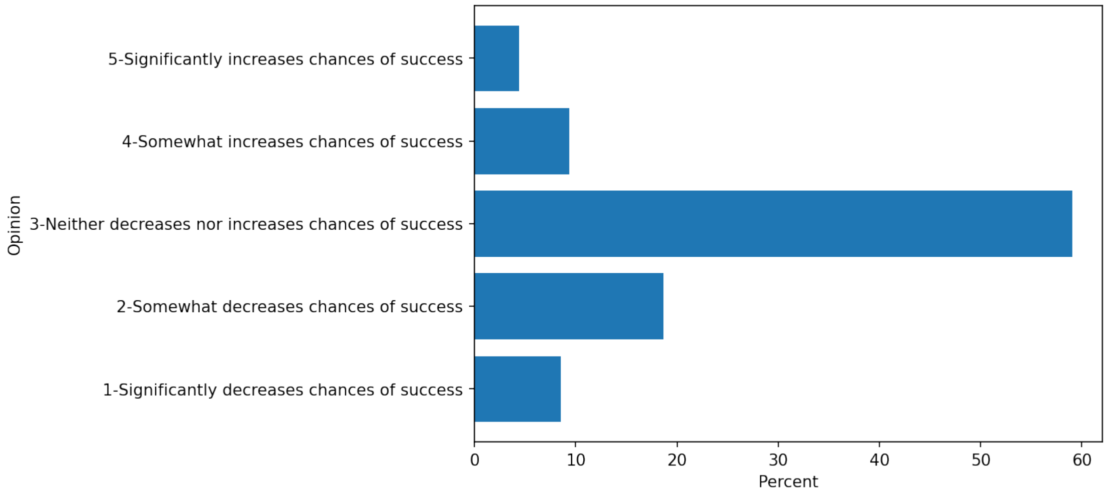


Figure 37: Opinions of survey participants on including software development costs in funding proposals.

This shows that while 60% of the survey respondents felt that including software development costs on a grant was neither detrimental or helpful, 27% felt it could harm their applications. Only 14% felt that it would increase their likelihood of receiving funding. Further analysis of this data examining only the opinions of survey participants who had applied for funding explicitly including costs for software development showed roughly the same percentages.

Survey participants were also invited to provide details on why they held their views on including software development costs in funding proposals. Among the respondents who chose "Neither decreases nor increases chances of success" the main themes were: "Depends on the grant/proposal/call/project/reviewers", "software development costs are hidden as postdocs/PhDs/research staff time", and "as long as it is properly justified". For the respondents who chose "Significantly decreases the chances of success" the most common themes were: "Reviewers/funders don't appreciate/value research software development as scientific research" and "importance is placed on the science and not development of research software". Several responses also cited certain funders as having no history of funding research software development and/or maintenance. For the respondents who chose "Significantly increases the chances of success", most argued that the inclusion of software development costs should significantly increase the chances of success, rather than basing their answers on experience.

Large-scale computing: the case for greater UK coordination provides explicit recommendations on how job security, salary structures, and progression opportunities should be improved to retain talent and respond to career path and progression challenges. This report concludes with steps that should be taken across academia and the public sector with relation to funding to retain large-scale computing professionals. These steps include the following: "Staffing costs should be fully accounted for during procurement to ensure facilities have adequate staffing for the life of a computing system" (Government Office for Science, 2021).

Similarly, the ExCALIBUR Research Software Engineer Knowledge Integration Landscape Review includes in its recommendations that "UKRI should ensure that it supports the message that RSEs are a highly valued resource at Universities, National Laboratories and other research organisations by providing clear guidance for inclusion of RSEs on grants" (Parsons et al., 2021).

A final question on funding asked: "Do you know who funds the computational infrastructure you use for your research?" 73% answered "Yes" (see Appendix E, Q16), with the percentage giving that answer increasing with seniority (see Appendix G, Q16 x Q6). This is to be expected as senior staff are more likely to apply for funding.

Major findings:

- Slightly more survey participants had applied for funding that explicitly includes costs for software development from EPSRC, NERC, and STFC than from BBSRC and MRC.
- The most common research software funding mechanism is standard research grants.
- Slightly more respondents felt that it was detrimental to include software development costs on a grant than those who thought it was helpful—but most were neutral on this topic.

5. KEY FINDINGS AND RECOMMENDATIONS

There are a number of recommendations arising from the analysis contained in section 4 that may assist the UK in maintaining its world-leading position with regards to research software.

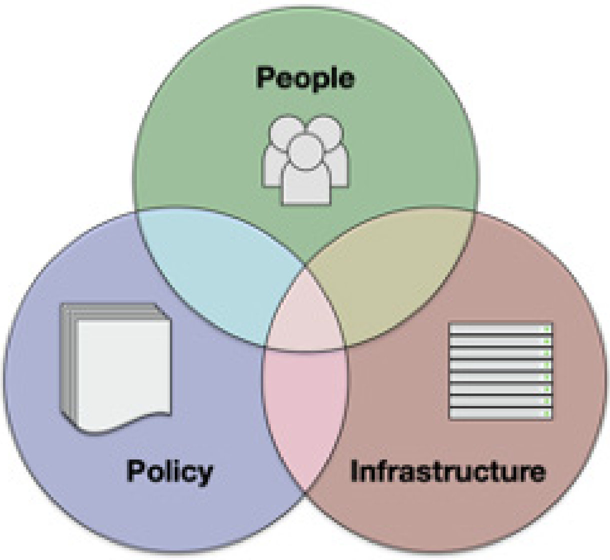
The ARDC's framework for ensuring that research software is recognised as a first-class output of research identifies three levels at which coordinated action is needed: to see software (increase visibility), to shape software (increase fitness for purpose), and to sustain software (increase sustainability) (ARDC, 2022). For the UK, many of the recommendations contained in this report focus on shaping and sustaining, with the latter arguably presenting the greatest challenge since there are few international exemplars to learn from. The UK has both the privilege and the challenge of being able to break new ground. However, some of this report's recommendations relate to the seeing and shaping of software, and for these there are many examples of best practice, both locally and internationally, which are not being implemented in a coordinated way in the UK.

The following recommendations are designed to aid in the creation of a cohesive national strategy on research software to address software, skills, and personnel gaps across the research landscape.

People, infrastructure, and policy recommendations:

People: All parts of the research community must understand the requirement for a wide variety of roles that support the research computing ecosystem. Better recognition and career pathways are vital to ensuring that there are enough of these people in the UK to support the infrastructure.

Provide unified funding requirements across UKRI Research Councils that align with international standards, including relevant UNESCO and OECD recommendations and international vision.



Develop a national roadmap for coordinated access to, and training in, all levels of UK research computing infrastructure. This should include a focus on enabling personnel to transition between levels, both in their use of the research computation and in their confidence/competency with the software.

Figure 38: Key recommendations.

People recommendations:

4. All parts of the research community must understand that a wide variety of roles support the research ecosystem. Better recognition and appropriate career pathways are vital to ensuring that there are enough skilled people in the UK to support research. This is the joint responsibility of a number of key stakeholders, including Principal Investigators, universities, funders, industry, and government.

   1.7. Enable detailed analysis of how to professionalise RSE roles, building on existing initiatives such as the international RSE survey. This analysis should profile RSE as a career through case studies, careers fairs, etc., and it should occur in tandem with legitimisation of other roles such as Data Stewards, etc.

   1.8. Facilitate collaboration between government, funders, and employers (particularly universities and Principal Investigators, and potentially industry) to create national policies aimed at improving standards of employment for RSEs in academia around length of contract, pay standards, mobility between academia and industry, and professional development, by means of standard assessment criteria. Investigate whether contract terms such as a minimum of two years (or over) for RSEs improve job satisfaction and retention.

   1.9. Support further research by UKRI and relevant professional bodies (RLUK, CILIP, etc) to identify why some research computing-adjacent roles, such as Research Librarian and Data Steward, are not as visible or utilised by researchers engaged in scientific computing. Consider whether this suggests changes are required to make these roles more relevant to researchers whose work is computationally intensive.

   1.10. Facilitate collaboration between government and universities to create a training strategy that empowers all research roles and enables them to take the time to learn the skills needed for modern research. It must build a pathway through data and software training that starts at the undergraduate level, builds through Masters and PhD study, and continues during employment. The strategy should:

   - Incentivise professional development opportunities for early career researchers.
   - Identify and address any gender disparities among the people who receive training.
   - Identify and address disciplinary differences, such as the lower software development undertaken by MRC and BBSRC researchers.

   Cultivate the most in-demand skills through teaching and training, such as honours/Masters courses.

   1.11. Enable regular information collection exercises to be undertaken to allow UKRI to track current training provision and highlight skills that will be vital in the future.

   1.12. Conduct a large-scale study of gender inequalities in scientific computing alongside an analysis of EDI initiatives, with the goal of producing a set of recommendations for funders on reducing inequalities. This could assist in addressing challenges and frame recommendations on how to use existing best practice within the UK and internationally to improve EDI outcomes. Leaders in this field, such as Advance HE and the Royal Society, should be brought in as partners in this study.

Infrastructure recommendations:

5. Develop a national roadmap for a coordinated access to, and training in, all levels of UK research computing infrastructure, involving universities, funders, industry, and government. This should include a focus on enabling personnel to transition between levels, both in their use of research computing and in their confidence/competency with the software.

   5.1. Provide unified guidance across UKRI Research Councils that support recognition of software as critical digital research infrastructure, and implementation of international standards on software citation and FAIR Principles for Research Software.

   5.2. Undertake further analysis to understand what the barriers are to researchers running their experiments on Tier 2 and larger research computing infrastructure. This should include consideration for increasing availability of RSEs to help researchers in bridging this gap, to ensure that a wide range of researchers benefit from investments in exascale, HPC, and cloud computing.

   5.3. Recognise the need to support users of both less advanced and more advanced infrastructure through access to RSEs and training, and encourage international cooperation in order to improve access to larger scale resources and different architectures and technologies.

Policy recommendations:

6. Provide unified funding requirements across UKRI Research Councils that align with international standards, including relevant UNESCO and OECD recommendations, and lay out the following international vision:

   > Research software must be recognised as a key element of research.

   > The development and maintenance of research software must be supported.

   > Research software must be as open and/or as FAIR as other components, so that the research it enables can be trusted and replicated.

   6.1 Incorporate into these unified funding requirements a framework in line with the 10 Simple Rules for Funding Scientific Open Source Software, and include:

   > Specific programmes for maintenance.

   > Encouragement of reuse and/or contribution to existing platforms.

   > A variety of sizes of funding.

   6.2. Continue to lead and/or contribute to international efforts to develop standards and practices that solve challenges faced by the UK research community.

These recommendations identify where further research is necessary to help understand systemic imbalances and leverage points in the community, and to identify how to integrate more detailed analysis with initiatives (such as those designed to improve talent pipelines).

REFERENCES

Ainsworth, R. (2021), *Highlights from Collaborations Workshop 2021,* Software Sustainability Institute, https://software.ac.uk/blog/2021-05-24-highlights-collaborations-workshop-2021

Akhmerov, A., Cruz, M., Drost, N., Hof, C., Knapen, T., Kuzak, M., Martinez-Ortiz, C., Turkyilmaz-van der Velden, Y., & van Werkhoven, B. (2019), *Raising the Profile of Research Software,* Zenodo, https://doi.org/10.5281/ZENODO.3378572

Almarzouq, B. (2022), Reflection on Open Science practices and Research Software in the Kingdom in alignment with Saudi Arabia's Vision 2030, *Asia Pacific Advanced Network (APAN) 53,* Bangladesh, https://docs.google.com/presentation/d/1Ow6aIFxhOWRpB6AxFbm5uYpEAiSQyo7uXdwrZpMiQas/edit#slide=id.p

Anzt, H., Bach, F., Druskat, S., Löffler, F., Loewe, A., Renard, B. Y., Seemann, G., Struck, A., Achhammer, E., Aggarwal, P., Appel, F., Bader, M., Brusch, L., Busse, C., Chourdakis, G., Dabrowski, P. W., Ebert, P., Flemisch, B., Friedl, S., ... Weeber, R. (2021), An environment for sustainable research software in Germany and beyond: Current state, open challenges, and call for action, *F1000Research,* 9, 295, https://doi.org/10.12688/f1000research.23224.2

ARDC (2022), *A National Agenda for Research Software,* Zenodo, https://doi.org/10.5281/zenodo.6378082

Asmi, A., Ryan, L., Salmon, E., Kubiak, C., Battaglia, S., Förster, M., Dupré, J., Walter, H., Stirling, W. G., Clausen, K., Knudsen, M., Ferreira-Aulu, M., Shabanova-Danielyan, E., Wang, W., Luukkanen, J., Kaivo-oja, J., Baldovin, C., Timmer, M., van der Meer, R., ... Sundet, J. (2019), *International Research Infrastructure Landscape 2019 (Version 1),* Zenodo, https://doi.org/10.5281/ZENODO.3539254

Barker, M. (2021), *ReSA People Roadmap Report,* Zenodo, https://doi.org/10.5281/ZENODO.5633318

Barker, M., & Buchhorn, M. (2022), *Research Software Capability in Australia,* Zenodo, https://doi.org/10.5281/ZENODO.6335998

Barker, M., Chue Hong, N. P., Katz, D. S., Leggott, M., Treloar, A., van Eijnatten, J., & Aragon, S. (2021), *Research software is essential for research data, so how should governments respond?,* Zenodo, https://doi.org/10.5281/zenodo.5762703

Barker, M., & Katz, D. S. (2022), *Overview of research software funding landscape,* Zenodo, https://doi.org/10.5281/ZENODO.6102487

BBSRC (2020), *Review of Data-Intensive Bioscience,* UKRI, https://www.ukri.org/wp-content/uploads/2020/11/BBSRC-201120-ReviewOfDataIntensiveBioscience.pdf

BCS (2022), *BCS landscape review: Computing qualifications in the UK,* https://www.bcs.org/policy-and-influence/education/bcs-landscape-review-computing-qualifications-in-the-uk/

Benjamin, R. (2019), *Race After Technology: Abolitionist Tools for the New Jim Code,* John Wiley & Sons

Bourne, P. E., Bonazzi, V., Brand, A., Carroll, B., Foster, I., Guha, R. V., Hanisch, R., Keller, S. A., Kennedy, M. L., Kirkpatrick, C., Mons, B., Nusser, S. M., Stebbins, M., Strawn, G., & Szalay, A. (2022), Playing catch-up in building an open research commons, *Science,* 377(6603), 256–258, https://doi.org/10.1w126/science.abo5947

Brett, A., Croucher, M., Haines, R., Hettrick, S., Hetherington, J., Stillwell, M., & Wyatt, C. (2017), *Research Software Engineers: State Of The Nation Report 2017,* Zenodo, https://doi.org/10.5281/ZENODO.495360

Campbell, L. G., Mehtani, S., Dozier, M. E., & Rinehart, J. (2013), Gender-Heterogeneous Working Groups Produce Higher Quality Science, *PLoS ONE,* 8(10), e79147, https://doi.org/10.1371/journal.pone.0079147

Carver, J. C., Weber, N., Ram, K., Gesing, S., & Katz, D. S. (2022), A survey of the state of the practice for research software in the United States, *PeerJ Computer Science,* 8, e963, https://doi.org/10.7717/peerj-cs.963

ChEESE (2022), *Exascale-compatible geohazard mitigation services developed to help authorities in decision-making,* https://cheese-coe.eu/media/news/press-release-exascale-compatible-geohazard-mitigation-services-developed-help

CHORUS (2022), *Software Citation Policies Index,* https://www.chorusaccess.org/resources/software-citation-policies-index/

Choudhury, S. (2021), *University Open Source Program Offices,* https://ospoplusplus.com/blog/university-open-source-program-offices/

Chue Hong, N. P., Cohen, J., & Jay, C. (2021), Understanding Equity, Diversity and Inclusion Challenges Within the Research Software Community, *Computational Science – ICCS 2021 (Vol. 12747, pp. 390–403),* https://doi.org/10.1007/978-3-030-77980-1_30

Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., & Honeyman, T. (2022), *FAIR Principles for Research Software (FAIR4RS Principles),* Zenodo, https://doi.org/10.15497/RDA00068

Cleaver, J., Derrick, G., & Hettrick, S. (2022), *What does Open Research mean for the future of the REF? The next REF could better recognise the diversity of roles contributing to research,* HEPI, https://www.hepi.ac.uk/2022/05/30/what-does-open-research-mean-for-the-future-of-the-ref-the-next-ref-could-better-recognize-the-diversity-of-roles-contributing-to-research/

Cohen, J., & Woodbridge, M. (2020), *RSEs in Research? RSEs in IT?: Finding a suitable home for RSEs,* arxiv.org, https://doi.org/10.48550/ARXIV.2010.10477

Culina, A., van den Berg, I., Evans, S., & Sánchez-Tójar, A. (2020), Low availability of code in ecology: A call for urgent action, *PLOS Biology,* 18(7), e3000763, https://doi.org/10.1371/journal.pbio.3000763

DARE UK Consortium (2021), *UK Data Research Infrastructure Landscape,* Zenodo, https://doi.org/10.5281/ZENODO.5584696

Department for Business, Energy & Industrial Strategy (2021), *R&D People and Culture Strategy: People at the heart of R&D,* GOV.UK, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1004685/r_d-people-culture-strategy.pdf

Department of Education, Skills and Employment (2022), *2021 National Research Infrastructure Roadmap,* GOV.UK, https://www.dese.gov.au/national-research-infrastructure/resources/2021-national-research-infrastructure-roadmap

DFG (2022), *Guidelines for Safeguarding Good Research Practice. Code of Conduct,* Zenodo, https://doi.org/10.5281/ZENODO.6472827

Dietrich, M. (2022), *Digital Infrastructure Makes a Big Impact on Research Excellence,* Bloodstone Solutions, https://www.bloodstonesolutions.org/Big-impact

Dietrich, M. (2019), *Mapping Effective eResearch Services: The International Situation is Intensifying,* eResearch NZ, https://eresearchnz.figshare.com/articles/presentation/Mapping_eResearch_Ecosystems_the_international_situation_is_intensifying_/8066942/1

Digital Research Alliance of Canada (2021), *Research Software Current State Assessment,* https://alliancecan.ca/sites/default/files/2022-03/rs_current_state_report_1.pdf

Dunbar-Hester, C. (2020), *Hacking Diversity: The Politics of Inclusion in Open Technology Cultures,* Princeton University Press

EOSC (2021a), *Charter of the EOSC Task Force on Infrastructures for Quality Research Software,* https://www.eosc.eu/sites/default/files/tfcharters/eosca_tfinfrastructureforqualityresearchsoftware_draftcharter_20210614.pdf

EOSC (2021b), *Strategic Research and Innovation Agenda of the European Open Science Cloud (Version 1.0 15),* https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf

EPSRC (2018), *EPSRC software infrastructure strategy 2018,* UKRI, https://www.ukri.org/wp-content/uploads/2022/01/EPSRC-07012022-SoftwareInfrastructureStrategy2018.pdf

EPSRC (2020), *Understanding our portfolio: A gender perspective,* UKRI, 2020, https://www.ukri.org/publications/epsrc-understanding-our-portfolio-a-gender-perspective/

EPSRC (2021a), *Software for Research Communities Sift Panel 3,* Grants on the Web, https://gow.epsrc.ukri.org/NGBOViewPanel.aspx?PanelId=1-CP1SMD

EPSRC (2021b), *Software for Research Communities Sift Panel 1,* Grants on the Web, https://gow.epsrc.ukri.org/NGBOViewPanel.aspx?PanelId=1-CPQJFJ

EPSRC (2021c), *Software for Research Communities Sift Panel 2,* Grants on the Web, https://gow.epsrc.ukri.org/NGBOViewPanel.aspx?PanelId=1-CQ067P

EPSRC (2022a), *EPSRC expectations for equality, diversity and inclusion,* UKRI, https://www.ukri.org/publications/epsrc-expectations-for-equality-diversity-and-inclusion/

EPSRC (2022b), *Summary of Proposals by Quarter,* Grants on the Web, 2022, https://gow.epsrc.ukri.org/NGBOViewProposalSummary.aspx

EPSRC (2022c), *Software for Research Communities Full Proposal Prioritisation Panel,* Grants on the Web, https://gow.epsrc.ukri.org/NGBOViewPanel.aspx?PanelId=1-DCALVT

ESRC (2022), *ESRC Data Infrastructure Strategy 2022-27,* UKRI, https://www.ukri.org/wp-content/uploads/2022/06/ESRC-090622-DataInfrastructureStrategy2022To2027.pdf

European Commission & EOSC Executive Board (2021), *Digital skills for FAIR and Open Science: Report from the EOSC Executive Board Skills and Training Working Group,* Publications Office, https://data.europa.eu/doi/10.2777/59065

Ferenbaugh, C., Gesing, S., Hettrick, S., & Katz, D. S. (2022), *A Report on the SC21 Research Software Engineers in HPC (RSE-HPC-2021) Workshop,* BSSw, https://bssw.io/blog_posts/a-report-on-the-sc21-research-software-engineers-in-hpc-rse-hpc-2021-workshop

Ferreira da Silva, R., Casanova, H., Chard, K., Laney, D., Ahn, D., Jha, S., Goble, C., Ramakrishnan, L., Peterson, L., Enders, B., Thain, D., Altintas, I., Babuji, Y., Badia, R., Bonazzi, V., Coleman, T., Crusoe, M., Deelman, E., Di Natale, F., … Wozniak, J. (2021), *Workflows Community Summit: Bringing the Scientific Workflows Community Together,* Zenodo, 15 March 2021, https://doi.org/10.5281/ZENODO.4606958

Finkel, H. (2022), *ASCR Software-Stewardship Request for Information (RFI),* US Department of Energy, https://science.osti.gov/-/media/ascr/ascac/pdf/meetings/202203/ASCAC_202203_Finkel-RFI-Codesign-PDES.pdf

Flick, U. (2018), *Doing Triangulation and Mixed Methods (2nd ed.),* Sage Publications Ltd

Freitas, A., & Curry, E. (2016), Big Data Curation, in J. M. Cavanillas, E. Curry, & W. Wahlster, *New Horizons for a Data-Driven Economy,* Springer, https://link.springer.com/book/10.1007/978-3-319-21569-3

G7 Research Compact (2021), https://www.g7uk.org/wp-content/uploads/2021/06/G7-2021-Research-Compact-PDF-356KB-2-pages-1.pdf

Gabelica, M., Bojčić, R., & Puljak, L. (2022), Many researchers were not compliant with their published data sharing statement: Mixed-methods study, *Journal of Clinical Epidemiology,* S089543562200141X, https://doi.org/10.1016/j.jclinepi.2022.05.019

Goodey, G., & Hardeman, M. (2021), Three key findings from this year's State of Open Data survey. In Digital Science, *The State of Open Data 2021,* pp. 9–11, https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2021/17061347

Government Office for Science (2021), *Large-scale computing: The case for greater UK coordination,* GOV.UK, https://www.gov.uk/government/publications/large-scale-computing-the-case-for-greater-uk-coordination

GOV.UK (2021), *Writing about ethnicity,* December 2021, https://www.ethnicity-facts-figures.service.gov.uk/style-guide/writing-about-ethnicity#bame-and-bme

Graps, A. (2022), *How is China Educating a Quantum Workforce?,* Quantum Computing Report, https://quantumcomputingreport.com/how-is-china-educating-a-quantum-workforce/

Gruenpeter, M., Katz, D. S., Lamprecht, A.-L., Honeyman, T., Garijo, Struck, A., Niehues, A., Martinez, P. A., Castro, L. J., Rabemanantsoa, T., Chue Hong, N. P., Martinez-Ortiz, C., Sesink, L., Liffers, M., Fouilloux, A. C., Busse, C., Peroni, S., Martinez Lavanchy, P., Todorov, I., & Sinha, M. (2021), *Defining Research Software: A controversial discussion (Version 1),* Zenodo, https://doi.org/10.5281/ZENODO.5504016

Hardey & Leng (2020), *Linking Professional Skills to RSE Career Paths,* Zenodo, 8 October 2020, https://doi.org/10.5281/ZENODO.4073299

Harman, Y., & Darab, S. (2012), A call for slow scholarship: A case study on the intensification of academic life and its implications for pedagogy, *Review of Education, Pedagogy, and Cultural Studies,* 34(1–2), 49–60,

HESA (n.d.), *Cost centres (2012/13 onwards),* https://www.hesa.ac.uk/support/documentation/cost-centres/2012-13-onwards

HESA (2013), *JACS 3.0: Principal subject codes,* https://www.hesa.ac.uk/support/documentation/jacs/jacs3-principal

HESA (2021a), *Staff Record (2004/05—2020/21),* https://www.hesa.ac.uk/collection/c20025

HESA (2021b), *Table 3—HE staff by activity standard occupational classification 2020-21,* https://www.hesa.ac.uk/data-and-analysis/staff/table-3

Hettrick, S. (2014), *It's impossible to conduct research without software, say 7 out of 10 UK researchers,* Software Sustainability Institute, https://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers

Hettrick, S. (2020a), *How many RSEs? SORSE workshop: What do we (not) know about RSE?,* https://slides.com/simonhettrick/how-many-rses/fullscreen

Hettrick, S. (2020b), *Hidden REF reveals unsung heroes,* Research Professional News, https://www.researchprofessionalnews.com/rr-news-uk-views-of-the-uk-2020-2-hidden-ref-reveals-unsung-heroes/

Hettrick, S., Antonioletti, M., Carr, L., Chue Hong, N. P., Crouch, S., De Roure, D., Emsley, I., Goble, C., Hay, A., Inupakutika, D., Jackson, M., Nenadic, A., Parkinson, T., Parsons, M. I., Pawlik, A., Peru, G., Proeme, A., Robinson, J., & Sufi, S. (2014), *UK Research Software Survey 2014 [Data set],* Zenodo, https://doi.org/10.5281/zenodo.14809

Hettrick, S., Bast, R., Crouch, S., Wyatt, C., Collin, O., Botzki, A., Carver, J., Cosden, I., D'Andrea, F., Dasgupta, A., Godoy, W., Gonzalez-Beltran, A., Hamster, U., Henwood, S., Holmvall, P., Janosch, S., Lestang, T., May, N., Philips, J., ... Zhang, Q. (2022), *International RSE Survey 2022 (v0.9.1),* Zenodo, https://doi.org/10.5281/ZENODO.6884882

Houston, D., Meyer, L. H., & Paewai, S. (2006), Academic staff workloads and job satisfaction: Expectations and values in academe, *Journal of Higher Education Policy and Management,* 28(1), 17–30.

Hucka, M., & Graham, M. J. (2018), Software search is not a science, even among scientists: A survey of how scientists and engineers find software, *Journal of Systems and Software,* 141, 171–191, https://doi.org/10.1016/j.jss.2018.03.047

Ison, J., Ienasescu, H., Chmura, P., Rydza, E., Ménager, H., Kalaš, M., Schwämmle, V., Grüning, B., Beard, N., Lopez, R., Duvaud, S., Stockinger, H., Persson, B., Vařeková, R. S., Raček, T., Vondrášek, J., Peterson, H., Salumets, A., Jonassen, I., ... Brunak, S. (2019), The bio.tools registry of software tools and data resources for the life sciences, *Genome Biology,* 20(1), 164, https://doi.org/10.1186/s13059-019-1772-6

Janacek, S (2022), *UKRI funding programs for research software,* [slides not publicly available, permission was granted for the image used in this report].

Japan Science and Technology Agency (2017), *Strategy,* https://www.jst.go.jp/EN/about/strategy.html

Katz, D. S., McHenry, K., & Lee, J. S. (2021), *Senior level RSE career paths (with an s),* Zenodo, https://doi.org/10.5281/ZENODO.5531839

Katz, D. S., McInnes, L. C., Bernholdt, D. E., Mayes, A. C., Hong, N. P. C., Duckles, J., Gesing, S., Heroux, M. A., Hettrick, S., Jimenez, R. C., Pierce, M., Weaver, B., & Wilkins-Diehr, N. (2019), Community Organizations: Changing the Culture in Which Research Software Is Developed and Sustained. *Computing in Science & Engineering,* 21(2), 8–24, https://doi.org/10.1109/MCSE.2018.2883051

Kelle, U., & Erzeberge, C. (2004), Qualitative and Quantitative Methods: Not in Opposition, in U. Flick, E. von Kardoff, & I. Steinke, *A Companion to Qualitative Research* (pp. 172–177), Sage Publications.

Kipnis, D. (2021), Open data and the life sciences: The turning point, in Digital Science, *The State of Open Data 2021* (pp. 24–25) https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2021/17061347

Laird, J. (2021), *Announcing the panel on diversity and inclusion at Collaborations Workshop 2021,* Software Sustainability Institute, https://software.ac.uk/news/announcing-panel-diversity-and-inclusion-collaborations-workshop-2021

LERU (2022), *Possible research career paths in the United Kingdom (England),* https://www.leru.org/academic-career-maps-in-europe

Liang, T., Liu, C., Lin, T., & Lin, B. (2007), Effect of team diversity on software project performance, *Industrial Management & Data Systems,* 107(5), 636–653, https://doi.org/10.1108/02635570710750408

Maisuria, A., & Helmes, S. (2019), *Life for the Academic in the Neoliberal University,* Routledge

Martinez, P. A., Barker, M., Struck, A., Castro, L. J., Erdmann, Garijo, D., Gesing, S., Loewe, A., & Moldón, J. (2022), *A Survey on Adoption Guidelines for the FAIR4RS Principles,* Zenodo, https://doi.org/10.5281/ZENODO.6374598

Martinez, P. A., Gustafsson, J. R., & Crosby, V. (2022), *Visible Research Software Interest Group Terms of Reference,* https://docs.google.com/document/d/1wOOIlEywsSyK1OhLUEojBWJSwhpJLyS8jEsJsWsZP0M/edit?usp=embed_facebook

Martinez-Ortiz, C., Katz, D. S., Lamprecht, A.-L., Barker, M., Loewe, A., Fouilloux, A., Wyngaard, J., Garijo, D., Moldon, J., Castro, L. J., Wheeler, D., Albers, J. R. D., & Lee, A. (2022), *FAIR4RS: Adoption support,* Zenodo, https://doi.org/10.5281/ZENODO.6258365

MI Talent, UKRI, & Research England (2022), *The TALENT Commission: Technical skills, roles and careers in UK higher education and research,* https://www.mitalent.ac.uk/write/mediauploads/commission/The_TALENT_Commission_report_singles.pdf

Miller, R. B., & Hollist, C. S. (2007), Attrition Bias, in *N. J. Salkind (Ed.), Encyclopedia of Measurement and Statistics* (pp. 58–60), Sage Publications Ltd

Nangia, U., & Katz, D. S. (2017), *Track 1 Paper: Surveying the U.S. National Postdoctoral Association Regarding Software Use and Training in Research,* https://doi.org/10.6084/m9.figshare.5328442.v3

National Academies of Sciences, Engineering, and Medicine (2021), *Next Generation Earth Systems Science at the National Science Foundation* (p. 26042), National Academies Press, https://doi.org/10.17226/26042

NERC (2021), *Digitally Enabled Environmental Science: NERC Digital Strategy 2021-2030,* UKRI, https://www.ukri.org/wp-content/uploads/2022/05/NERC-170522-NERCDigitalStrategy-FINAL-WEB.pdf

Nogrady, B. (2019), Female researchers in Australia less likely to win major medical grants than males [news], *Nature,* https://www.nature.com/articles/d41586-019-03038-w

NSF (2022), *Quantum Information Science and Engineering Research at NSF,* https://www.nsf.gov/mps/quantum/quantum_research_at_nsf.jsp#Ongoing

NWO (2021), *National Roadmap for Large-scale Research Infrastructure,* https://www.nwo.nl/sites/nwo/files/media-files/National%20Roadmap%20for%20Large-scale%20Research%20Infrastructure%202021_0.pdf

OECD (2020), *Building digital workforce capacity and skills for data-intensive science,* OECD Science, Technology and Industry Policy Papers No. 90, https://www.oecd-ilibrary.org/science-and-technology/building-digital-workforce-capacity-and-skills-for-data-intensive-science_e08aa3bb-en

OECD (2021), *Recommendation of the Council concerning Access to Research Data from Public Funding,* https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347

Office for Artificial Intelligence, Department for Digital, Culture, Media & Sport, & Department for Business, Energy & Industrial Strategy (2021), *National AI Strategy,* GOV.UK, https://www.gov.uk/government/publications/national-ai-strategy

Office of the Chief Science Advisor of Canada (2020), *Roadmap for Open Science,* https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97992.html

Oliveira, D. F. M., Ma, Y., Woodruff, T. K., & Uzzi, B. (2019), Comparison of National Institutes of Health Grant Amounts to First-Time Male and Female Principal Investigators, *JAMA,* 321(9), 898, https://doi.org/10.1001/jama.2018.21944

Ouvrir la Science (2021), *Second National Plan for Open Science,* https://www.ouvrirlascience.fr/second-national-plan-for-open-science

Park, H., & Wolfram, D. (2019), Research software citation in the Data Citation Index: Current practices and implications for research software sharing and reuse. *Journal of Informetrics,* 13(2), 574–582, https://doi.org/10.1016/j.joi.2019.03.005

Parsons, M., Basden, A., Bower, R., Chue Hong, N. P., Constanzo, D., Witt, S., Del Debbio, L., Greed, G., Ham, D., Hasnip, P., Hines, A., Hettrick, S., Laizet, S., Thiyagalingam, J., Weinzierl, T., Wells, G., Woodley, S., & Wood, N. (2021), *ExCALIBUR Research Software Engineer Knowledge Integration Landscape Review,* Zenodo, https://doi.org/10.5281/ZENODO.4986062

Pawar, M. (2004), *Data Collecting Methods and Experiences,* New Dawn Press Group.

Philippe, O. (2018), *What do we know about RSEs? Results from our international surveys,* Software Sustainability Institute, https://www.software.ac.uk/blog/2018-03-12-what-do-we-know-about-rses-results-our-international-surveys

Professionalising Data Stewardship IG (2020), *Professionalising Data Stewardship IG Charter,* Research Data Alliance, https://www.rd-alliance.org/sites/default/files/case_statement/Update%20October_Professionalising_data_stewardship_charter_proposal_20201021_0.pdf

Responses to the Request for Information on Stewardship of Software for Scientific and High-Performance Computing (2021), https://www.osti.gov/servlets/purl/1843576/

RSE International Survey 2022 (2022), https://softwaresaved.github.io/international-survey-2022/section/sociodemography/

Smith, A. M., Katz, D. S., Niemeyer, K. E., & FORCE11 Software Citation Working Group (2016), Software citation principles, *PeerJ Computer Science,* 2, e86, https://doi.org/10.7717/peerj-cs.86

Softwaresaved/international-survey (2019), *GitHub,* https://github.com/softwaresaved/international-survey

SPARC Europe, & Digital Curation Centre (2019), *An Analysis of Open Science Policies in Europe v4,* Zenodo, https://doi.org/10.5281/ZENODO.3379705

SSI (2021), *Carpentry Programmes,* Software Sustainability Institute, https://www.software.ac.uk/programmes-and-events/carpentry-programmes

SSI (2022a), *Closing soon: Survey of digital methods & software for economics and social sciences,* Software Sustainability Institute, https://software.ac.uk/news/closing-soon-survey-digital-methods-software-economics-and-social-sciences

SSI (2022b), *Reminder: Survey of digital/software requirements for the AHRC research community,* Software Sustainability Institute, https://www.software.ac.uk/news/reminder-survey-digitalsoftware-requirements-ahrc-research-community

Stanford Software Survey (2020), https://stanford-rc.github.io/stanford-software-survey

Strasser, C., Hertweck, K., Greenberg, J., Taraborelli, D., & Vu, E. (2022), *10 Simple Rules for Funding Scientific Open Source Software,* Zenodo, https://doi.org/10.5281/ZENODO.6611500

Teperek, M., Cruz, M., & Kingsley, D. (2022), Time to re-think the divide between academic and support staff, *Nature,* d41586-022-01081–01088, https://doi.org/10.1038/d41586-022-01081-8

Tracy, S. J. (2020), *Qualitative Research Methods,* John Wiley & Sons Ltd

UKRI (2021), *Initiatives boost health and bioscience skills and industry,* https://www.ukri.org/news/initiatives-boost-health-and-bioscience-skills-and-industry/

UKRI (2022a), *Embed digital skills in arts and humanities research,* https://www.ukri.org/opportunity/embed-digital-skills-in-arts-and-humanities-research/

UKRI (2022b), *Policy framework on research data,* https://www.ukri.org/about-us/epsrc/our-policies-and-standards/policy-framework-on-research-data/scope-and-benefits/#contents-list

UKRI (2022c), *UKRI position on next phase of large scale compute investments,* https://www.ukri.org/what-we-offer/creating-world-class-research-and-innovation-infrastructure/digital-research-infrastructure/ukri-position-on-next-phase-of-large-scale-compute-investments/

UKRI (2022d), *Digital research infrastructure,* https://www.ukri.org/what-we-offer/creating-world-class-research-and-innovation-infrastructure/digital-research-infrastructure/

UNESCO (2021), *Recommendation on Open Science,* https://unesdoc.unesco.org/ark:/48223/pf0000378381.locale=en

US-RSE (2021), *Mission,* https://us-rse.org/mission/

Vasilescu, B., Posnett, D., Ray, B., van den Brand, M. G. J., Serebrenik, A., Devanbu, P., & Filkov, V. (2015), Gender and Tenure Diversity in GitHub Teams, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,* 3789–3798, https://doi.org/10.1145/2702123.2702549

VSNU, NFU, KNAW, NWO, & ZonMw (2019), *Room for everyone's talent,* http://vsnu.nl/recognitionandrewards/wp-content/uploads/2019/11/Position-paper-Room-for-everyone%E2%80%99s-talent.pdf

Wijnen, M. N., Massen, J. J. M., & Kret, M. E. (2021), Gender bias in the allocation of student grants, *Scientometrics,* 126(7), 5477–5488, https://doi.org/10.1007/s11192-021-03985-0

Wilkinson, M., Lawrence, B., Prescott, A., De Roure, D., Ford, D., Parsons, M., Wingate, B., Coveney, P., Morrell, S., Birney, E., Nilanjan Chakraborty, & Sijacki, D. (2021), *UKRI Supercomputing Science Case,* Zenodo, https://doi.org/10.5281/ZENODO.4985325

Woolston, C. (2022), Why science needs more research software engineers, *Nature,* https://doi.org/10.1038/d41586-022-01516-2

APPENDIX A: STUDY PERSONNEL

Professor Neil Chue Hong (Principal Investigator) is Director of the SSI and Professor of Research Software Policy and Practice at EPCC, which manages the UK national supercomputing facilities and Edinburgh International Data Facility. His recent work in this area includes the 2020 OECD report on Building digital workforce capacity and skills for data-intensive science, the section on software and computing in the 2021 STFC Particle Physics Technology Advisory Panel's "UK Technology R&D Roadmap for Accelerators, Detectors, and Computing", and Six Recommendations for Implementation of FAIR Practice. He is a member of the BBSRC Transformative Technologies SAP, Chair of the MetOffice/UKRI ExCALIBUR Steering Committee, and Editor-in-Chief of the Journal of Open Research Software.

Professor Simon Hettrick (Co-Investigator) works with stakeholders from across the research community to develop policies that support research software. Simon's research focuses on the use of software in the research community with the aim of understanding practices and demographics. In this role, he conducted the first study of software reliance in academia. He orchestrated the campaign that led to the recognition of RSEs. He is a leading expert in surveying the field of RSE and the project director for the International RSE Survey.

Dr Rebecca Taylor (Co-Investigator) is a sociologist with 16 years' experience in policy research, and a track record in managing and delivering projects for a range of funders including ESRC, Nuffield, Age UK, and Department for Work and Pensions. She is a qualitative methods specialist with expertise in longitudinal research and research ethics. She is a co-director of the Work Futures Research Centre and has an interest in digital labour, digital transformation, and sustaining digital infrastructure.

Dr Michelle Barker (Consultant) is a sociologist and the Director of ReSA. She is expert in open science, research software, digital workforce capability, and digital research infrastructure. She co-authored Research Software Capability in Australia and Digital Skills for FAIR and Open Science. As Chair of the OECD expert group she also co-authored Building digital workforce capability and skills for data-intensive science, which made policy recommendations to member states. Michelle is a former Director of the ARDC, where she led the strategic planning for the Australian government's $180 million investment in ARDC, and the national research software infrastructure investment.

Dr Elena Breitmoser (Project Manager) works at the EPCC, University of Edinburgh. Her background is theoretical astrophysics, and she has worked in HPC and data science as an RSE for many years.

Dr Ioanna Lampaki is a data scientist working at the EPCC, University of Edinburgh. Her background is in statistics and she has worked in a variety of data analytics and data science projects.

Dr Anthony Quinn is a mixed-methods researcher who conducts interdisciplinary research across Electronics & Computer Science and Medicine at the University of Southampton. Within the SSI, he has worked on sustaining digital infrastructure and open source software.

Dr Philippa Broadbent is an RSE in the Southampton Research Software Group at the University of Southampton. Her background is in cognitive and health psychology.

We would also like to thank Philip Grylls from the University of Southampton for his contributions.

APPENDIX B: SURVEY INFORMATION SHEET AND QUESTIONS

INTRODUCTION AND WELCOME

Welcome to this survey on software, skills and infrastructure required for research involving computing run by the University of Edinburgh and University of Southampton. This survey is divided into 6 brief sections and should take approximately 20 minutes to complete. If you require any assistance whilst completing this survey then please get in touch with us, software-survey@software.ac.uk. This study was certified according to the EPCC Research Ethics Process. Please take time to read the following information carefully and keep it for your records.

The purpose of this study is to deliver a better understanding of the software and skills required for research computing (including High Performance Computing (HPC), High-Throughput Computing (HTC)/Cloud Computing, Artificial Intelligence/ Machine Learning and Data Science) in the United Kingdom and recommendations for how policies and support for these should be structured. The results of this study aim to contribute towards EPSRC and UKRI policy and help direct funding.

Taking part in the study, risks and benefits

You are invited to participate in this study because you have been identified, or identify, as someone who uses, or supports the use of research computing, e.g., a user, researcher, research software engineer, infrastructure provider, or trainer/educator.

Participation in this study is entirely up to you. You can withdraw from the study at any time, without giving a reason. Your rights will not be affected. If you wish to withdraw, contact the Principal Investigator, Neil Chue Hong. We will stop using your data in any publications or presentations submitted after you have withdrawn consent. However, we will keep copies of your original consent, and of your withdrawal request. If you decide to take part in this study you will be answering questions regarding the use of large-scale computing, software and tools you may be using in your work along with questions regarding required skills and gaps you may have identified related to available support and required skills in the field.

There are no significant risks associated with participation in this study. On completion of the survey, you may choose to be entered into a draw to receive shopping vouchers.

What will happen to the results of this study?

A report and pseudonymised datasets will be published as research outputs. Where a low number of participants in a category might allow identification even after replacement of easily attributable identifiers, only aggregate data will be published. With your consent, information can also be used for future research. Your data may be archived for a minimum of two years. The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymised: we will remove any information that could, in our assessment, allow anyone to identify you.

Data protection, rights and confidentiality

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your personal data will only be viewed by the research team. Pseudonymised data will be shared with our funder, EPSRC, and the partners on this project. All electronic data will be stored on a password-protected encrypted computer, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, or Sharepoint) and all paper records will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk.

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance with Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk. For general information about how we use your data, go to: edin.ac/privacy-research.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Mr. Neil Chue Hong, n.chuehong@epcc.ed.ac.uk. If you wish to make a complaint about the study, please contact the Chair of the University of Edinburgh's College of Science and Engineering Research Ethics & Integrity Committee, Prof. Andy Mount, a.mount@ed.ac.uk, or fill out the Research Misconduct Informal Reporting Form. (https://www.ed.ac.uk/science-engineering/research/research-ethics/research-misconduct). When you contact us, please provide the study title and detail the nature of your complaint.

Updated information

If the research project changes in any way, an updated Participant Information Sheet will be made available on the Software Sustainability Institute website: https://www.software.ac.uk/.

Consent

By proceeding with the study, I agree to all of the following statements:

- I have read and understood the above information.
- I understand that my participation is voluntary, and I can withdraw at any time.
- I consent to my anonymised data being used in academic publications and presentations.
- I consent to my data being used in future ethically approved research.
- I consent to the results of the study being shared and published as research outputs.
- I consent to my anonymised data being shared with EPSRC and the project partners.
- I consent to all of the above

1. I consent to all of the above.

YOU AND YOUR CAREER

2. Which organisation do you work for?

3. Are you based in the United Kingdom?

4. In which disciplines do you work? Select all that apply.

- Agriculture & related subjects
- Architecture, building & planning
- Biological sciences
- Business & administrative studies
- Computer science
- Creative arts & design
- Education
- Engineering & technology
- Historical & philosophical studies
- Languages
- Law
- Mass communications & documentation
- Mathematical sciences
- Medicine and dentistry
- Physical sciences
- Social studies
- Subjects allied to medicine
- Veterinary science
- Combined (multidisciplinary)

For details of what subjects are classified in each of these areas, please see https://www.hesa.ac.uk/support/documentation/jacs/jacs3-principal.

5. What is currently your primary role? Select only one.

- Research
- Teaching
- Research Software Engineer
- Technician
- Research Support
- Management
- Other

5a. If you selected Other, please specify:

6. At what career stage are you currently? Select the one career stage that most accurately describes your role.

Phase 1 – Junior (e.g., PhD candidate, Junior Research Software Engineer)

Phase 2 – Early (e.g., Research Assistant/Associate/Fellow, first grant holder, Lecturer, Research Software Engineer)

Phase 3 – Mid / Recognised (e.g., Senior Lecturer, Reader, Senior Researcher, Senior Research Fellow, Senior Research Software Engineer, Research Software Group Leader, Senior Data Scientist)

Phase 4 – Established / Experienced / Senior (e.g., Professor, Director of Research Computing, Distinguished Engineer, Chief Data Scientist)

7. Where have you applied for funding, or who funds your work?

	Have applied to	Have been funded by
AHRC	☐	☐
BBSRC	☐	☐
EPSRC	☐	☐
ESRC	☐	☐
European Commission/ European research council	☐	☐
Innovate UK	☐	☐
MRC	☐	☐
NERC	☐	☐
STFC	☐	☐
Other	☐	☐

7a. If you selected Other, please specify where have you applied for and where you got funding from.

SOFTWARE USE

The following questions help us understand the variety of software being used to support research, and how this is discovered and chosen.

By "software", we mean any software or digital tool you have used in the course of your research that has helped you undertake your research or produce a research output (e.g., a publication).

This can be anything from a short script of code to help you clean your data, to web/mobile apps, or to a fully-fledged software suite or specialised toolset.

It includes code that you have written yourself or code written by someone else.

8. How important is software to your research

> Not important at all
> Somewhat important
> Important
> Very important
> Essential

9. How do you discover the software you use in your research? Select all that apply.

> Recommended by colleagues
> Heard about it in a paper / talk / event
> Search engine (e.g., Google)
> Recommended / provided by my organisation
> My research does not currently require new software
> Other

9a. If you selected Other, please specify:

10. Are you free to use the software you want to use?

> Yes
> No

10a. If you answered No in the above question what is the reason? Select all that apply.

> Because I have to use the software that has historically been used in my group/field
> Because there are legal constraints to what I am allowed to use (e.g., I am not allowed to use open source software)
> Because I don't have resources to purchase the software I want to use
> Because the software I want to use is not available to me for a reason not listed above

11. What are the names of (up to three) pieces of research software you've used to support your research in the last month? We are interested in the software that you believe is most important to your work. Please separate your answers using semicolons.

SOFTWARE DEVELOPMENT

The following questions help us understand how people working in research create or modify the software they use.

12. Do you develop software (i.e. write your own code) for your research?

> No
> Yes

12a. What programming language(s) do you usually use for developing code? Please separate multiple answers by semicolon.

> Yes
> No

12.b.i. Who was the funder? Select all that apply.

> AHRC
> BBSRC
> EPSRC
> ESRC
> European Commission/European Research Council
> MRC
> NERC
> STFC
> Innovate UK
> My institution
> Other

12.b.i.a. If you selected Other, please specify:

12.b.ii. What type of funding did you apply for? Select all that apply.

> Software grant
> Standard research grant
> Programme grant
> Fellowship
> Any other type of research grant

13. What is your opinion on including software development costs in a funding proposal?

20. Significantly decreases chances of success
21. Somewhat decreases chances of success
22. Neither decreases nor increases chances of success
23. Somewhat increases chances of success
24. Significantly increases chances of success

13a. Why do you think this? Could you briefly describe any experiences you've had? Please use a maximum of 100 words.

14. Over the last five years, which of the following practices have been part of your standard research process? Select all that apply.

- I haven't shared my research software
- I share my research software with my collaborators
- I share my research software with individuals/groups that have requested access
- I share my research software publicly
- I license my research software to allow it to be shared
- I cite my software in publications and other outputs from my research
- I create a unique identifier (e.g., a DOI) to make my research software findable
- I record my software as a research output (e.g., in ResearchFish, Pure, Sympletic)

DIGITAL RESEARCH INFRASTRUCTURE

The following questions help us understand what infrastructure people use, the benefits of and the barriers to moving to a different infrastructure.

15. Which of the following do you currently use or plan to use in the next five years for the computational parts of your research?

	Haven't heard of this	Currently using	Plan to use	Don't plan to use
A laptop/desktop	☐	☐	☐	☐
A server/cluster operated by your research group/department	☐	☐	☐	☐
A central research computing service operated by your institution/organisation	☐	☐	☐	☐
Computing services at a data centre or at a data safe haven (including trusted research environments)	☐	☐	☐	☐
An Academic Cloud Computing/High-Throughput Computing service	☐	☐	☐	☐
A Commercial Cloud provider, e.g., Microsoft Azure or Amazon Web Services	☐	☐	☐	☐
A UK Tier 2 high performance computing service	☐	☐	☐	☐
A UK Tier 1 (national) high performance computing service (e.g., ARCHER2)	☐	☐	☐	☐
An Exascale HPC service	☐	☐	☐	☐
Specialised Machine Learning hardware	☐	☐	☐	☐
Other	☐	☐	☐	☐

15a. If you selected Other, please specify.

16. Do you know who funds the computational infrastructure you use for your research?

- Yes
- No

16a. Do you use any non-UKRI funded computational infrastructure in preference to a UKRI one? If so, what are the main reasons?

- Because the resources are provided by collaborators
- Because the infrastructure I need is outside the UK
- Because UKRI provided infrastructure is not sufficient
- Because of lower costs
- Other

16a.i. If you selected Other, please specify:

17. Over the last five years, once you've completed a research project/study, which of the following have you used to store the data that were generated after your research has been completed? Select all that apply.

- On a laptop/desktop (including external hard drives)
- In a shared cloud service (e.g., personal Google Drive, Dropbox)
- In a shared area provided by your organisation (shared network drive, institutional Sharepoint, Teams, Google Drive, etc.)
- In a data repository provided by your organisation (Pure, ePrints, etc.)
- In a data repository outside of your organisation (UK Data Service, Zenodo, etc.)
- Other

17a. If you selected Other, please specify:

18. Over the last five years, which of the following statements best apply to how you share your research data? Select all that apply.

- I haven't shared my research data
- I have only shared my research data with colleagues
- I make my research data available on request
- I share my research data via a personal or project website
- I have published my research data in a repository
- I do not have research data to share
- Other

18a. If you selected Other, please specify:

19. In the next year, do you feel that your research will be limited by access to computational and/or data research infrastructure?

- Don't know
- No
- Yes

19a. If you selected Yes, does anything prevent you from using larger scale computational and data resources ? Please use a maximum of 100 words.

20. Who do you ask for advice on choosing the computational and/or data research infrastructure (computing, storage and network) you use?

- I don't need help choosing digital research infrastructure
- My PI /Supervisor
- My colleagues/peers
- My local RSE group
- My organisation's central/departmental IT services
- My organisation's research computing team
- Online resources and guides
- Other

20a. If you selected Other, please specify:

21. Would you like to expand on any of the answers you've given in this section?

ROLES, SKILLS AND POLICIES

The following questions help us understand the prevalence of roles, skills and policies related to the use of computing in research.

22. Which of the following supporting roles are you aware of, and which do you currently work with?

	Haven't heard of this	Currently using	Plan to use
Research Software Engineers	☐	☐	☐
Research Data Managers	☐	☐	☐
Data Stewards	☐	☐	☐
IT Services/Helpdesk	☐	☐	☐
Research Librarian	☐	☐	☐
Research Manager /Research Officer	☐	☐	☐
Other	☐	☐	☐

22a. If you selected Other, please specify:

23. Over the last five years, have you learned new skills or knowledge in relation to any of the following? Select all that apply.

- Data collection (e.g., new technological tools, web scraping)
- Data analysis (e.g., statistical, tools visualison tools)
- Data management/storage (e.g., ethics, GDPR, repositories)
- Data sharing (e.g., licensing, archiving, creating DOIs)
- None of the above

24. Over the last five years, have you learned new skills or knowledge in relation to any of the following? Select all that apply.

- Training to use a specific piece of software
- Software requirements collection (e.g., landscape review, usage study)
- Software development (e.g., new languages, design, testing)
- Large-scale computing architectures and platforms (e.g., HPC, GPUs, distributed computing)
- Software sharing (e.g., licensing, repositories, distribution, deployment)
- None of the above

25. Where did you acquire your skills and knowledge about data and/or software? Select all that apply.

- Undergraduate courses
- Masters/Doctoral training (completed during PGR and PGT courses)
- Training course provided in-house by my organisation(while employed)
- Training course provided externally (academic and commercial providers)
- Workshops and Events (e.g., at a conference)
- Learning on the job-self taught (e.g., websites, embedded help/training)
- Learning on the job-peer taught (e.g., learning from project teams, members, colleagues, forums)
- Other

25a. If you selected Other, please specify:

26. Thinking about your current skill set, what is the main data, software or computing skill that you would like to obtain to improve your research/work?

27. Do any of the following issues prevent you from gaining the skill you entered above? Select all that apply.

- I'm not aware of any training or opportunities to learn this skill
- I'm aware of training/learning opportunities but I have no time to engage
- There are only external training/learning opportunities, but I don't have funding to attend them
- My institution doesn't support the infrastructure necessary to make use of the skill
- My supervisor/manager does not support me
- None of the above
- Other

27a. If you selected Other, please specify:

28. Is there anything that you would like to change to enable the use of computing in research? Please use a maximum of 100 words.

APPENDIX C: SURVEY RESPONSES

APPENDIX C: GROUP INTERVIEW INFORMATION SHEET, CONSENT FORM AND TOPIC GUIDE

Participant Information Sheet

Study Title: Software Skills for Large Scale Computing: collective evidence to develop a National Research Software Strategy

Researchers: Professor Simon Hettrick, Dr Rebecca Taylor, Dr Anthony Quinn

ERGO number: 71904

You are being invited to take part in the above research study. To help you decide whether you would like to take part or not, it is important that you understand why the research is being done and what it will involve. Please read the information below carefully, and ask questions if anything is not clear or you would like more information before you decide to take part in this research. You may like to discuss it with others but it is up to you to decide whether or not to take part. If you are happy to participate you are asked to electronically sign a consent form, sent to you alongside this Participant Information Sheet. Signed forms will be transferred via safesend https://safesend.soton.ac.uk

What is the research about?

Our research study will explore software and roles and skills that are required to support large-scale research computing. Through research with early career researchers, established researchers and research software engineers, we will seek to identify immediate needs and barriers within different research communities as well as relevant funding and policies that are currently in place.

Why have I been asked to participate?

You indicated to us that you would be willing to participate in follow up research after completing our survey. We are keen to hear the views of a variety of researchers including early career, established researchers, research software engineers or those with a specific knowledge of large scale computing. We are interested in understanding your role(s), skills(s) and thoughts as well as the software that you use or wish to use as part of your work.

What will happen to me if I take part?

We will hold an online focus group session that will last between 60-90 minutes. There should be 4 or 5 other participants taking part in the same session as you. You will be asked to talk about your role, skills, experience of research and software and the available infrastructure and support for what you do. You can discuss and compare experiences with other participants in the group. We will seek to organise our groups as much as possible on the basis of demographic variables such as age, gender, ethnicity. In some cases (where it is difficult to assemble a group or if a participant prefers), we will conduct a one-to-one interview. Within these interviews, you will be asked the same questions you would have been asked within a focus group. The session will be audio recorded. You will receive a £50 shopping voucher as a thank you for your participation.

Are there any benefits in my taking part?

We need to understand more about the software, roles and skills required for large scale research computing. By gathering this important knowledge, we seek to develop a sustainable and accessible research infrastructure across disciplines and relevant communities. Your contribution can enable us to develop a more resilient software infrastructure. As a token of our gratitude, we will provide you with a shopping voucher [£50 GBP] after your participation in a focus group/ interview.

Are there any risks involved?

There are no risks involved in participation.

What data will be collected?

At the start of the session, you will be given an opportunity to ask questions. You will be individually asked if it is ok for the researcher to begin recording. An audio recording of the session will be transcribed and pseudonymised so that it does not include any identifiable information. The audio recording will be destroyed once the transcription has been completed. The focus group and interview transcripts will then be stored within a password protected folder on the University of Edinburgh Sharepoint. These data will be analysed using analysis software for qualitative data.

Will my participation be confidential?

Your participation and the information that we collect about you during the course of the research be kept strictly confidential. At the start of the focus group, participants will be asked not to discuss any aspect of the session after it has been completed.

Your personal details will remain strictly confidential. Only members of the research team at the University of Southampton and the University of Edinburgh, or responsible members of either institution, will be given access to data about you for research analysis, monitoring purposes and/ or to carry out an audit of the study to ensure the research is complying with applicable regulations. Individuals from regulatory authorities (people who check that we are carrying out the study correctly) may require access to your data. All of these people have a duty to keep your information, as a research participant, strictly confidential.

We will transcribe the session. Identifiable information within the transcripts such as names of people or affiliations will be de-identified to hide the identity of study participants. All data will be stored on University of Edinburgh project SharePoint which is only accessible to the research teams at both institutions. The findings will be disseminated in a report to the EPSRC. The funder will not know who we have talked to and any quotes used in the reporting will be anonymised.

Do I have to take part?

No; it is entirely up to you to decide whether or not you take part. If you decide that you want to voluntarily take part then you will need to electronically sign the consent form attached to this document and return it to the study team via safesend https://safesend.soton.ac.uk. Please do not hesitate to get in touch if you have any questions in relation to this participant information sheet or the consent form.

What happens if I change my mind?

You have the right to change your mind and withdraw at any time, prior to completion of the study in June 2022 or publication of findings (whichever comes first). This is because it will not be possible to identify and retract your data after this point. If you do decide to withdraw your participation then this can be done by contacting a member of the research team without a reason, without penalty and without your participant rights being affected. You will still receive your shopping voucher if you have completed a focus group or an interview.

What will happen to the results of the research?

Research findings may be made available in reports or publication however these will not include any identifiable information. The results of this study will be presented within a report to the funder (EPSRC), will form the basis of a National Research Software Strategy and may be written up for conference or journal publication. In line with best practice, we ask for consent for your pseudonymised contribution to be stored in a repository (currently Pure) at the University of Southampton and available for future research as required by the funder EPSRC.

Where can I get more information?

You are most welcome to email the Principal Investigator for this study at: sjh@ecs.soton.ac.uk

What happens if there is a problem?

If you have a concern or query about any aspect of this study, you should speak to the researchers on this study who will do their best to answer your questions. If you remain unhappy or have a complaint about any aspect of this study, please contact the University of Southampton Research Integrity and Governance Manager (02380595058, rgoinfo@soton.ac.uk**).**

Data Protection Privacy Notice

The University of Southampton conducts research to the highest standards of research integrity. As a publicly-funded organisation, the University has to ensure that it is in the public interest when we use personally-identifiable information about people who have agreed to take part in research. This means that when you agree to take part in a research study, we will use information about you in the ways needed, and for the purposes specified, to conduct and complete the research project. Under data protection law, 'Personal data' means any information that relates to and is capable of identifying a living individual. The University's data protection policy governing the use of personal data by the University can be found on its website (https://www.southampton.ac.uk/legalservices/what-we-do/data-protection-and-foi.page).

This Participant Information Sheet tells you what data will be collected for this project and whether this includes any personal data. Please ask the research team if you have any questions or are unclear what data is being collected about you.

Our privacy notice for research participants provides more information on how the University of Southampton collects and uses your personal data when you take part in one of our research projects and can be found **here.** Any personal data we collect in this study will be used only for the purposes of carrying out our research and will be handled according to the

University's policies in line with data protection law. If any personal data is used from which you can be identified directly, it will not be disclosed to anyone else without your consent unless the University of Southampton is required by law to disclose it.

Data protection law requires us to have a valid legal reason ('lawful basis') to process and use your Personal data. The lawful basis for processing personal information in this research study is for the performance of a task carried out in the public interest. Personal data collected for research will not be used for any other purpose.

For the purposes of data protection law, the University of Southampton is the 'Data Controller' for this study, which means that we are responsible for looking after your information and using it properly. The University of Southampton will keep identifiable information about you for 10 years after the study has finished after which time any link between you and your information will be removed.

To safeguard your rights, we will use the minimum personal data necessary to achieve our research study objectives. Your data protection rights – such as to access, change, or transfer such information - may be limited, however, in order for the research output to be reliable and accurate. The University will not do anything with your personal data that you would not reasonably expect.

If you have any questions about how your personal data is used, or wish to exercise any of your rights, please consult the University's data protection webpage (**https://www.southampton.ac.uk/legalservices/what-we-do/data-protection-and-foi.page**) where you can make a request using our online form. If you need further assistance, please contact the University's Data Protection Officer **(data.protection@soton.ac.uk).**

Thank you for taking the time to read this.

CONSENT FORM

Study title: Software and Skills for Large Scale Computing: collecting evidence to develop a National Research Software Strategy

Researchers: Prof Simon Hettrick, Dr Rebecca Taylor, Dr Anthony Quinn

ERGO number: 71904

Please indicate your agreement to each of these items by providing your initials and signing and dating the form.

Signed forms should be returned via safesend https://safesend.soton.ac.uk.

	Your Initials
I have read and understood the information sheet (08/04/2022, Version No. 3) and have had the opportunity to ask questions about the study.	
I agree to take part in this research project and agree for my data to be used for the purpose of this study.	
I understand my participation is voluntary and that I may withdraw for any reason without my participation rights being affected.	
I understand that I must keep the focus group discussion confidential and that my anonymity cannot be guaranteed in a focus group but that any information collected by the researchers will be kept confidential and all participants will be asked to keep the discussions confidential.	
I give permission for my focus group contributions to be audio recorded and for a pseudonymised transcript to be deposited in the researcher's data space (University of Edinburgh SharePoint) and repository (Pure) as described in the Participant Information Sheet so it can be used for future research and learning within the EPSRC community.	
I understand that I may be quoted directly in research reports but that I will not be directly identified (e.g. my name will not be used and identifiable details will be changed).	
I understand that I can withdraw from this study prior to completion of the study or publication of findings (whichever comes first). I understand that should I withdraw from the study after this point, then the information collected about me up to this point may still be used for the purposes of achieving the objectives of the study only.	

Name of participant (print name)...

Signature of participant (digital signature)..

Date...

Name of researcher ....

Signature of researcher (digital signature) ....

Date...

TOPIC GUIDE

Introduction

Each participant introduces themselves/institution/role/career stage.

1. Software

Thinking about software and tools you use in your current (last five years) research:

> What do you use? (Probe on: types of code/ programming languages, web/ apps, data analysis visualisation tools, versioning software, libraries)

> Why do you use these tools? What are the challenges of using these tools? (Probe on: power, memory, disk space, technical support)

2. Skills and training

Now we're going to think about how you acquired the skills to use and/ or develop the current types of software and tools you use in your research

> How and when have you learned how to use these tools in your research/ work? (Probe on: doctoral training, professional training internal and external.) (Probe on motivations for training/ costs of training.)

> Are there other forms of training you haven't done or wouldn't do? Why?

3. Thinking about your institution

> If you required software or particular tools, would you know how to access or request this? What about training?

> What about if you needed particular software expertise on a project? How might your institution help with that (or not)?

4. What about beyond your institution?

> Are there other forms of digital research infrastructure that you use? What scale?

> Would there be benefits to your work of being able to use computational infrastructure of a larger scale? Is there a need to move to the next level of computing?

5. Thinking about your funding and funding in general [mainly in the UK]

> What is your view on the support available via its funding (Probe on: the structure of calls and bid requirements? Reviewers and funding panels? Availability of advice and support?)

> How do you think software grants should be reviewed?

> What kind of things encourage or discourage you from submitting a funding bid?

6. Final thoughts

> What about software in future? Where do you see software in your field going?

> What skills would be useful for you in the future and how would you want to acquire them? (Probe on: How should training for that be delivered?)

> How do you best communicate the value of [your] research software?

APPENDIX D: STAKEHOLDER INTERVIEW INFORMATION SHEET AND QUESTIONS

PARTICIPANT INFORMATION SHEET

Project title: Software and Skills for Large Scale Computing

Principal investigator: Mr. Neil Chue Hong

Researchers collecting data: Michelle Barker, Neil Chue Hong, Elena Breitmoser

Funder (if applicable): EPSRC

This study was certified according to the EPCC Research Ethics Process. Please take time to read the following information carefully. You should keep this document for your records.

Who are the researchers?

- Mr. Neil Chue Hong, EPCC, University of Edinburgh
- Prof. Simon Hettrick, School of Electronics and Computer Science, University of Southampton
- Dr. Rebecca Taylor, Department of Sociology. Social Policy and Criminology, University of Southampton
- Dr. Anthony Quinn, Faculty of Social Sciences & Department of Electronics and Computer Science, University of Southampton
- Dr. Elena Breitmoser, EPCC, University of Edinburgh
- Dr. Ioanna Lampaki, EPCC, University of Edinburgh
- Dr. Philip Grylls, School of Electronics and Computer Science, University of Southampton
- Dr. Michelle Barker, Open Science Consultant

What is the purpose of the study?

The purpose of this study is to deliver a better understanding of the software and skills required for large-scale research computing (including High Performance Computing (HPC), High Throughput Computing (HTC)/Cloud Computing, Artificial Intelligence/Machine Learning and Data Science) in the United Kingdom and recommendations for how policies and support for these should be structured.

Why have I been asked to take part?

You have been identified, or identify, as a stakeholder in large-scale research computing, e.g., an infrastructure provider, research consortia, and/or funder.

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, without giving a reason. Your rights will not be affected. If you wish to withdraw, contact the PI. We will stop using your data in any publications or presentations submitted after you have withdrawn consent. However, we will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

If you decide to take part in this study then a 1-hour interview will be organised. The interview questions will be provided in advance of the interview. You will be answering questions regarding challenges and opportunities related to the use of large-scale computing, software and tools, and required skills. Additional ad hoc questions may also be asked during the interview to provide clarifying information. The video will be recorded for the purposes of transcription only.

Compensation

Taking part in this study will not be compensated.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

What will happen to the results of this study?

A report will be published as a research output. Your answers will be used to inform the study's outcomes. Quotes or key findings will be anonymised: We will remove any information that could, in our assessment, allow anyone to identify you. If the researchers would like to identify you and any responses you made, then they will request your consent to do so after the interview.

With your consent, information can also be used for future ethically approved research. Your data may be archived for a minimum of 2 years. The results of this study may be summarised in published articles, reports and presentations.

DATA PROTECTION AND CONFIDENTIALITY

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your personal data will only be viewed by the research team. Pseudonymised data will be shared with our funder, EPSRC, and the partners on this project

All electronic data will be stored on a password-protected encrypted computer, or on the University of Edinburgh's secure encrypted cloud storage services (DataShare, ownCloud, or Sharepoint) and all paper records will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk, the exception being the interview recordings before they are transcribed.

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance with Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at **dpo@ed.ac.uk.**

For general information about how we use your data, go to: edin.ac/privacyresearch.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Mr. Neil Chue Hong, **n.chuehong@epcc.ed.ac.uk.**

If you wish to make a complaint about the study, please contact the Chair of the University of Edinburgh's College of Science and Engineering Research Ethics & Integrity Committee, Prof. Andy Mount, **a.mount@ed.ac.uk**, or fill out the Research Misconduct Informal Reporting Form. When you contact us please provide the study title and detail the nature of your complaint.

Updated information

If the research project changes in any way, an updated Participant Information Sheet will be made available on the Software Sustainability Institute website: https://www.software.ac.uk/research-studies

Consent

By proceeding with the study, I agree to all of the following statements:

- I have read and understood the above participant information for this study, I have had the opportunity to ask questions, and any questions I had were answered to my satisfaction.
- I understand that my participation is voluntary, and I can withdraw at any time. Withdrawing will not affect my rights.
- I consent to my anonymised data being used in academic publications and presentations.
- I consent to the results of the study being shared and published as research outputs.
- I consent to my anonymised data being shared with EPSRC and the project partners.

Please tick yes or no for each of the following statements and return this form prior to the interview.

1.	I agree to being video recorded.	Yes	No
2.	I have read and understood the participant information for this study.	Yes	No
3.	I allow my data to be used in future ethically approved research.	Yes	No
4.	I agree to take part in this study.	Yes	No

Name of person giving consent	Date	Signature
	dd/mm/yyyy	
Name of person giving consent	**Date**	**Signature**
	dd/mm/yyyy	

Stakeholder interview questions

25. Your name and role?
26. Your organisation and its role in the research landscape (e.g., national/regional, disciplinary focus, services provided, typical users)?
27. What have been some of the successes in the context that your organisation works in, in building the UK's computational skills base and software development community in the last few years?
28. What does your organisation/community need to do in the next five years to prepare for the next generation of digital research infrastructure?
29. What is needed next to aid researchers in their journey through large scale computing?
30. What would you like your workforce to look like in five years?
31. What would you like to see the UKRI do to facilitate good research practice/innovation around research software?
32. What software skills and development training will be needed and how could this be funded?
33. How could support for software use in research be appropriately funded?
34. What level/type of coordination would be beneficial to enable a national strategy for research software?

APPENDIX E: SURVEY RESPONSES - UNIVARIATE ANALYSES

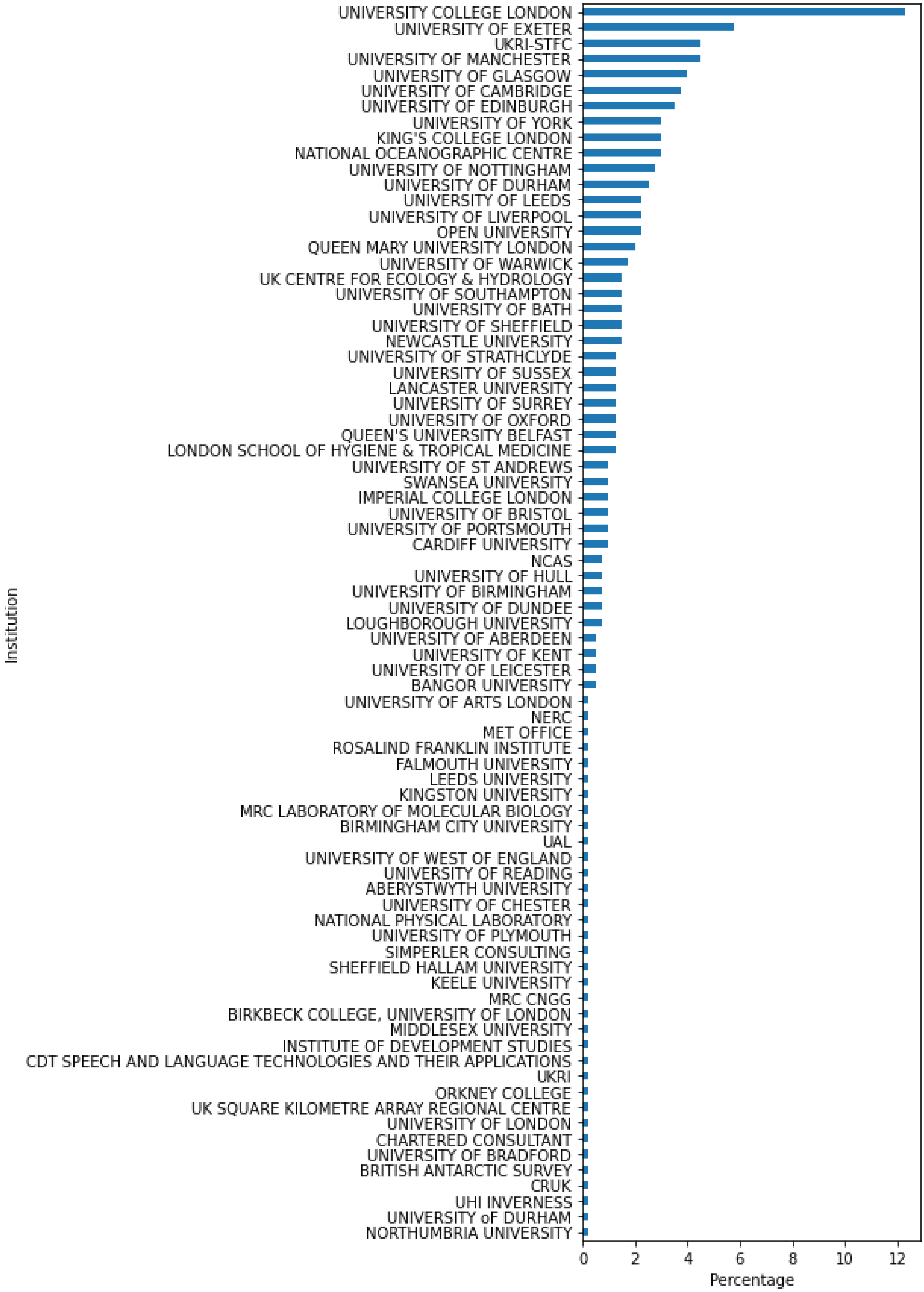
Q1. Consent

Q2. Which organisation do you work for? (N = 399)

Institution	N	Percent
UNIVERSITY COLLEGE LONDON	49	12.28
UNIVERSITY OF EXETER	23	5.76
UKRI-STFC	18	4.51
UNIVERSITY OF MANCHESTER	18	4.51
UNIVERSITY OF GLASGOW	16	4.01
UNIVERSITY OF CAMBRIDGE	15	3.76
UNIVERSITY OF EDINBURGH	14	3.51
KING'S COLLEGE LONDON	12	3.01
UNIVERSITY OF YORK	12	3.01
NATIONAL OCEANOGRAPHIC CENTRE	12	3.01
UNIVERSITY OF NOTTINGHAM	11	2.76
UNIVERSITY OF DURHAM	10	2.51
UNIVERSITY OF LIVERPOOL	9	2.26
OPEN UNIVERSITY	9	2.26
UNIVERSITY OF LEEDS	9	2.26
QUEEN MARY UNIVERSITY LONDON	8	2.01
UNIVERSITY OF WARWICK	7	1.75
NEWCASTLE UNIVERSITY	6	1.5
UNIVERSITY OF SOUTHAMPTON	6	1.5
UNIVERSITY OF SHEFFIELD	6	1.5
UNIVERSITY OF BATH	6	1.5
UK CENTRE FOR ECOLOGY and HYDROLOGY	6	1.5
QUEEN'S UNIVERSITY BELFAST	5	1.25
UNIVERSITY OF SURREY	5	1.25
UNIVERSITY OF OXFORD	5	1.25
UNIVERSITY OF SUSSEX	5	1.25
UNIVERSITY OF STRATHCLYDE	5	1.25

Institution	N	Percent
LANCASTER UNIVERSITY	5	1.25
LONDON SCHOOL OF HYGIENE and TROPICAL MEDICINE	5	1.25
IMPERIAL COLLEGE LONDON	4	1
UNIVERSITY OF BRISTOL	4	1
UNIVERSITY OF PORTSMOUTH	4	1
SWANSEA UNIVERSITY	4	1
CARDIFF UNIVERSITY	4	1
UNIVERSITY OF ST ANDREWS	4	1
UNIVERSITY OF HULL	3	0.75
UNIVERSITY OF BIRMINGHAM	3	0.75
UNIVERSITY OF DUNDEE	3	0.75
NCAS	3	0.75
LOUGHBOROUGH UNIVERSITY	3	0.75
BANGOR UNIVERSITY	2	0.5
UNIVERSITY OF ABERDEEN	2	0.5
UNIVERSITY OF KENT	2	0.5
UNIVERSITY OF LEICESTER	2	0.5
CDT SPEECH AND LANGUAGE TECHNOLOGIES AND THEIR APPLICATIONS	1	0.25
BIRMINGHAM CITY UNIVERSITY	1	0.25
SIMPERLER CONSULTING	1	0.25
UNIVERSITY OF BRADFORD	1	0.25
UK SQUARE KILOMETRE ARRAY REGIONAL CENTRE	1	0.25
MRC LABORATORY OF MOLECULAR BIOLOGY	1	0.25
ROSALIND FRANKLIN INSTITUTE	1	0.25
CHARTERED CONSULTANT	1	0.25
LEEDS UNIVERSITY	1	0.25
KINGSTON UNIVERSITY	1	0.25
UKRI	1	0.25

Institution	N	Percent
ORKNEY COLLEGE	1	0.25
UHI INVERNESS	1	0.25
BRITISH ANTARCTIC SURVEY	1	0.25
NERC	1	0.25
CRUK	1	0.25
UNIVERSITY oF DURHAM	1	0.25
MET OFFICE	1	0.25
MIDDLESEX UNIVERSITY	1	0.25
NATIONAL PHYSICAL LABORATORY	1	0.25
ABERYSTWYTH UNIVERSITY	1	0.25
UNIVERSITY OF PLYMOUTH	1	0.25
UNIVERSITY OF WEST OF ENGLAND	1	0.25
SHEFFIELD HALLAM UNIVERSITY	1	0.25
INSTITUTE OF DEVELOPMENT STUDIES	1	0.25
UNIVERSITY OF ARTS LONDON	1	0.25
UNIVERSITY OF READING	1	0.25
KEELE UNIVERSITY	1	0.25
NORTHUMBRIA UNIVERSITY	1	0.25
FALMOUTH UNIVERSITY	1	0.25
UAL	1	0.25
UNIVERSITY OF CHESTER	1	0.25
BIRKBECK COLLEGE UNIVERSITY OF LONDON	1	0.25
MRC CNGG	1	0.25
UNIVERSITY OF LONDON	1	0.25

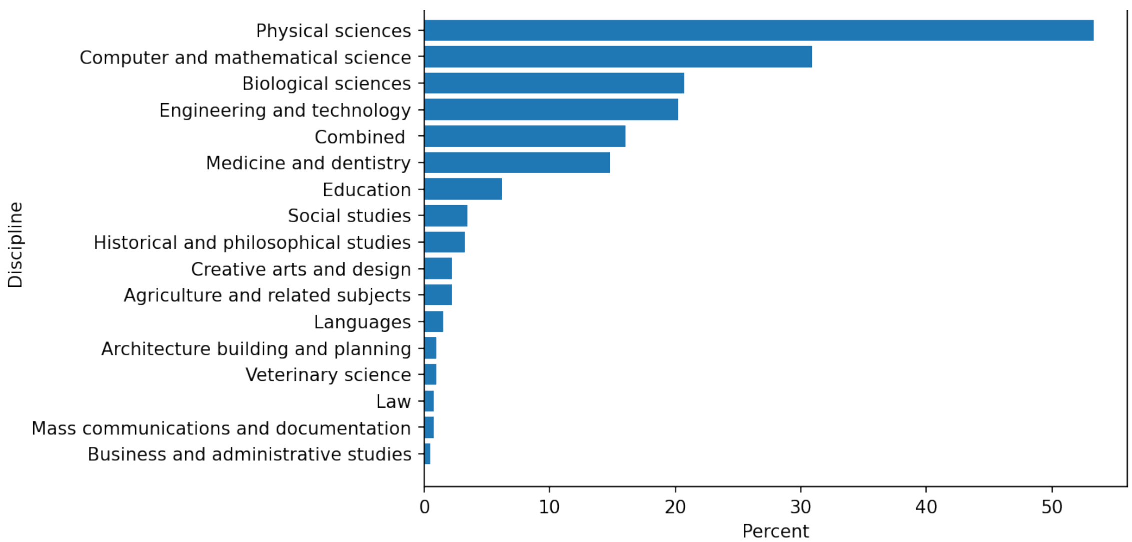

UNIVERSITY COLLEGE LONDON
UNIVERSITY OF EXETER
UKRI-STFC
UNIVERSITY OF MANCHESTER
UNIVERSITY OF GLASGOW
UNIVERSITY OF CAMBRIDGE
UNIVERSITY OF EDINBURGH
UNIVERSITY OF YORK
KING'S COLLEGE LONDON
NATIONAL OCEANOGRAPHIC CENTRE
UNIVERSITY OF NOTTINGHAM
UNIVERSITY OF DURHAM
UNIVERSITY OF LEEDS
UNIVERSITY OF LIVERPOOL
OPEN UNIVERSITY
QUEEN MARY UNIVERSITY LONDON
UNIVERSITY OF WARWICK
UK CENTRE FOR ECOLOGY & HYDROLOGY
UNIVERSITY OF SOUTHAMPTON
UNIVERSITY OF BATH
UNIVERSITY OF SHEFFIELD
NEWCASTLE UNIVERSITY
UNIVERSITY OF STRATHCLYDE
UNIVERSITY OF SUSSEX
LANCASTER UNIVERSITY
UNIVERSITY OF SURREY
UNIVERSITY OF OXFORD
QUEEN'S UNIVERSITY BELFAST
LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE
UNIVERSITY OF ST ANDREWS
SWANSEA UNIVERSITY
IMPERIAL COLLEGE LONDON
UNIVERSITY OF BRISTOL
UNIVERSITY OF PORTSMOUTH
CARDIFF UNIVERSITY
NCAS
UNIVERSITY OF HULL
UNIVERSITY OF BIRMINGHAM
UNIVERSITY OF DUNDEE
LOUGHBOROUGH UNIVERSITY
UNIVERSITY OF ABERDEEN
UNIVERSITY OF KENT
UNIVERSITY OF LEICESTER
BANGOR UNIVERSITY
UNIVERSITY OF ARTS LONDON
NERC
MET OFFICE
ROSALIND FRANKLIN INSTITUTE
FALMOUTH UNIVERSITY
LEEDS UNIVERSITY
KINGSTON UNIVERSITY
MRC LABORATORY OF MOLECULAR BIOLOGY
BIRMINGHAM CITY UNIVERSITY
UAL
UNIVERSITY OF WEST OF ENGLAND
UNIVERSITY OF READING
ABERYSTWYTH UNIVERSITY
UNIVERSITY OF CHESTER
NATIONAL PHYSICAL LABORATORY
UNIVERSITY OF PLYMOUTH
SIMPERLER CONSULTING
SHEFFIELD HALLAM UNIVERSITY
KEELE UNIVERSITY
MRC CNGG
BIRKBECK COLLEGE, UNIVERSITY OF LONDON
MIDDLESEX UNIVERSITY
INSTITUTE OF DEVELOPMENT STUDIES
CDT SPEECH AND LANGUAGE TECHNOLOGIES AND THEIR APPLICATIONS
UKRI
ORKNEY COLLEGE
UK SQUARE KILOMETRE ARRAY REGIONAL CENTRE
UNIVERSITY OF LONDON
CHARTERED CONSULTANT
UNIVERSITY OF BRADFORD
BRITISH ANTARCTIC SURVEY
CRUK
UHI INVERNESS
UNIVERSITY oF DURHAM
NORTHUMBRIA UNIVERSITY
Institution
0
2
4
6
8
10
12
Percentage

Q3. Are you based in the United Kingdom?

All of the graphs and tables of survey data contain only participants who indicated that they were based in the UK (N = 405).

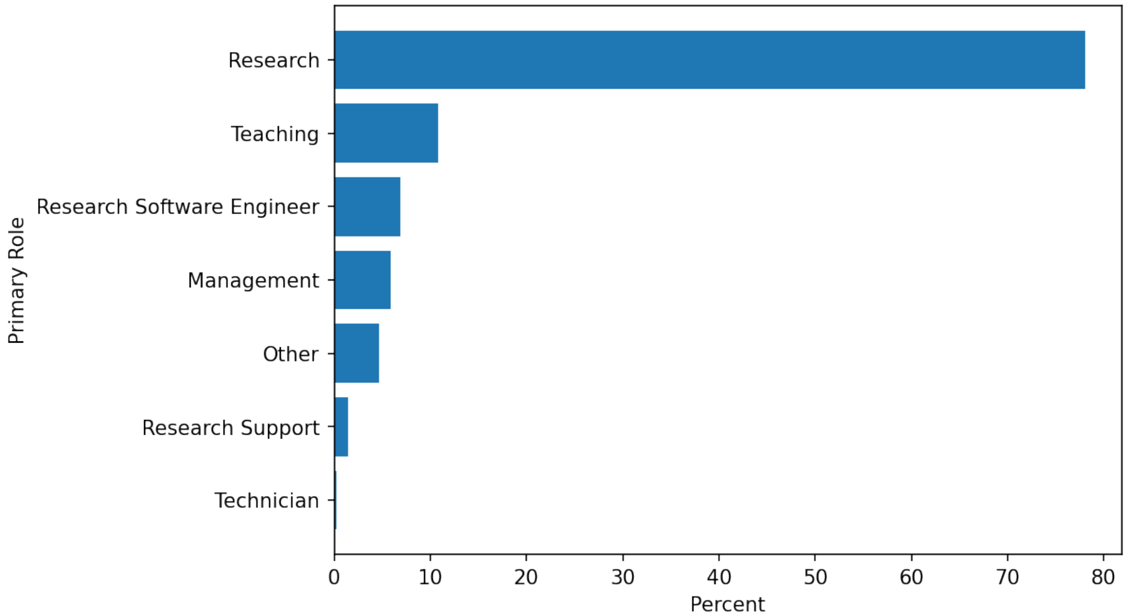
Q4. In which disciplines do you work? Select all that apply (N = 405)

Discipline	N	Percent
Physical sciences	216	53.33
Computer and mathematical science	125	30.86
Biological sciences	84	20.74
Engineering and technology	82	20.25
Combined	65	16.05
Medicine and dentistry	60	14.81
Education	25	6.17
Social studies	14	3.46
Historical and philosophical studies	13	3.21
Creative arts and design	9	2.22
Agriculture and related subjects	9	2.22
Languages	6	1.48
Architecture building and planning	4	0.99
Veterinary science	4	0.99
Law	3	0.74
Mass communications and documentation	3	0.74
Business and administrative studies	2	0.49



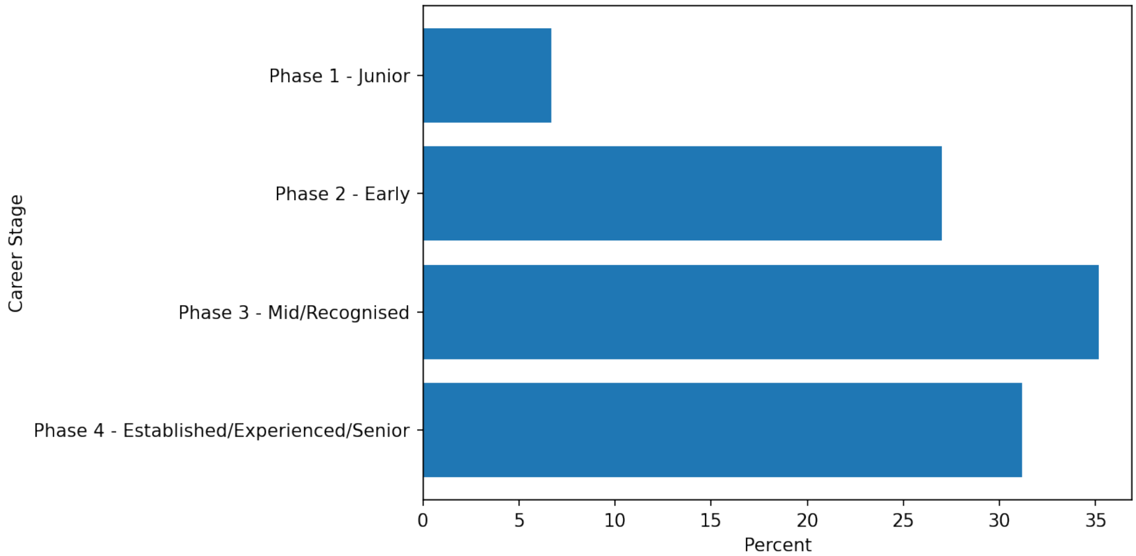
Q5. What is currently your primary role? Select only one (N = 405)

Primary Role	N	Percent
Research	316	78.02
Teaching	44	10.86
Research Software Engineer	28	6.91
Management	24	5.93
Other	19	4.69
Research Support	6	1.48
Technician	1	0.25



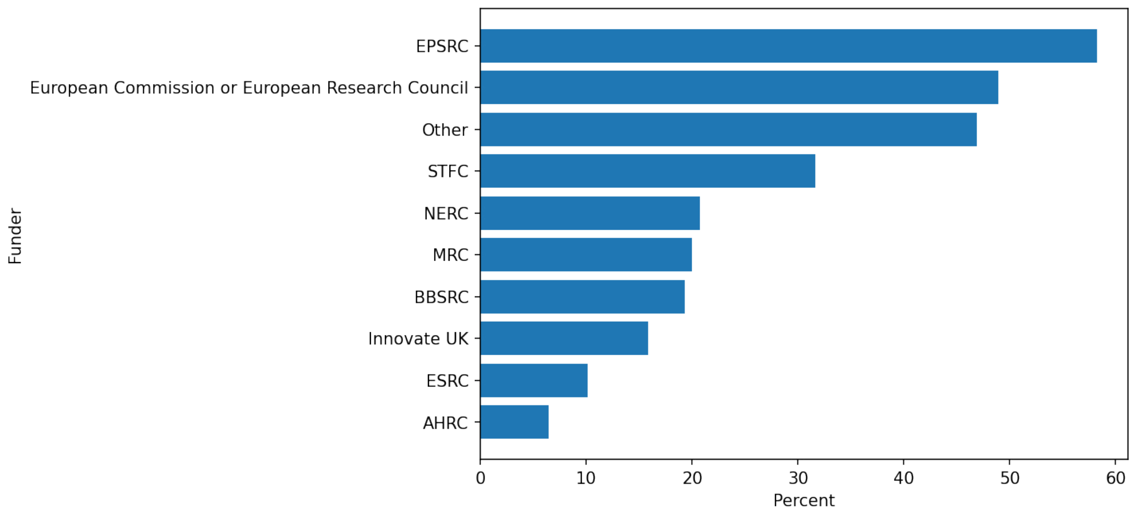
Q6. At what career stage are you currently? Select the one career stage that most accurately describes your role. (N = 404)

Career Stage	N	Percent
Phase 1 - Junior	27	6.68
Phase 2 - Early	109	26.98
Phase 3 - Mid/Recognised	142	35.15
Phase 4 - Established/Experienced/Senior	126	31.19



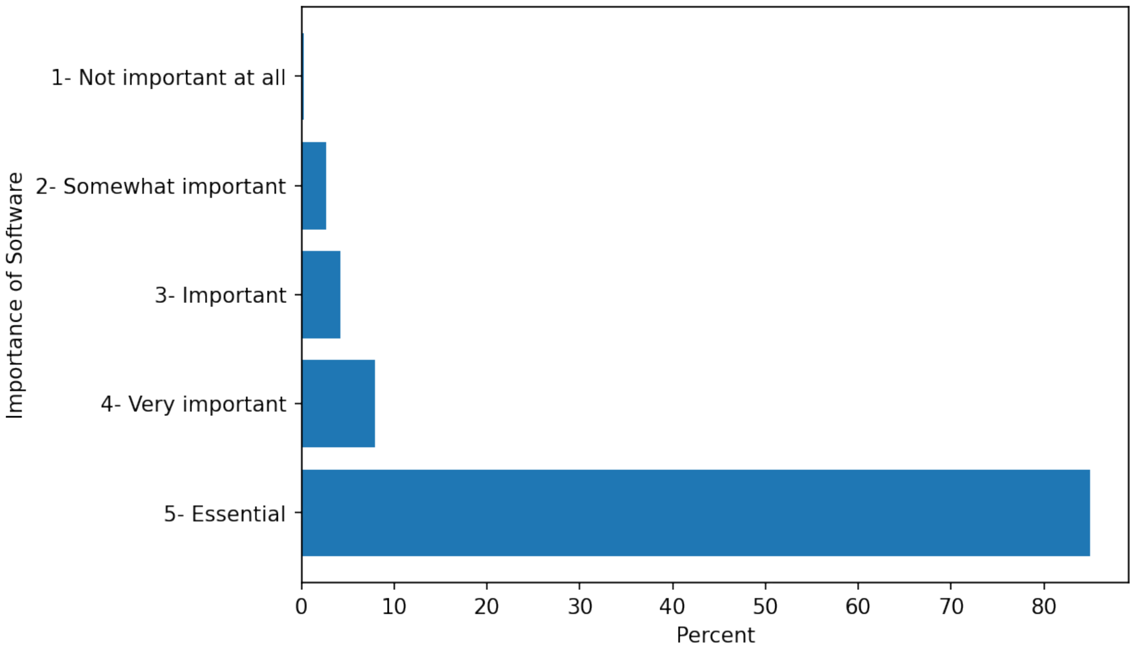
Q7. Where have you applied for funding, or who funds your work? (N = 405)

Funder	N	Percent
EPSRC	236	58.27
European Commission or European Research Council	198	48.89
Other	190	46.91
STFC	128	31.6
NERC	84	20.74
MRC	81	20
BBSRC	78	19.26
Innovate UK	64	15.8
ESRC	41	10.12
AHRC	26	6.42



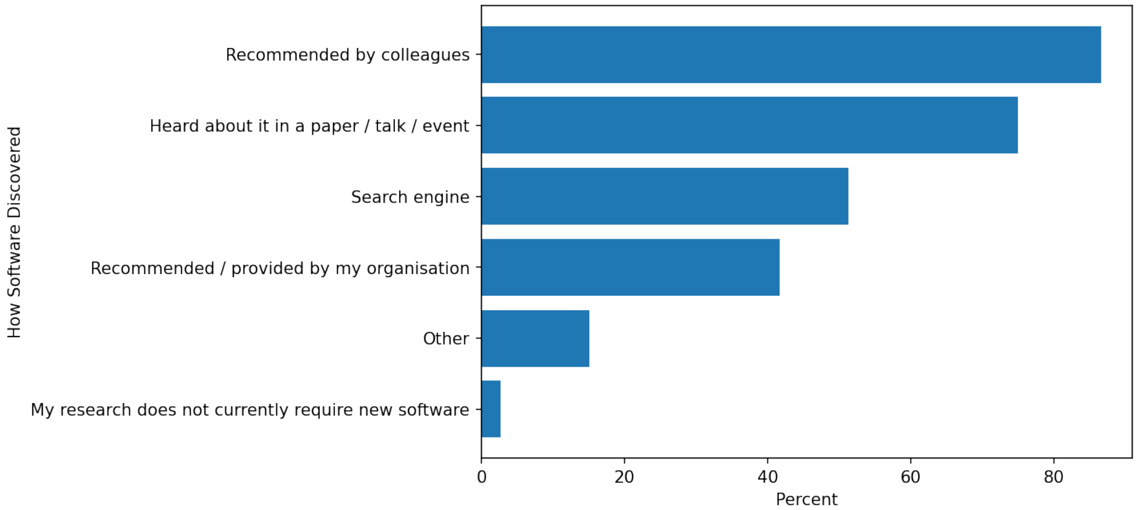
Q8. How important is software to your research? (N = 405)

Importance of Software	N	Percent
1 - Not important at all	1	0.25
2 - Somewhat important	11	2.72
3 - Important	17	4.2
4 - Very important	32	7.9
5 - Essential	344	84.94



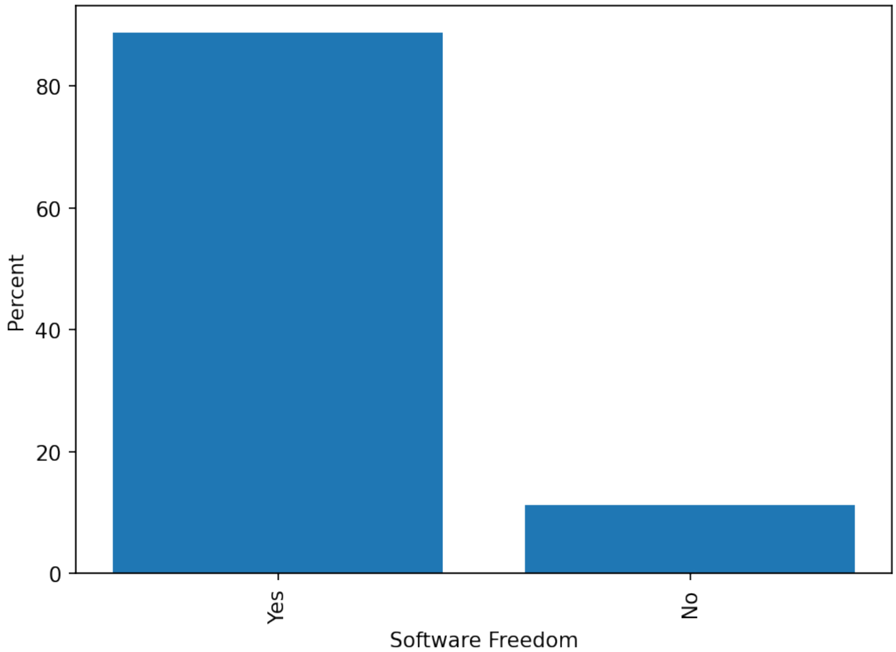
Q9. How do you discover the software you use in your research? Select all that apply.

How Software Discovered	N	Percent
Recommended by colleagues	351	86.67
Heard about it in a paper / talk / event	304	75.06
Search engine	208	51.36
Recommended / provided by my organisation	169	41.73
Other	61	15.06
My research does not currently require new software	11	2.72



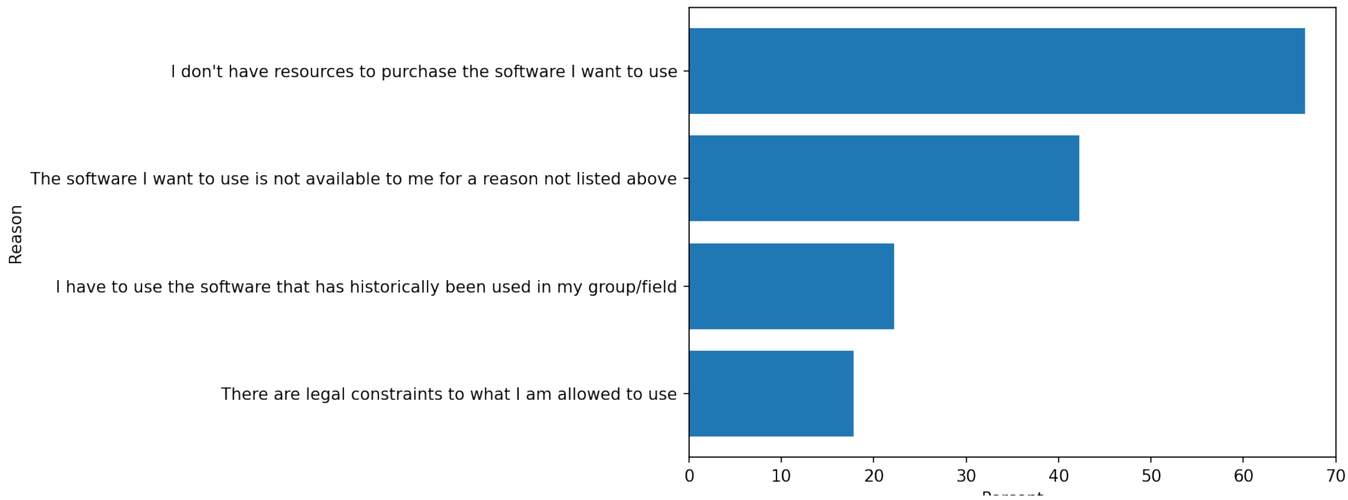
Q10. Are you free to use the software you want to use? (N = 400)

Software Freedom	N	Percent
Yes	355	88.75
No	45	11.25



Q10a. If you answered No in the above question what is the reason? Select all that apply (N = 45).

Reason	N	Percent
Because I don't have resources to purchase the software I want to use	30	66.67
Because the software I want to use is not available to me for a reason not listed above	19	42.22
Because I have to use the software that has historically been used in my group/field	10	22.22
Because there are legal constraints to what I am allowed to use	8	17.78

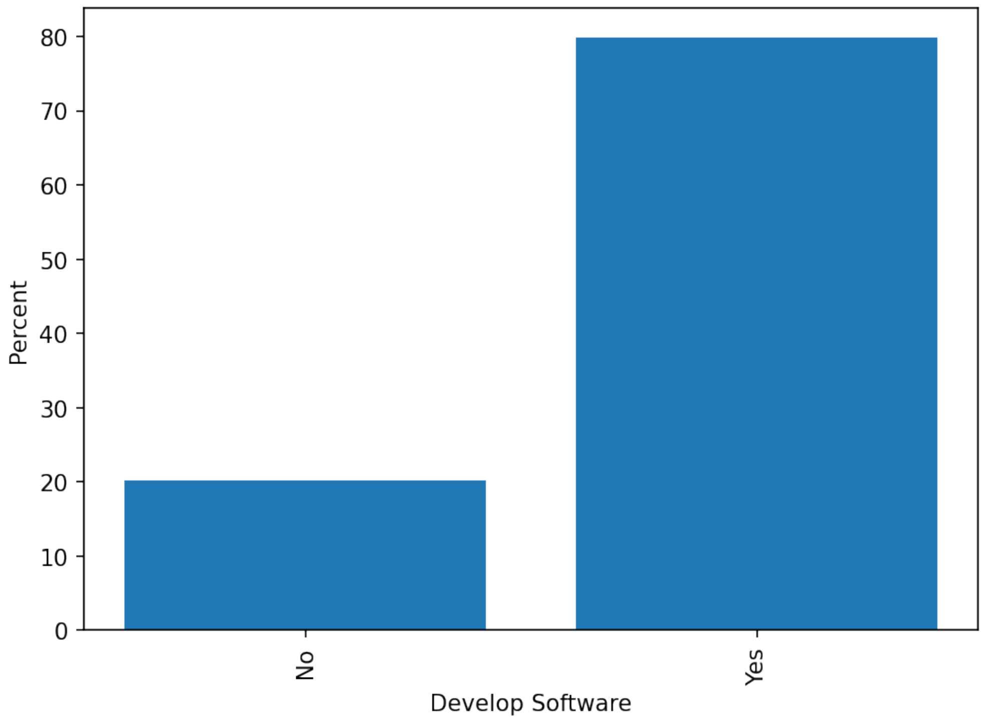


Q11. What are the names of (up to three) pieces of research software you've used to support your research in the last month? We are interested in the software that you believe is most important to your work. Please separate your answers using semicolon.

Software	N	Percent
python	76	19
matlab	46	11.5
r	39	9.75
latex	11	2.75
mathematica	11	2.75
stata	10	2.5
git	9	2.25
pytorch	9	2.25
amber	8	2
vasp	8	2
overleaf	7	1.75
imagej	7	1.75
gaussian	7	1.75
fiji	7	1.75
paraview	7	1.75
excel	7	1.75
github	6	1.5
docker	6	1.5
numpy	6	1.5
fortran	6	1.5

Q12. Do you develop software (i.e. write your own code) for your research? (N = 402)

Develop Software	N	Percent
No	81	20.15
Yes	321	79.85

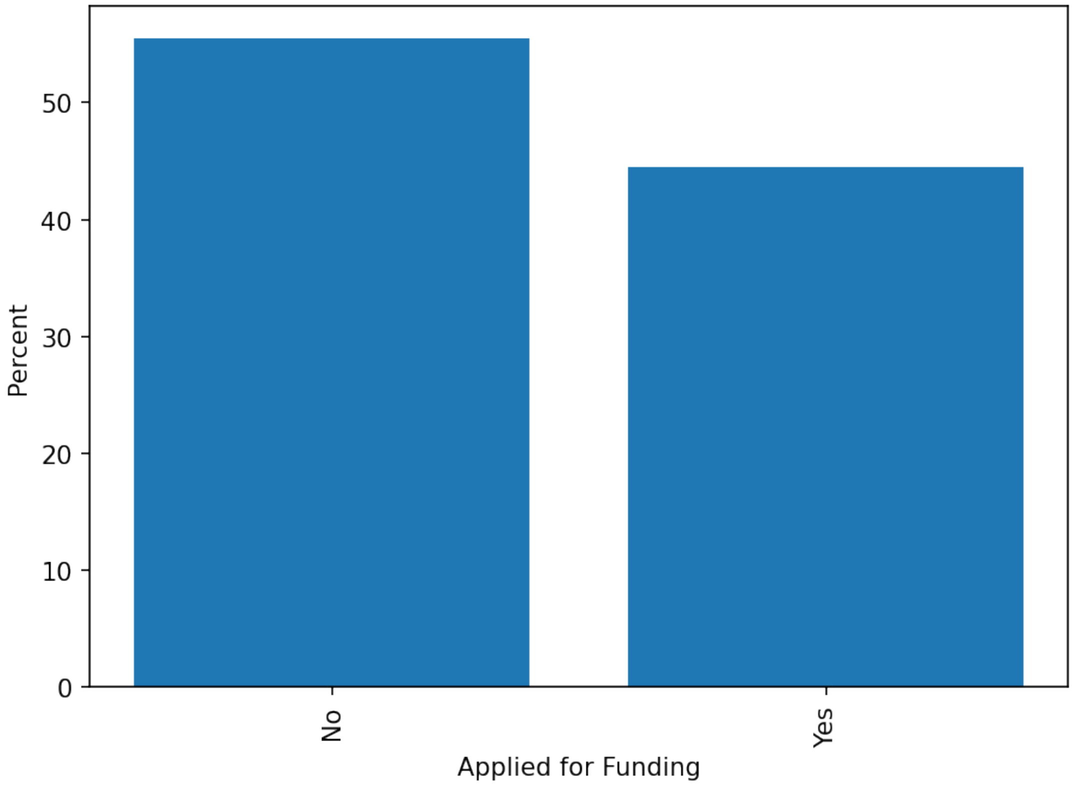


Q12a. What programming language(s) do you usually use for developing code? Please separate multiple answers by semicolon.

Software	N	Percent
python	235	59.19
fortran	98	24.69
c++	92	23.17
c	65	16.37
matlab	57	14.36
r	52	13.1
bash	28	7.05
java	26	6.55
perl	10	2.52
idl	8	2.02
javascript	8	2.02
rust	7	1.76
cuda	5	1.26
julia	5	1.26
c#	4	1.01
php	3	0.76
mathematica	3	0.76
fortran 90	3	0.76
shell	3	0.76
unix	3	0.76

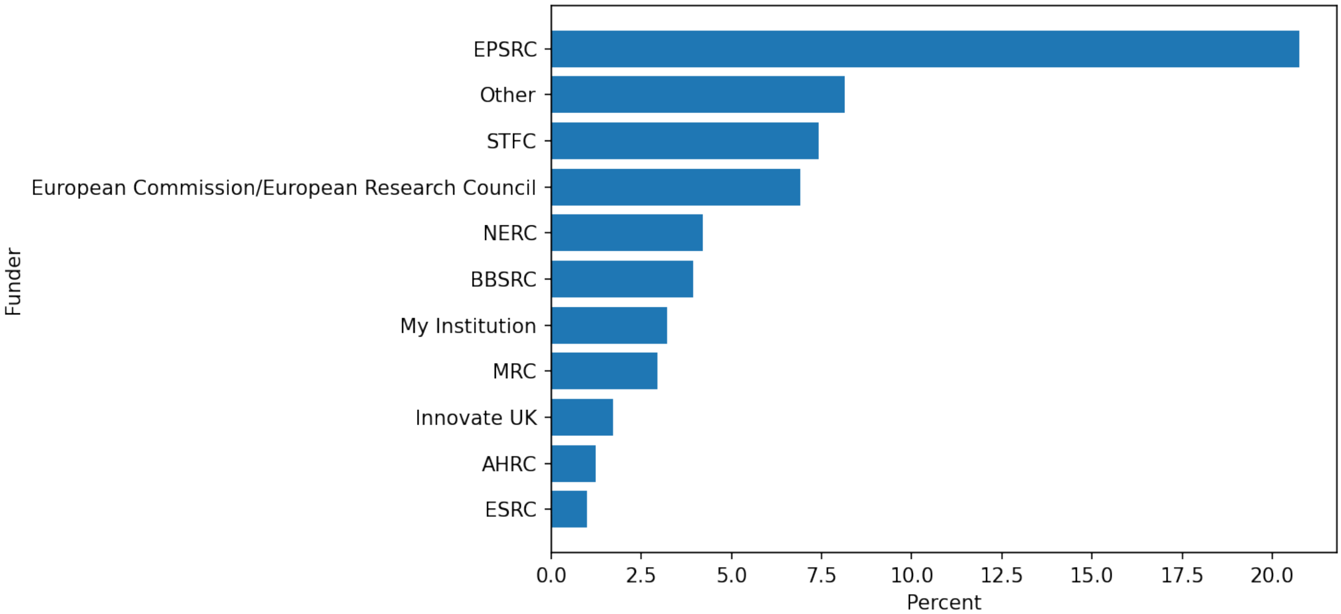
Q12b. Have you ever applied for funding that explicitly includes costs for software development? (N = 319)

Applied for Funding	N	Percent
No	177	55.49
Yes	142	44.51



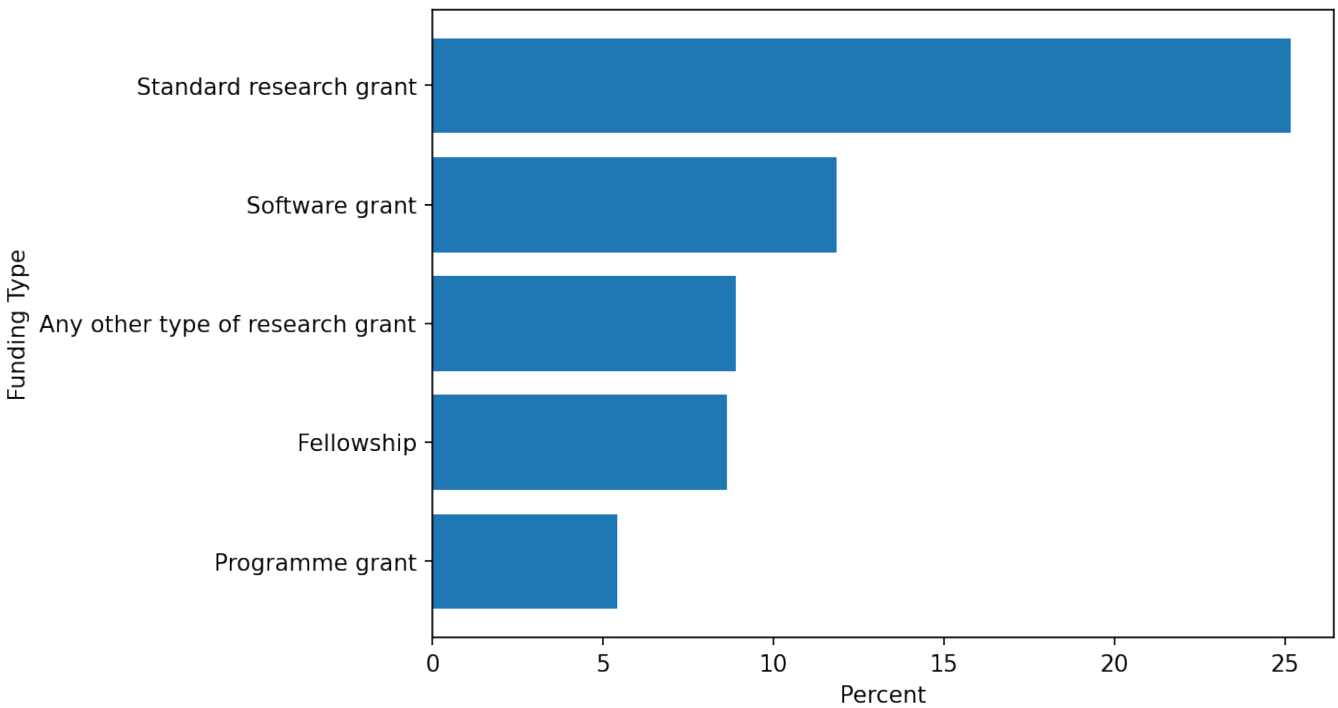
Q12b.i. Who was the funder? Select all that apply.(N = 142 - including only those who answered yes to the previous question)

Funder	N	Percent
EPSRC	84	59.15
Other	33	23.24
STFC	30	21.13
European Commission/European Research Council	28	19.72
NERC	17	11.97
BBSRC	16	11.27
My Institution	13	9.15
MRC	12	8.45
Innovate UK	7	4.93
AHRC	5	3.52
ESRC	4	2.82



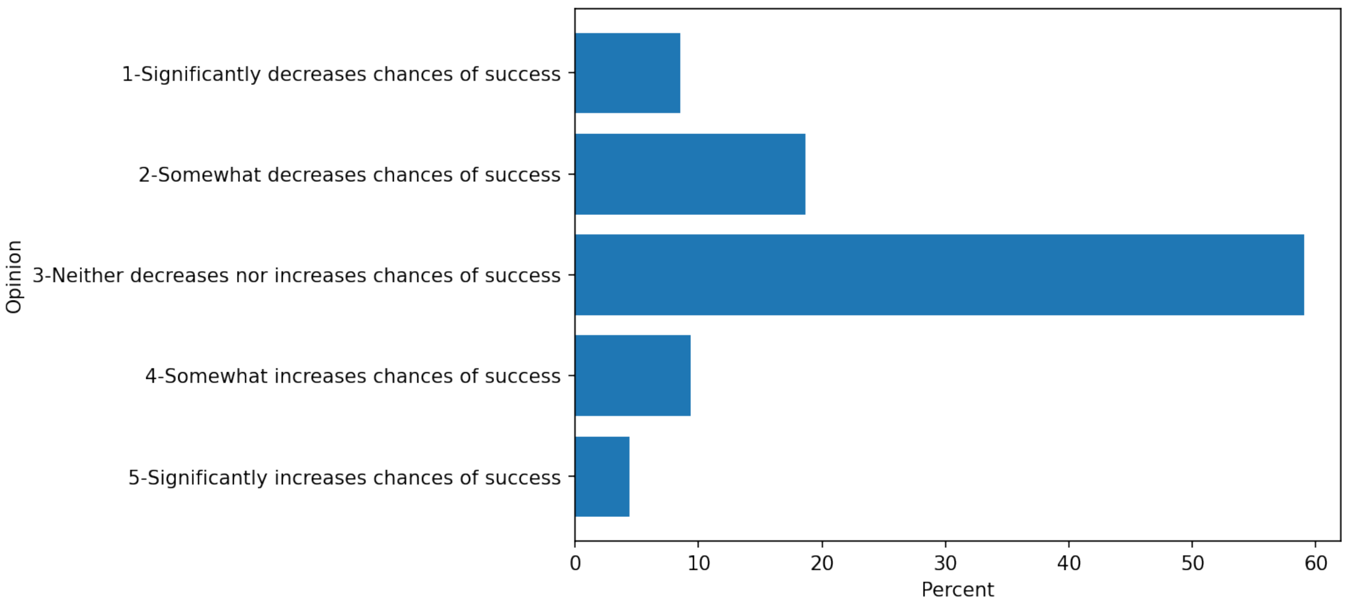
Q12b.ii. What type of funding did you apply for? Select all that apply.

Funding Type	N	Percent
Standard research grant	102	71.83
Software grant	48	33.8
Any other type of research grant	36	25.35
Fellowship	35	24.65
Programme grant	22	15.49



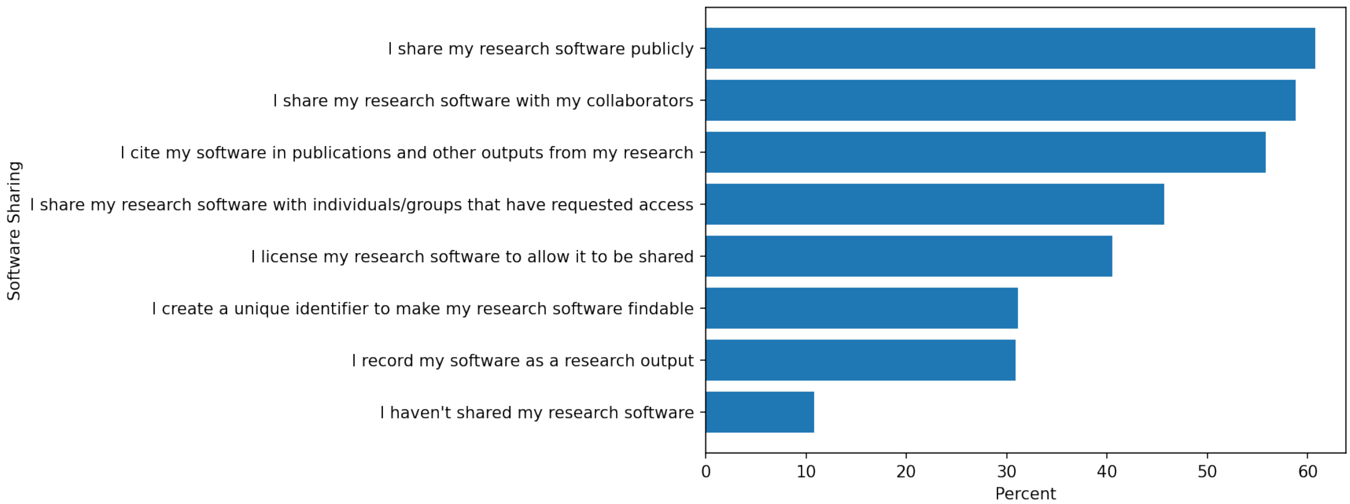
Q13. What is your opinion on including software development costs in a funding proposal? (N = 386)

Opinion	N	Percent
1 - Significantly decreases chances of success	33	8.55
2 - Somewhat decreases chances of success	72	18.65
3 - Neither decreases nor increases chances of success	228	59.07
4 - Somewhat increases chances of success	36	9.33
5 - Significantly increases chances of success	17	4.4



Q14. Over the last five years, which of the following practices have been part of your standard research process? Select all that apply. (N = 400)

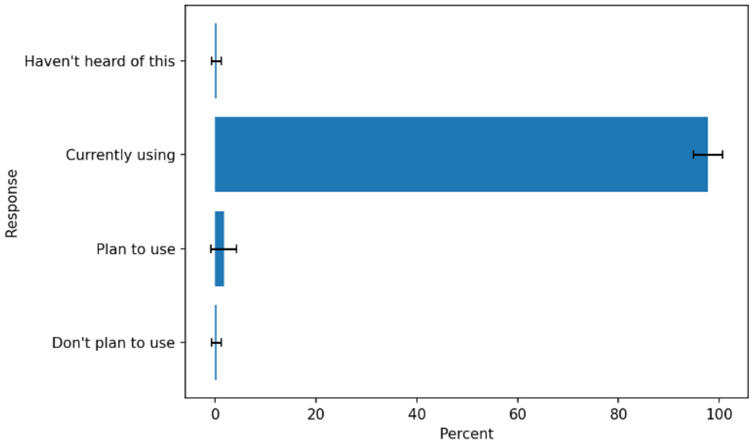
Software Sharing	N	Percent
I share my research software publicly	246	60.74
I share my research software with my collaborators	238	58.77
I cite my software in publications and other outputs from my research	226	55.8
I share my research software with individuals/groups that have requested access	185	45.68
I license my research software to allow it to be shared	164	40.49
I create a unique identifier to make my research software findable	126	31.11
I record my software as a research output	125	30.86
I haven't shared my research software	44	10.86



Q15. Which of the following do you currently use or plan to use in the next five years for the computational parts of your research? (N = 405)

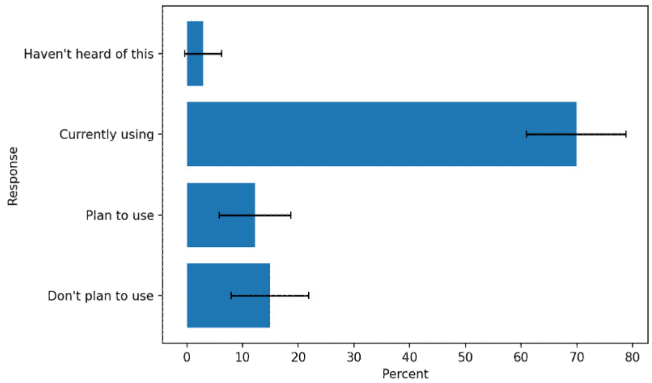
Q15.1. A laptop/desktop

	N	Percent
Haven't heard of this	1	0.25
Currently using	396	97.77
Plan to use	7	1.74
Don't plan to use	1	0.25



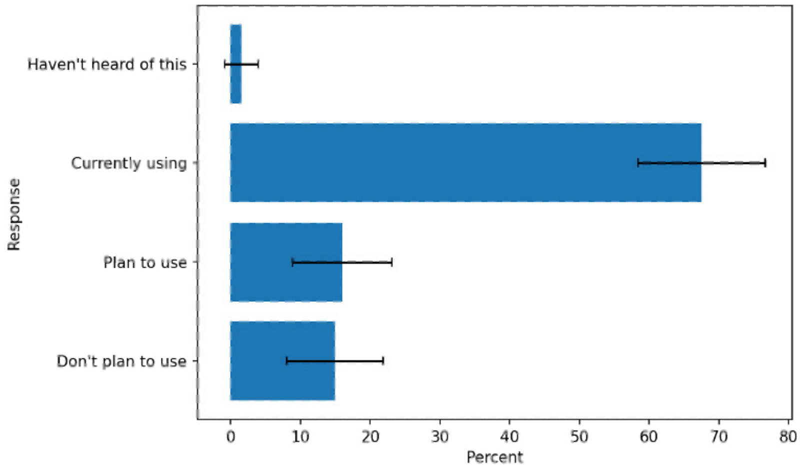
Q15.2. A server/cluster operated by your research group/department

	N	Percent
Haven't heard of this	12	2.93
Currently using	283	69.87
Plan to use	50	12.27
Don't plan to use	60	14.93



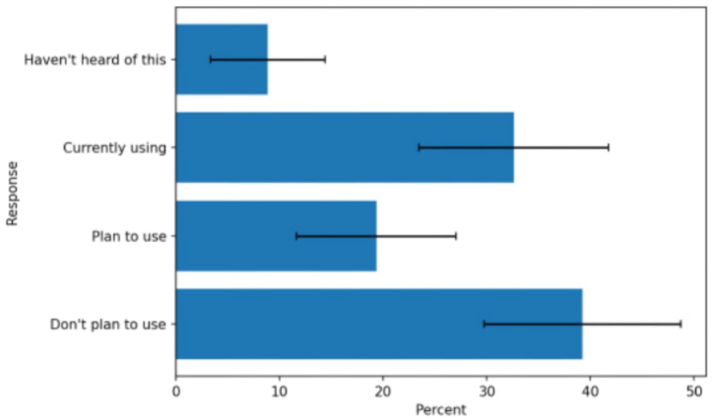
Q15.3. A central research computing service operated by your institution/organisation

	N	Percent
Haven't heard of this	6	1.55
Currently using	273	67.53
Plan to use	65	15.98
Don't plan to use	61	14.95



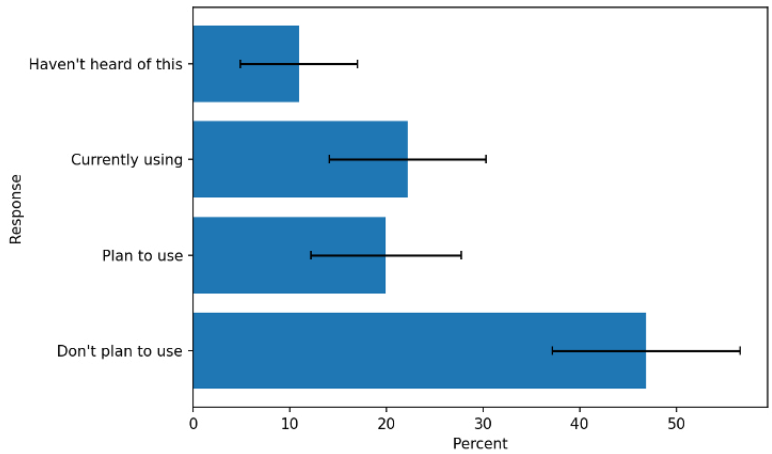
Q15.4. Computing services at a data centre or at a data safe haven (including trusted research environments)

Response	N	Percent
Haven't heard of this	36	8.84
Currently using	132	32.6
Plan to use	78	19.34
Don't plan to use	159	39.23



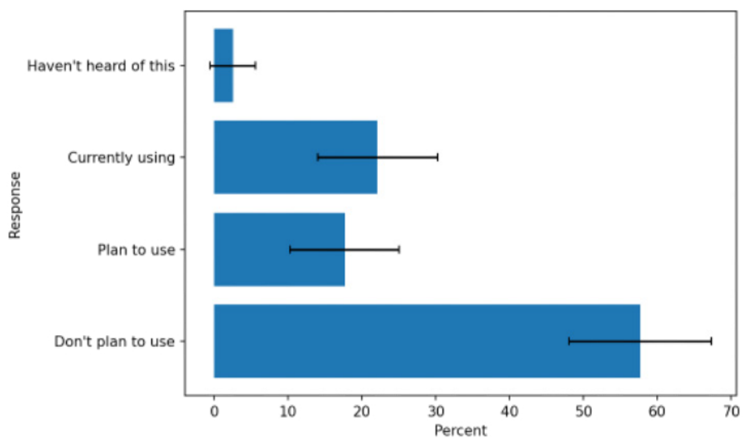
Q15.5. An Academic Cloud Computing/High-Throughput Computing service

Response	N	Percent
Haven't heard of this	44	10.96
Currently using	90	22.19
Plan to use	81	19.94
Don't plan to use	190	46.91



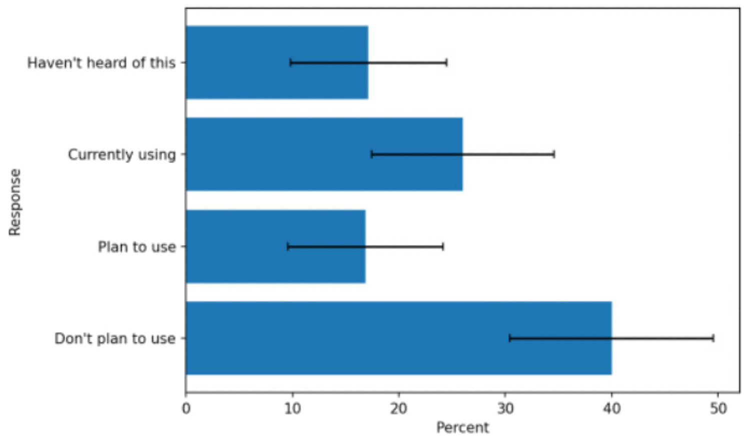
Q15.6. A Commercial Cloud provider, e.g., Microsoft Azure or Amazon Web Services

Response	N	Percent
Haven't heard of this	10	2.52
Currently using	90	22.13
Plan to use	71	17.65
Don't plan to use	234	57.7



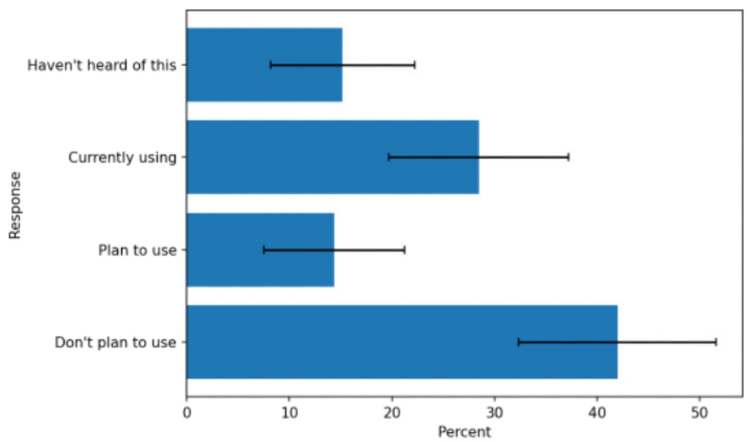
Q15.7. A UK Tier 2 high performance computing service

Response	N	Percent
Haven't heard of this	69	17.14
Currently using	105	26
Plan to use	68	16.86
Don't plan to use	162	40



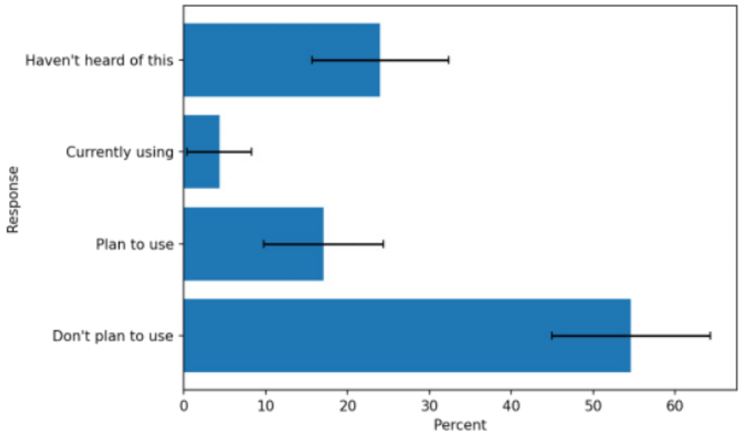
Q15.8. A UK Tier 1 (national) high performance computing service (e.g., ARCHER2)

Response	N	Percent
Haven't heard of this	62	15.19
Currently using	115	28.45
Plan to use	58	14.36
Don't plan to use	170	41.99



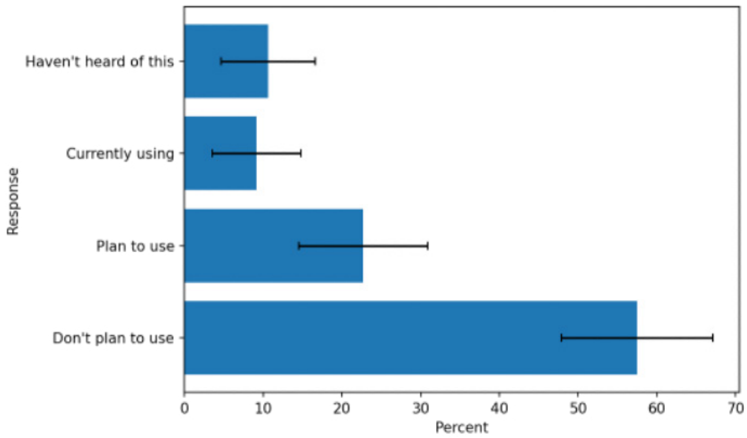
Q15.9. An Exascale HPC service

Response	N	Percent
Haven't heard of this	97	23.99
Currently using	18	4.34
Plan to use	69	17.05
Don't plan to use	221	54.62



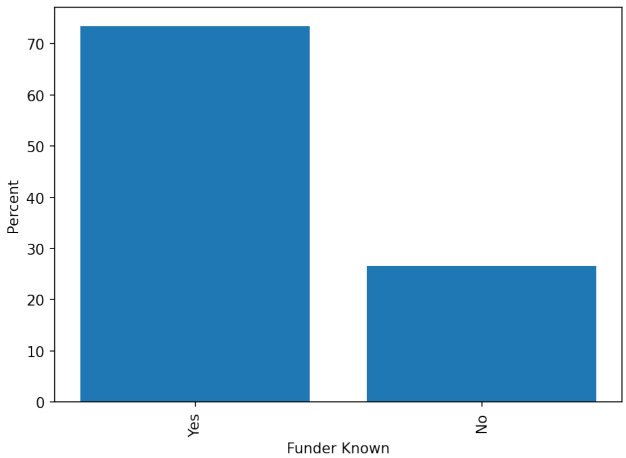
Q15.10. Specialised Machine Learning hardware

Response	N	Percent
Haven't heard of this	43	10.62
Currently using	37	9.14
Plan to use	92	22.71
Don't plan to use	233	57.52



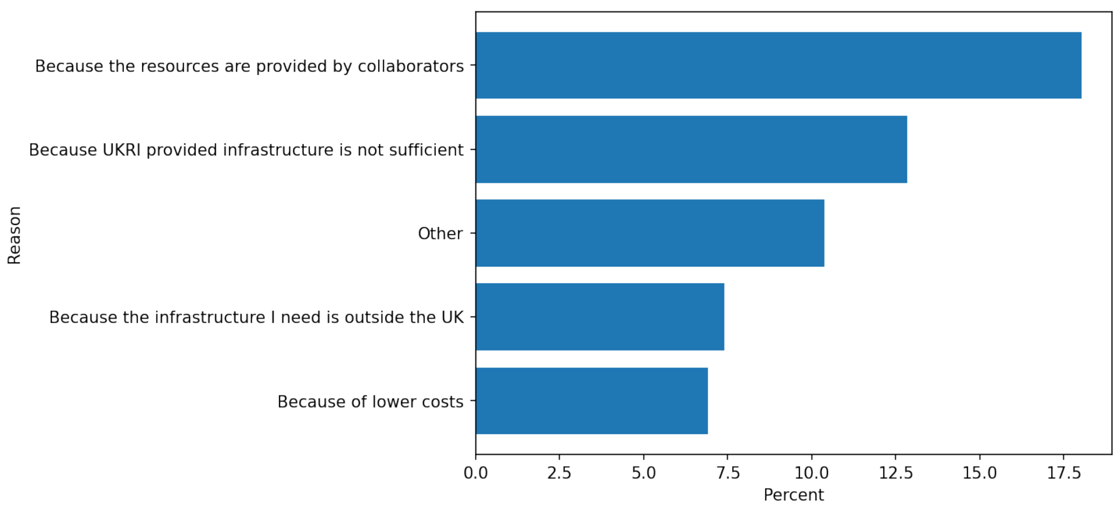
Q16. Do you know who funds the computational infrastructure you use for your research? (N = 403)

Funder Known	N	Percent
Yes	296	73.45
No	107	26.55



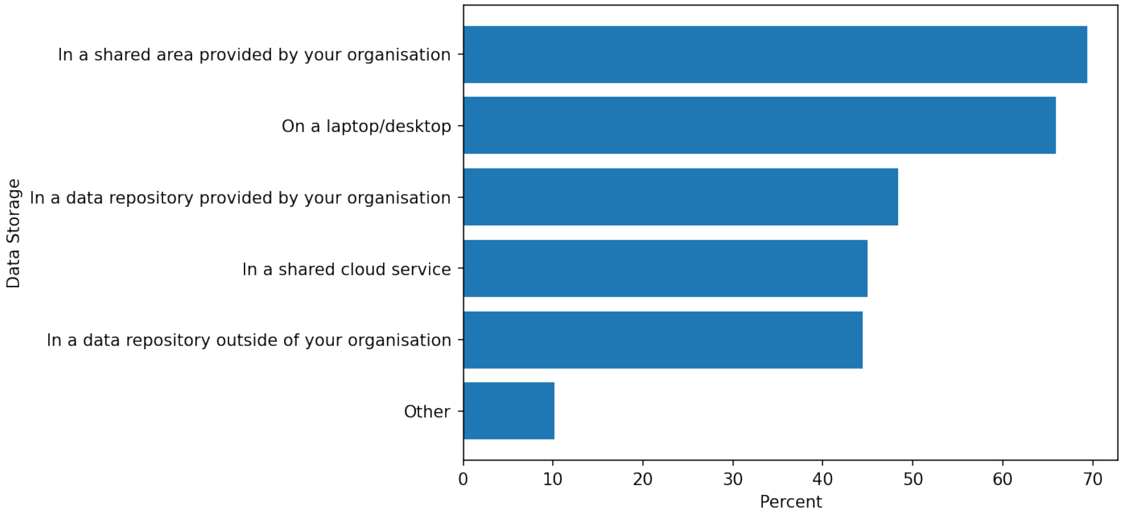
Q16.a. Do you use any non-UKRI funded computational infrastructure in preference to a UKRI one? If so, what are the main reasons? (N = 405)

Reason	N	Percent
Because the resources are provided by collaborators	73	18.02
Because UKRI provided infrastructure is not sufficient	52	12.84
Other	42	10.37
Because the infrastructure I need is outside the UK	30	7.41
Because of lower costs	28	6.91



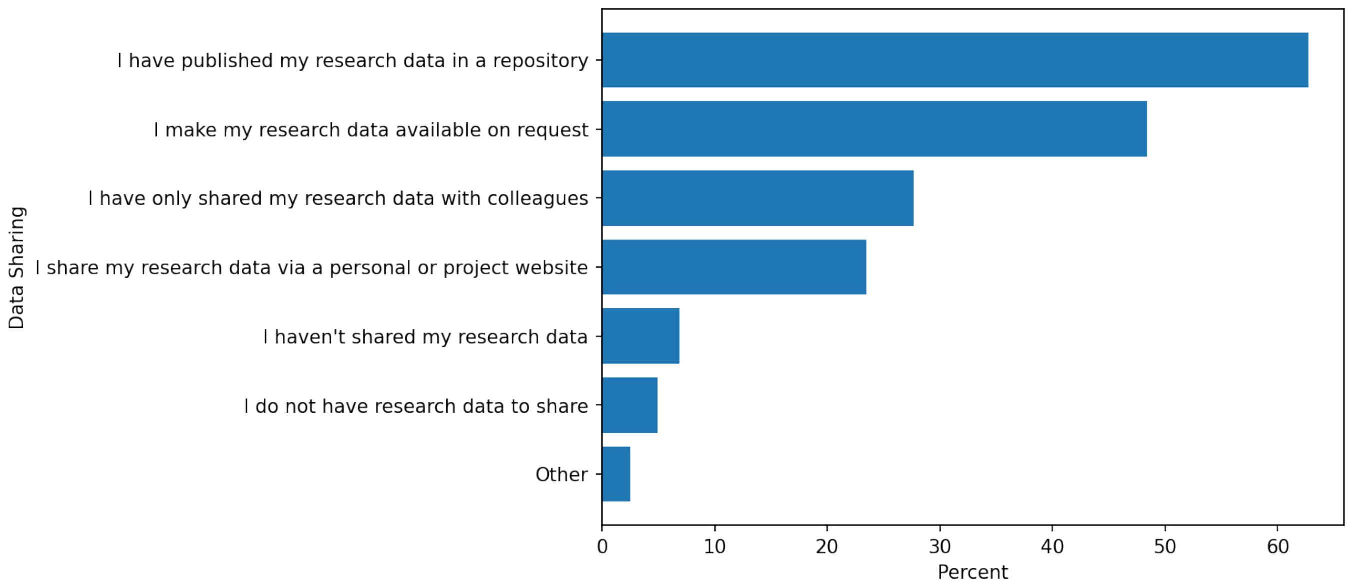
Q17. Over the last five years, once you've completed a research project/study, which of the following have you used to store the data that were generated after your research has been completed? Select all that apply. (N = 405)

Data Storage	N	Percent
In a shared area provided by your organisation	281	69.38
On a laptop/desktop	267	65.93
In a data repository provided by your organisation	196	48.4
In a shared cloud service	182	44.94
In a data repository outside of your organisation	180	44.44
Other	41	10.12



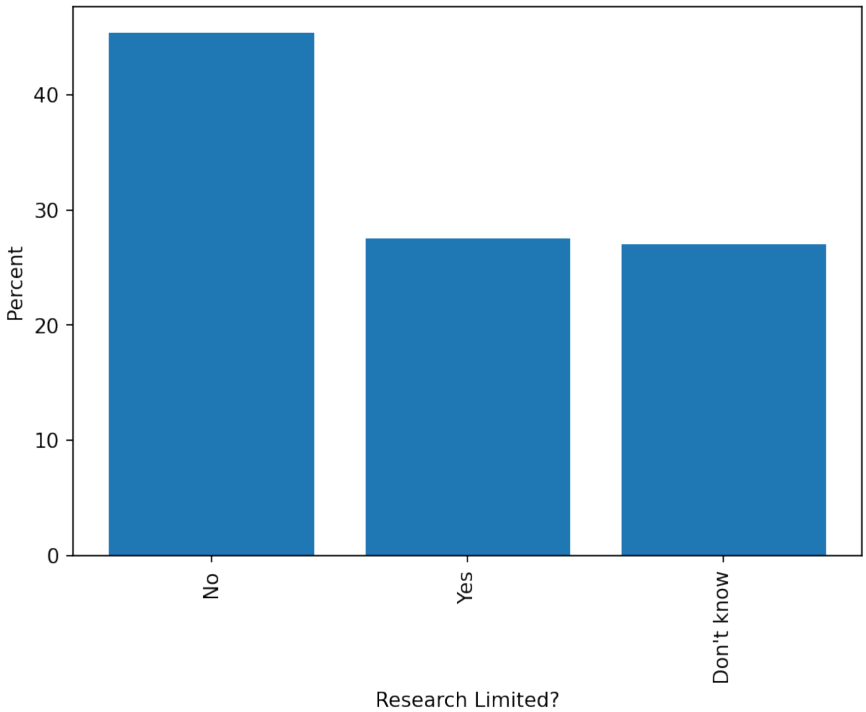
Q18. Over the last five years, which of the following statements best apply to how you share your research data? Select all that apply. (N = 405)

Data Sharing	N	Percent
I have published my research data in a repository	254	62.72
I make my research data available on request	196	48.4
I have only shared my research data with colleagues	112	27.65
I share my research data via a personal or project website	95	23.46
I haven't shared my research data	28	6.91
I do not have research data to share	20	4.94
Other	10	2.47



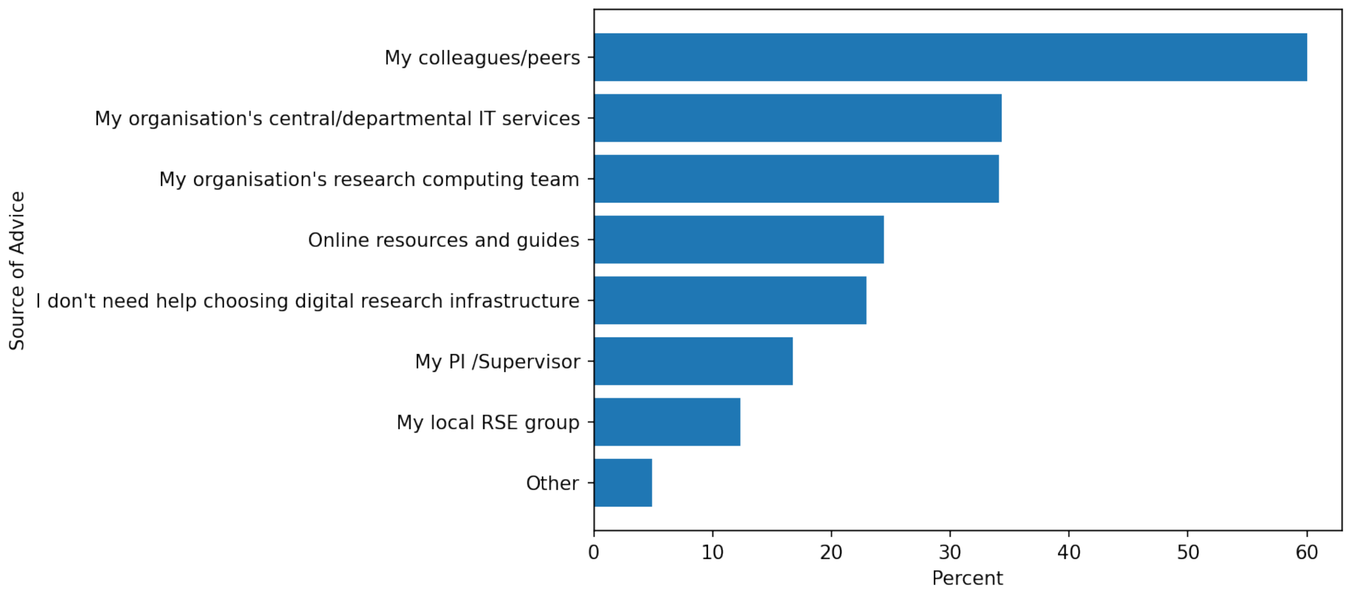
Q19. In the next year, do you feel that your research will be limited by access to computational and/or data research infrastructure? (N = 398)

Research Limited?	N	Percent
No	183	45.41
Yes	111	27.54
Don't know	109	27.05



Q20. Who do you ask for advice on choosing the computational and/or data research infrastructure (computing, storage and network) you use? (N = 405)

Source of Advice	N	Percent
My colleagues/peers	243	60
My organisation's central/departmental IT services	139	34.32
My organisation's research computing team	138	34.07
Online resources and guides	99	24.44
I don't need help choosing digital research infrastructure	93	22.96
My PI /Supervisor	68	16.79
My local RSE group	50	12.35
Other	20	4.94

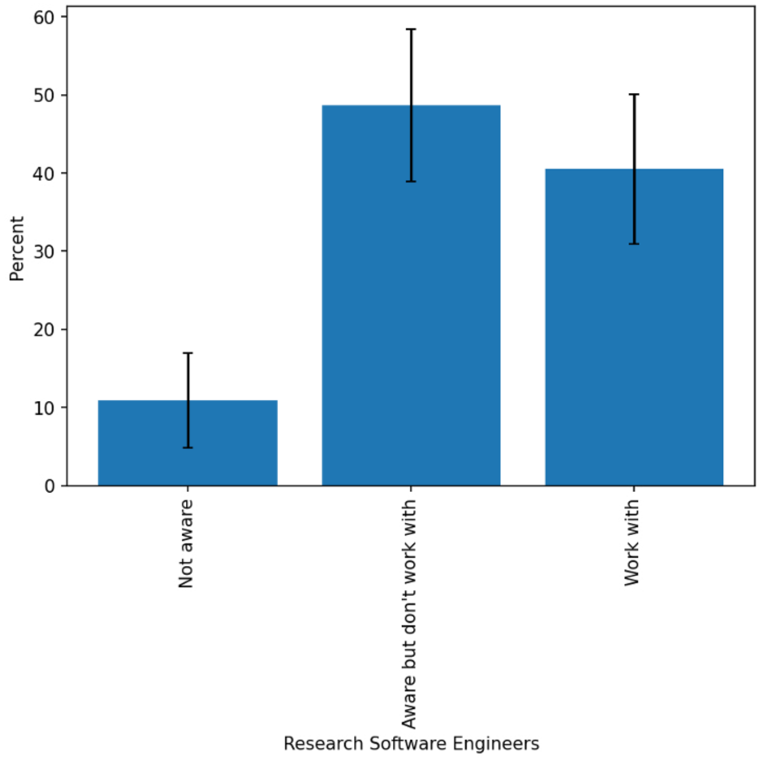


Q21. Would you like to expand on any of the answers you've given in this section? Free text response.

Q22.Which of the following supporting roles are you aware of, and which do you currently work with? (N = 405)

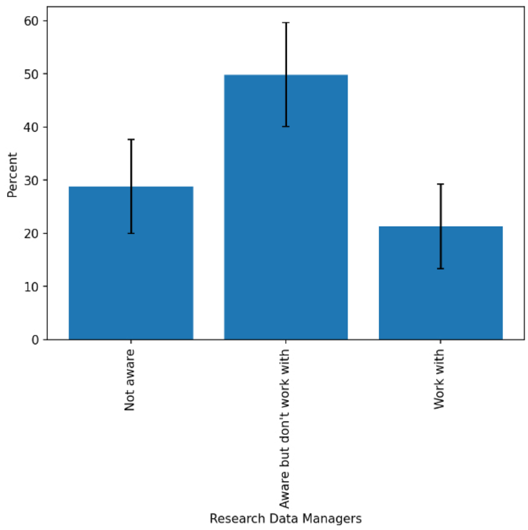
Q22.1. Research Software Engineers

Research Software Engineers	N	Percent
Aware but don't work with	192	48.61
Work with	160	40.51
Not aware	43	10.89



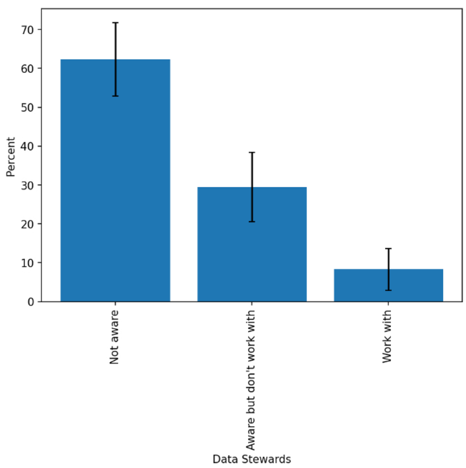
Q22.2. Research Data Managers

Research Data Managers	N	Percent
Aware but don't work with	192	49.87
Not aware	111	28.83
Work with	82	21.3



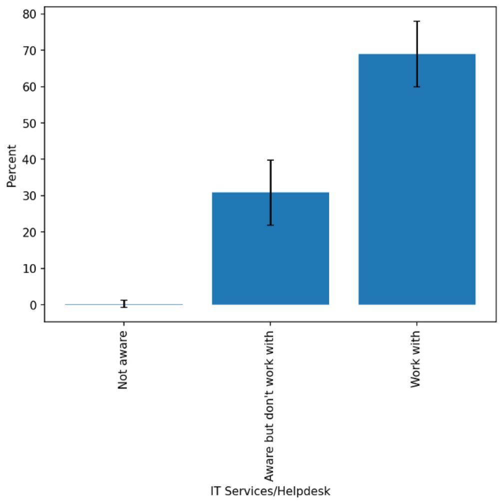
Q22.3. Data Stewards

Data Stewards	N	Percent
Not aware	239	62.24
Aware but don't work with	113	29.43
Work with	32	8.33



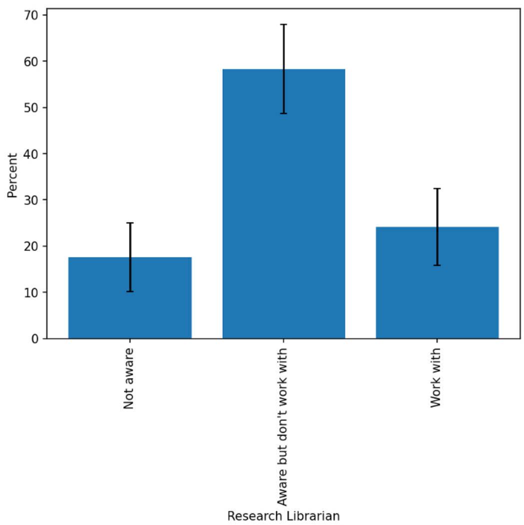
Q22.4. IT Services/ Helpdesk

IT Services/Helpdesk	N	Percent
Work with	275	68.92
Aware but don't work with	123	30.83
Not aware	1	0.25



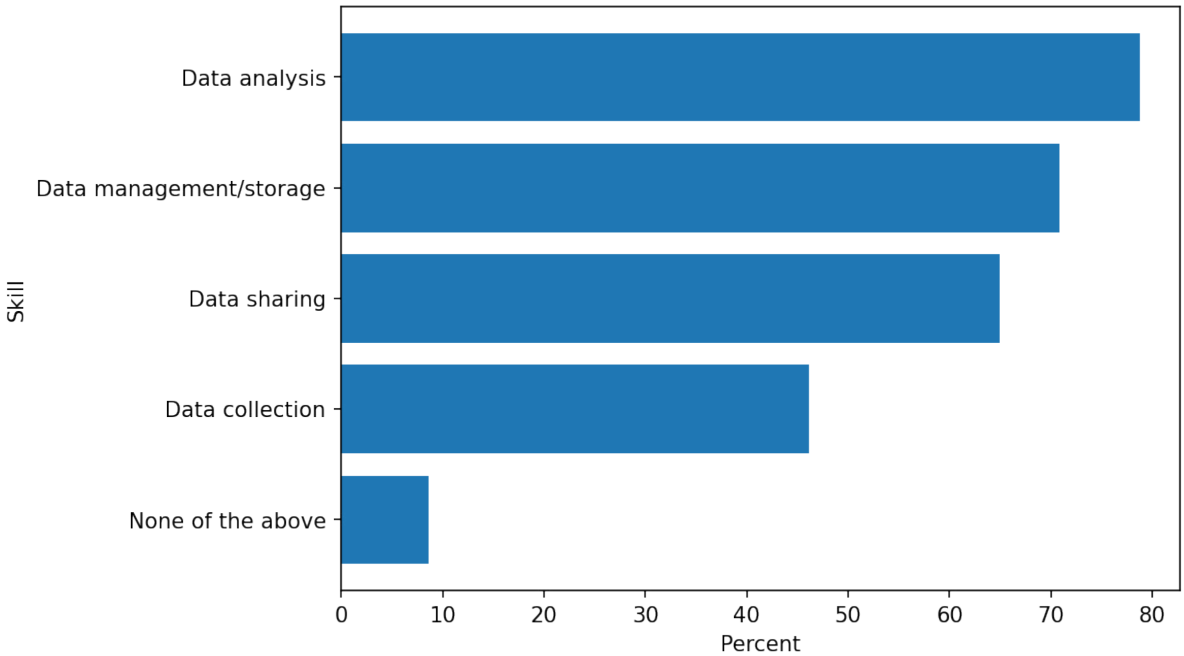
Q22.5. Research Librarian

Research Librarian	N	Percent
Aware but don't work with	225	58.29
Work with	93	24.09
Not aware	68	17.62



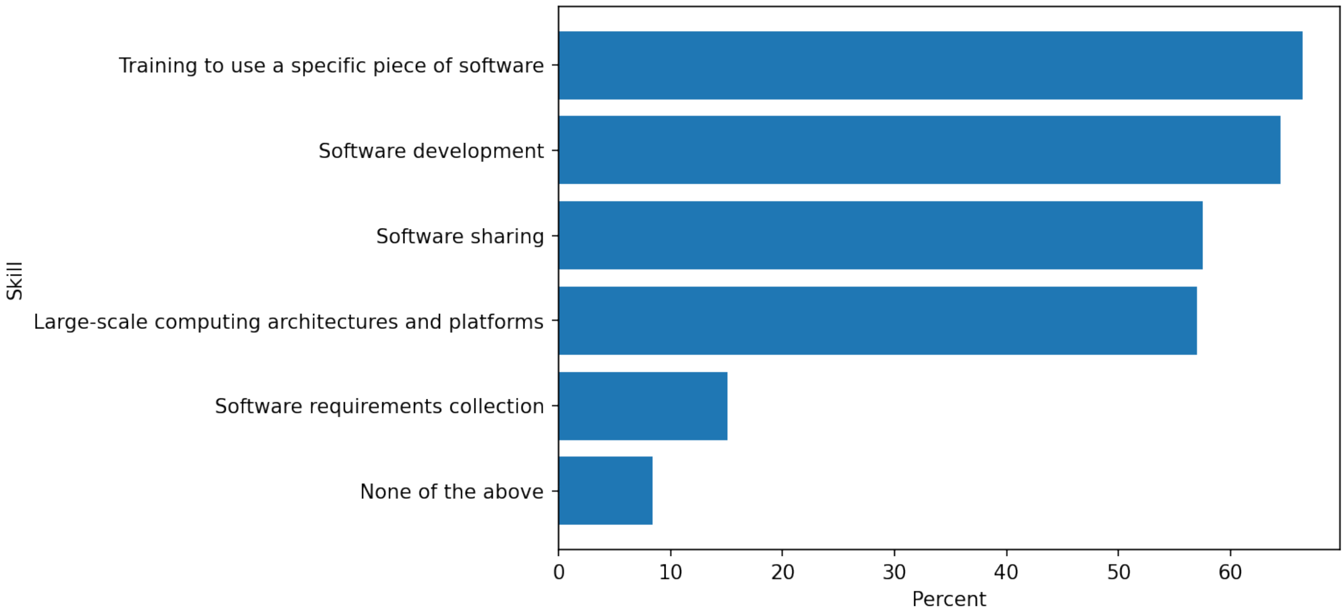
Q23. Over the last five years, have you learned new skills or knowledge in relation to any of the following? Select all that apply. (N = 405)

Skill	N	Percent
Data analysis	319	78.77
Data management/ storage	287	70.86
Data sharing	263	64.94
Data collection	187	46.17
None of the above	35	8.64



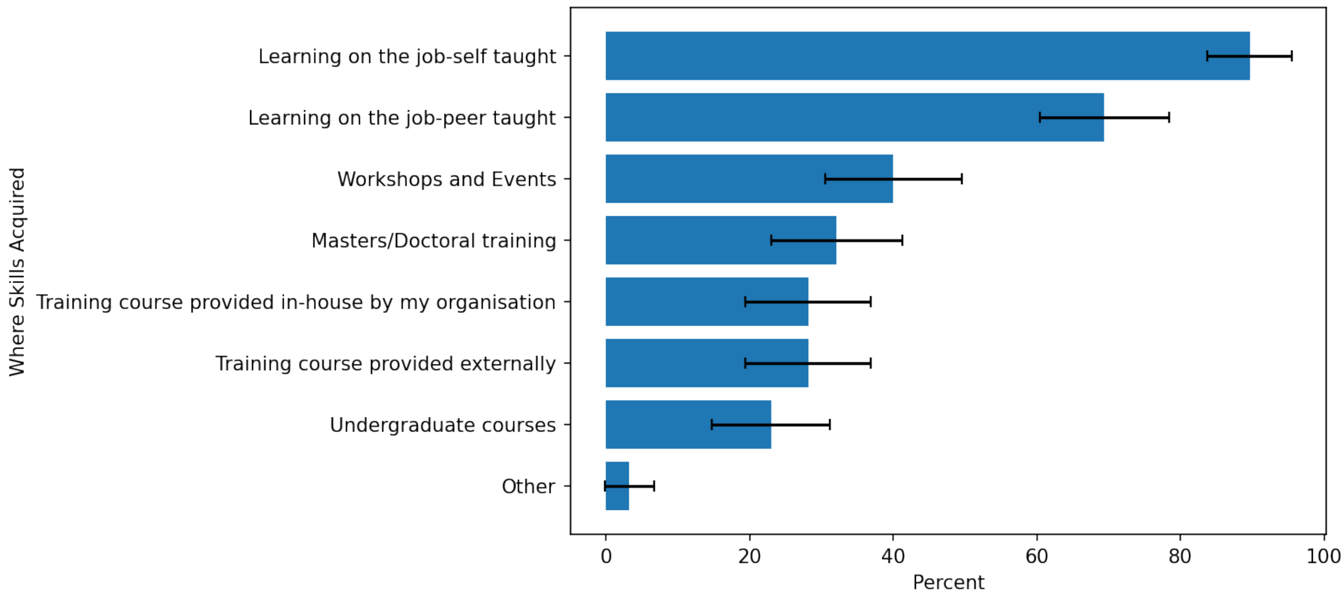
Q24. Over the last five years, have you learned new skills or knowledge in relation to any of the following? Select all that apply. (N = 405)

Skill	N	Percent
Training to use a specific piece of software	269	66.42
Software development	261	64.44
Software sharing	233	57.53
Large-scale computing architectures and platforms	231	57.04
Software requirements collection	61	15.06
None of the above	34	8.4



Q25. Where did you acquire your skills and knowledge about data and/or software? Select all that apply. (N = 405)

Where Skills Acquired	N	Percent
Learning on the job-self taught	363	89.63
Learning on the job-peer taught	281	69.38
Workshops and Events	162	40
Masters/Doctoral training	130	32.1
Training course provided in-house by my organisation	114	28.15
Training course provided externally	114	28.15
Undergraduate courses	93	22.96
Other	13	3.21

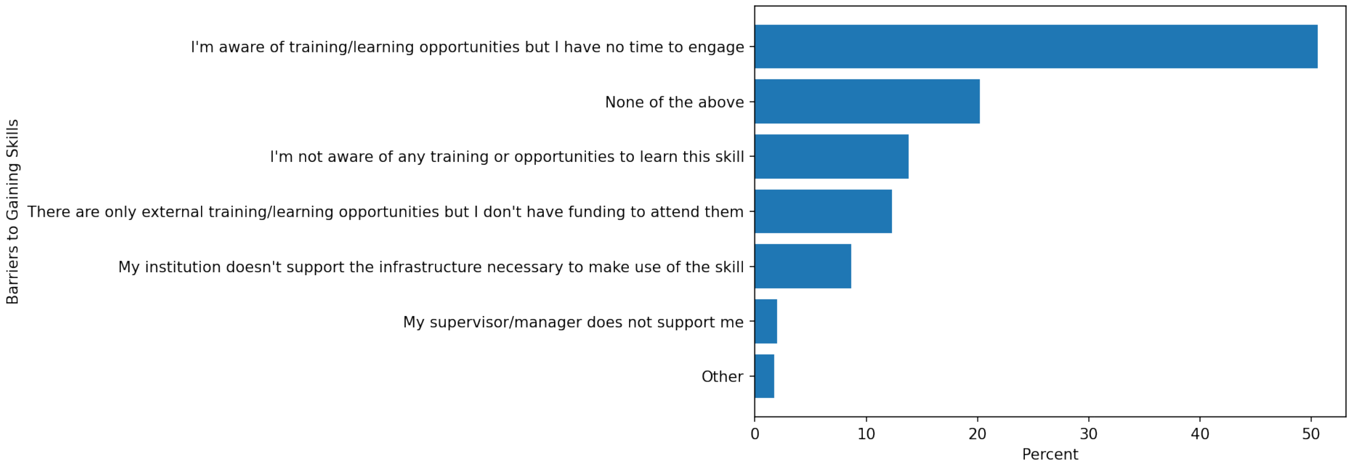


Q26. Thinking about your current skill set, what is the main data, software or computing skill that you would like to obtain to improve your research/work? Free text response.

ML/DL/AI/data mining	58
Codina/Programming improvement/good practices/documentation	48
Python	33
Data processing/management/access/sharing	31
None/Not applicable	29
GPUs/Cuda	19
Repositories/Version control/git	13
HPC/exascale	12
HTC/Cloud computing	11
Profiling/Optimisation	10
Statistics and statistical computing	10
Data Visualisation	9
Containers/K8s	8
C/C++	7
R	7
Web development/technologies	7
Julia	5
Testing/VUQ	5
Image Analysis	5
Database software/creation/management	5
Large-scale data	4
Front-end/UI	3
Funding/Business model	3
Modelling and Simulation	3
Quantum computing	3
leadership/management	2
Command Line Interface/Automation	2
Green Computing	1
Compilers	1
Collaboration	1

Q27. Do any of the following issues prevent you from gaining the skill you entered above? Select all that apply. (N = 405)

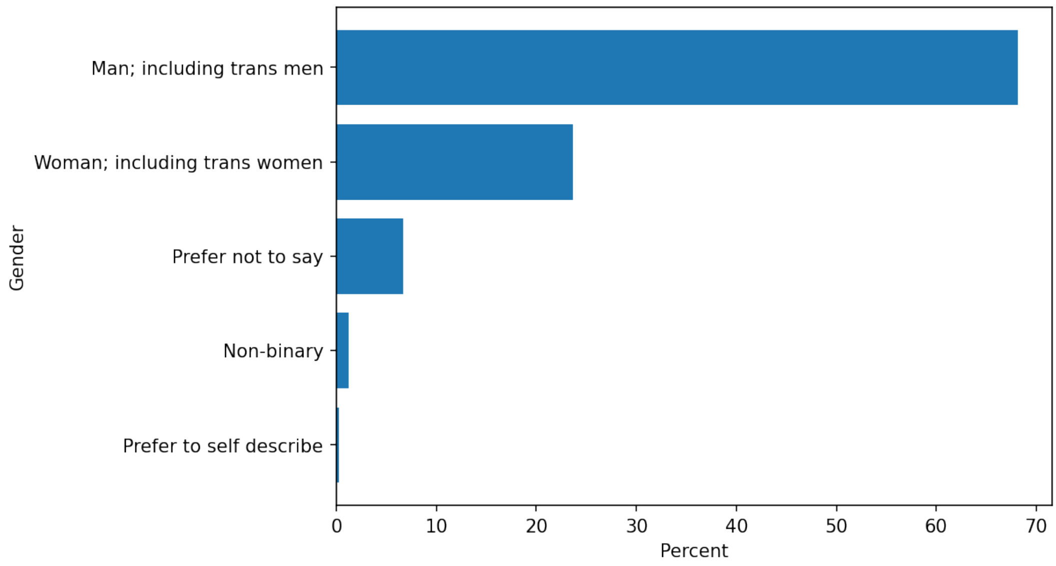
Barriers to Gaining Skills	N	Percent
I'm aware of training/learning opportunities but I have no time to engage	205	50.62
None of the above	82	20.25
I'm not aware of any training or opportunities to learn this skill	56	13.83
There are only external training/learning opportunities but I don't have funding to attend them	50	12.35
My institution doesn't support the infrastructure necessary to make use of the skill	35	8.64
My supervisor/manager does not support me	8	1.98
Other	7	1.73



Q28. Is there anything that you would like to change to enable the use of computing in research? Please use a maximum of 100 words. Free text response.

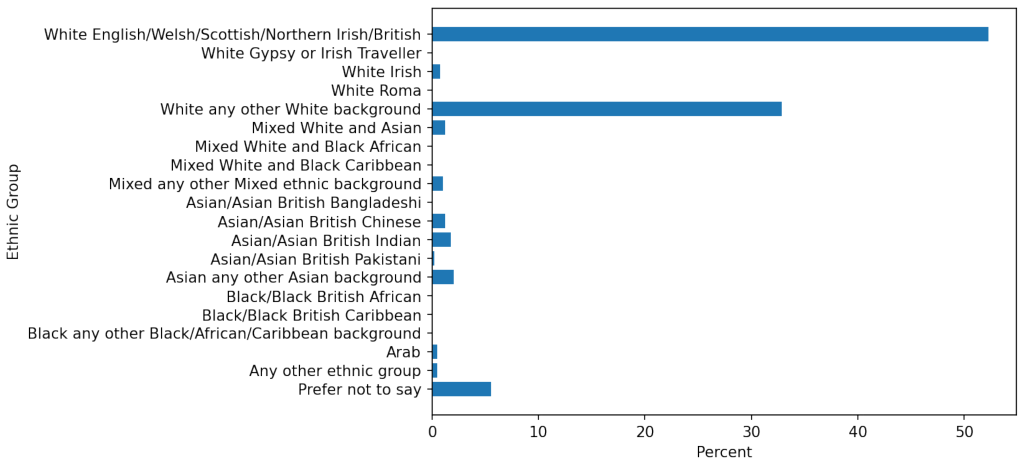
Q29. How would you describe your gender? (N = 402)

Gender	N	Percent
Man; including trans men	274	68.16
Woman; including trans women	95	23.63
Prefer not to say	27	6.72
Non-binary	5	1.24
Prefer to self-describe	1	0.25



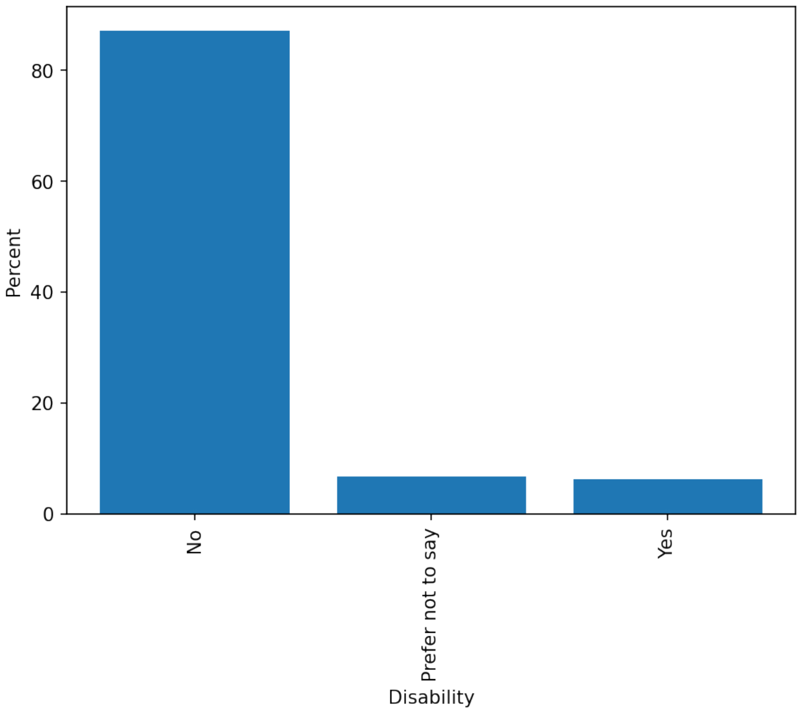
Q30. What is your ethnic group? (N = 404)

Ethnic Group	N	Percent
White English/Welsh/Scottish/Northern Irish/British	212	52.27
White Gypsy or Irish Traveller	0	0
White Irish	3	0.76
White Roma	0	0
White any other White background	133	32.83
Mixed White and Asian	5	1.26
Mixed White and Black African	0	0
Mixed White and Black Caribbean	0	0
Mixed any other Mixed ethnic background	4	1.01
Asian/Asian British Bangladeshi	0	0
Asian/Asian British Chinese	5	1.26
Asian/Asian British Indian	7	1.77
Asian/Asian British Pakistani	1	0.25
Asian any other Asian background	8	2.02
Black/Black British African	0	0
Black/Black British Caribbean	0	0
Black any other Black/African/Caribbean background	0	0
Arab	2	0.51
Any other ethnic group	2	0.51
Prefer not to say	22	5.56



Q31. Are your day-to-day activities limited because of a health problem or disability which has lasted, or is expected to last, at least 12 months? (N = 402)

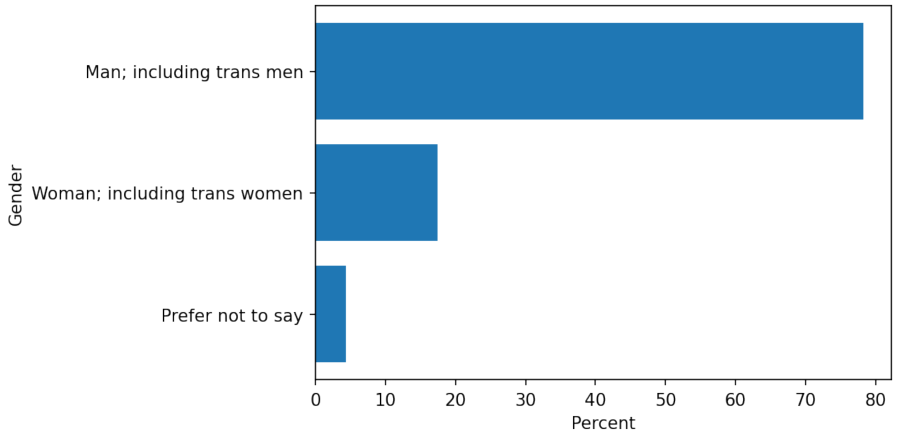
Disability	N	Percent
No	350	87.06
Prefer not to say	27	6.72
Yes	25	6.72



APPENDIX F: GROUP INTERVIEWS DEMOGRAPHIC DATA

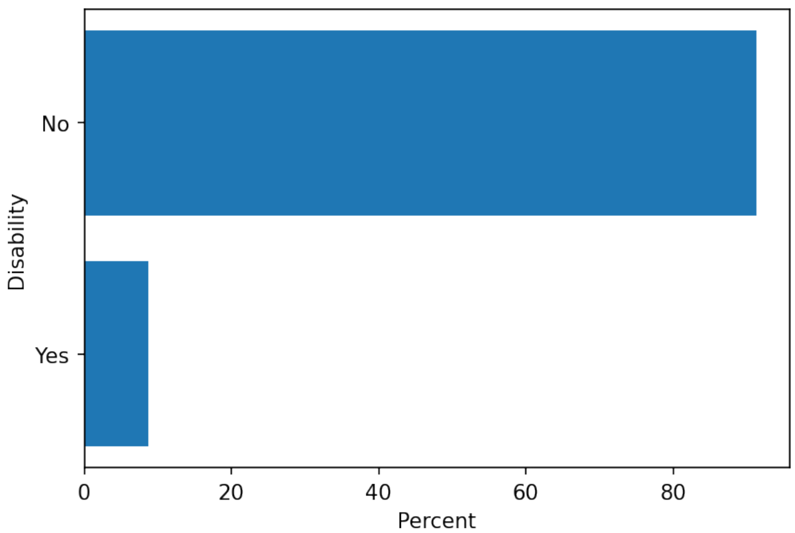
Gender

Gender	Percent
Man; including trans men	78.3
Woman; including trans women	17.4
Prefer not to say	4.3



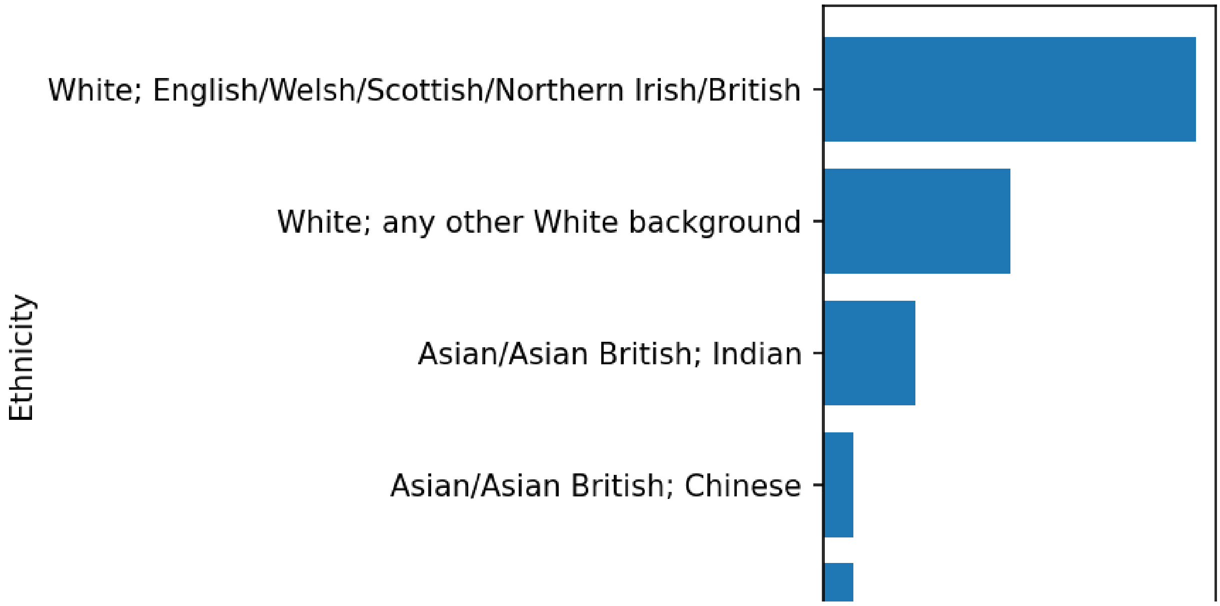
Disability

Disability	Percent
No	91.3
Yes	8.7



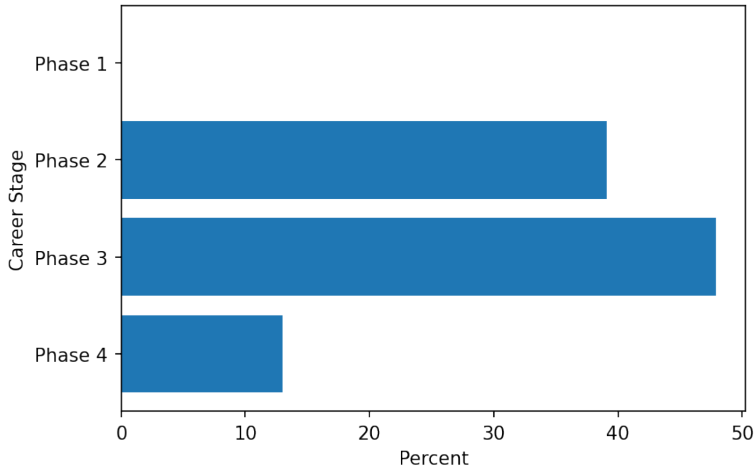
Ethnicity

Ethnicity	Percent
White; English/Welsh/Scottish/Northern Irish/British	52.27
White; any other White background	0
Asian/Asian British; Indian	0.76
Asian/Asian British; Chinese	0
Asian, any other Asian background	32.83



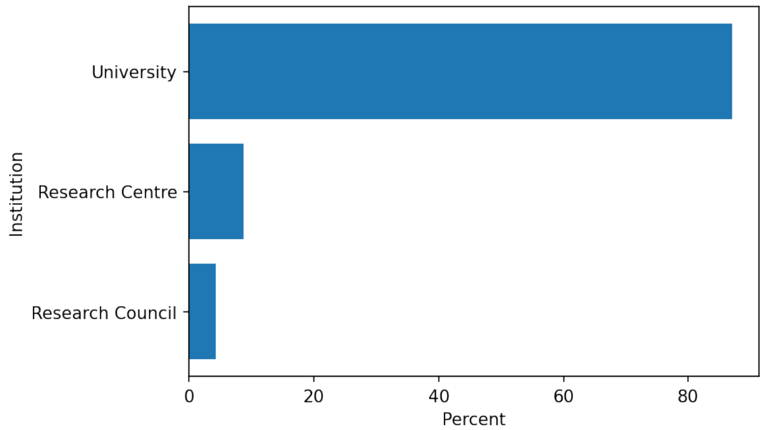
Career Stage

Career Stage	Percent
Phase 1	0
Phase 2	39.1
Phase 3	47.9
Phase 4	13



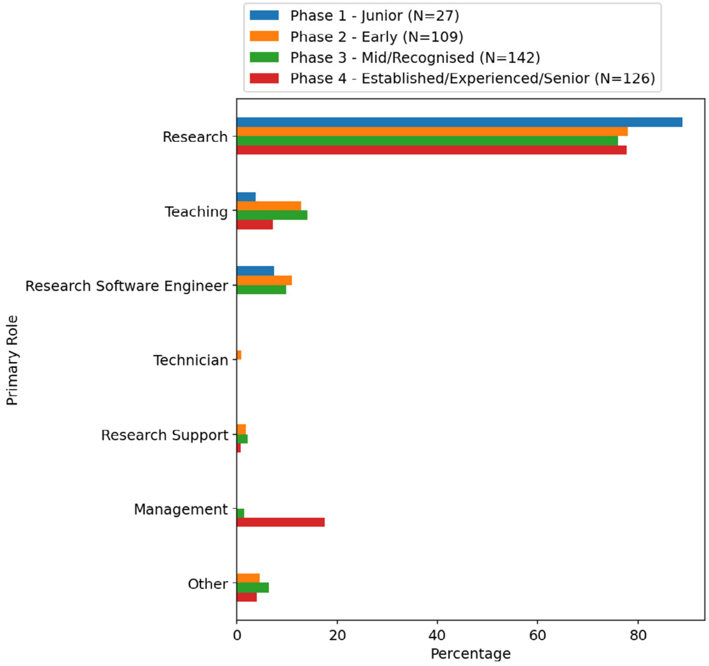
Institution

Institution	Percent
University	87
Research Centre	8.7
Research Council	4.3

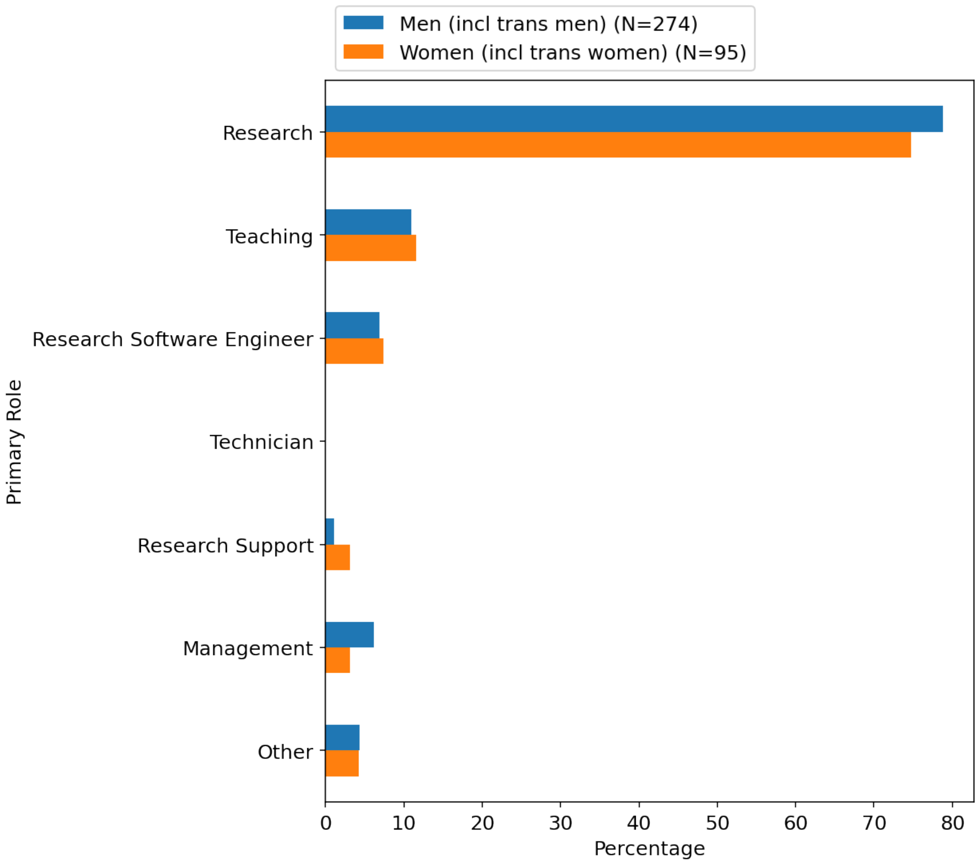


APPENDIX G: SURVEY RESPONSES - MULTIVARIATE ANALYSES

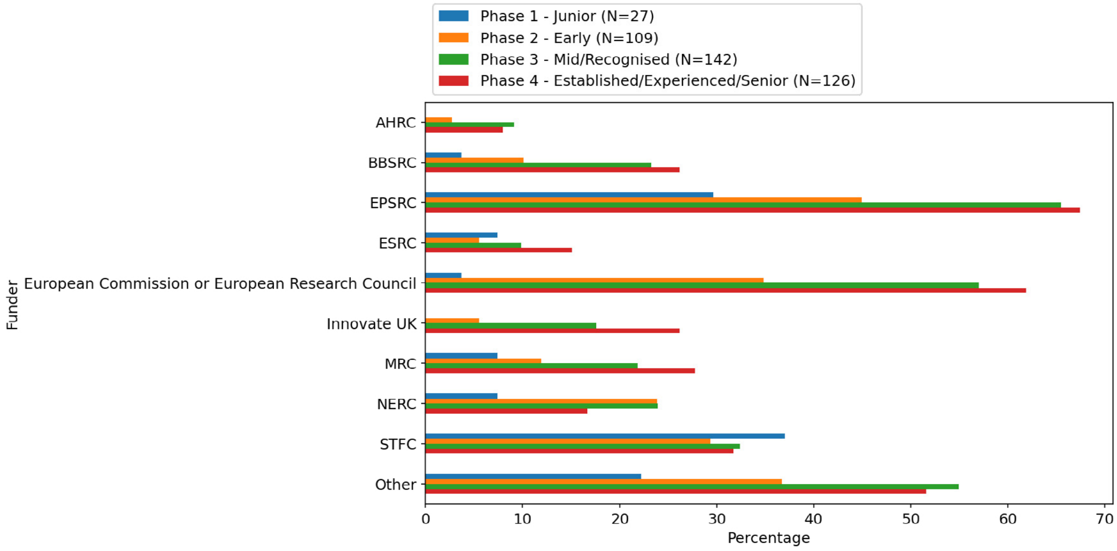
Q5 x Q6 (Primary role, by career stage)



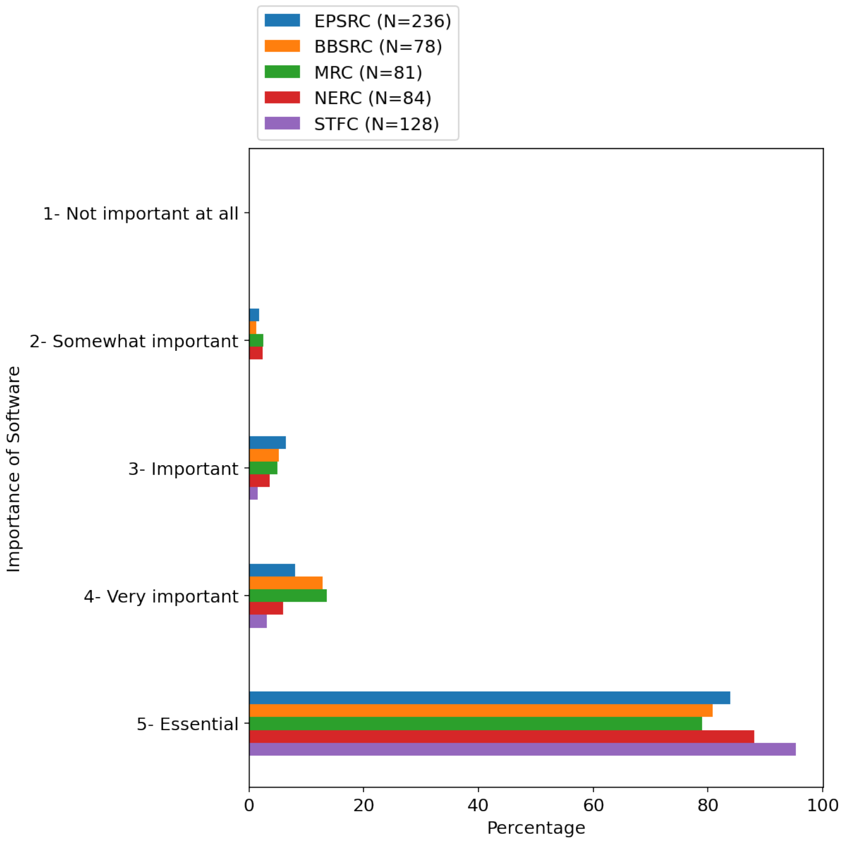
Q5 x Q29 (Primary role, by gender)



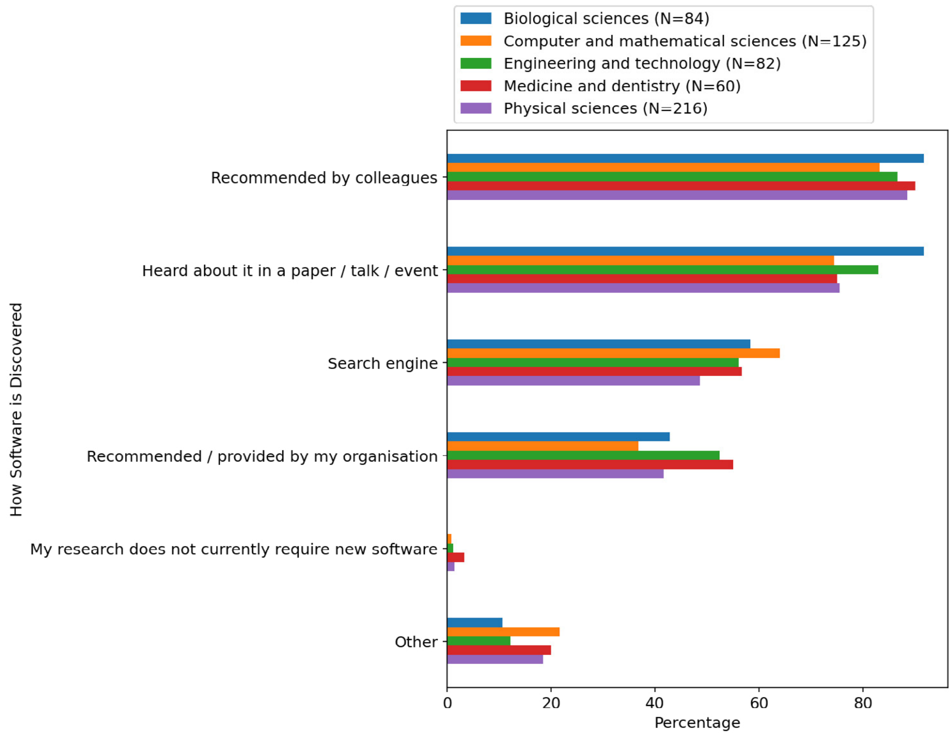
Q7 x Q6 (Funders that survey participants have applied to or who fund their work, by career stage)



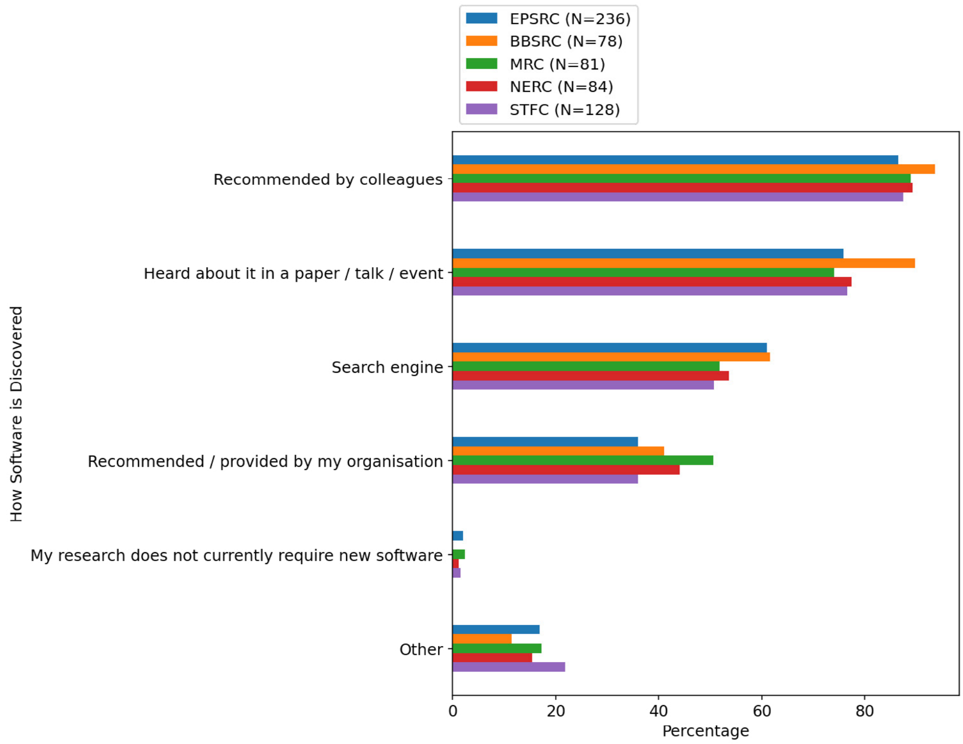
Q8 x Q7 (Importance of software, by funder)



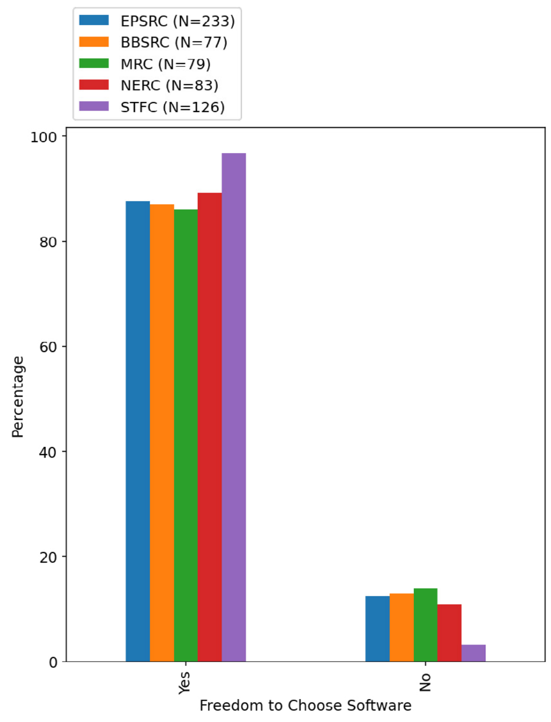
Q9 x Q4 (Software discovery, by discipline)



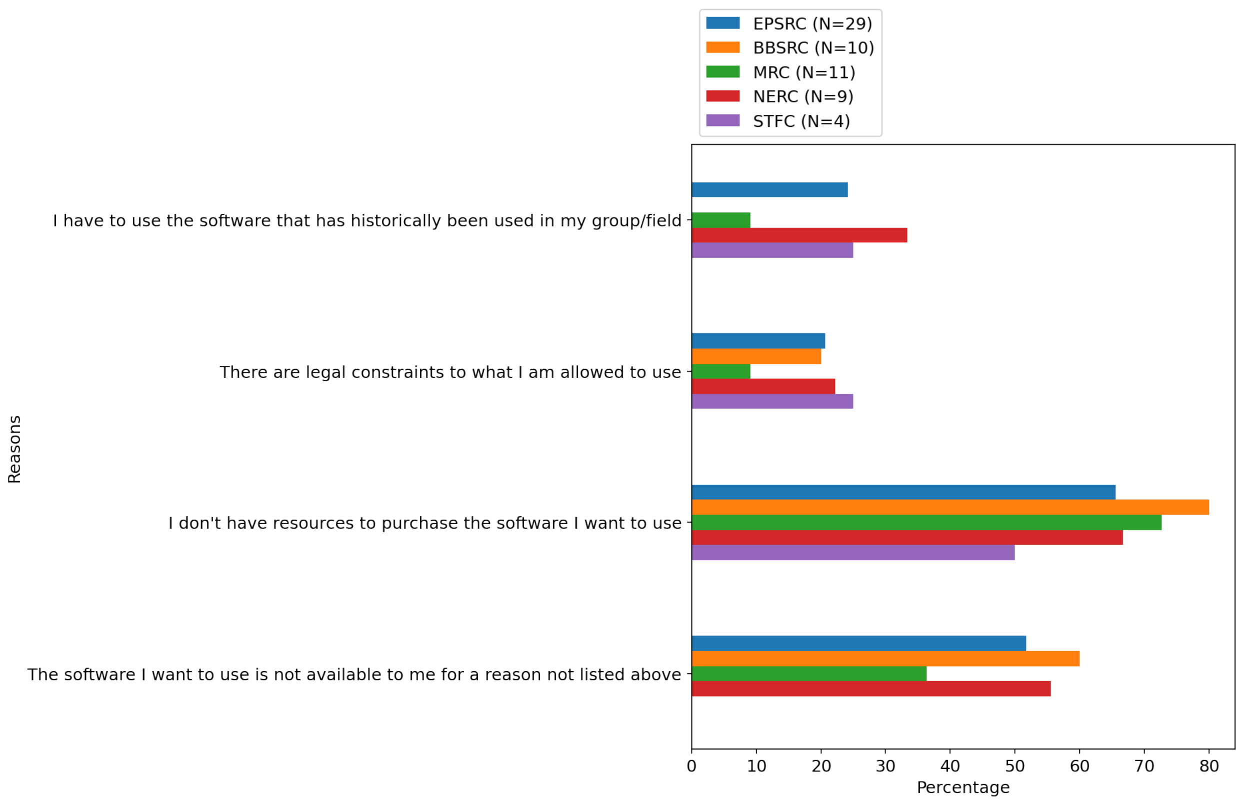
Q9 x Q7 (Software discovery, by funder)



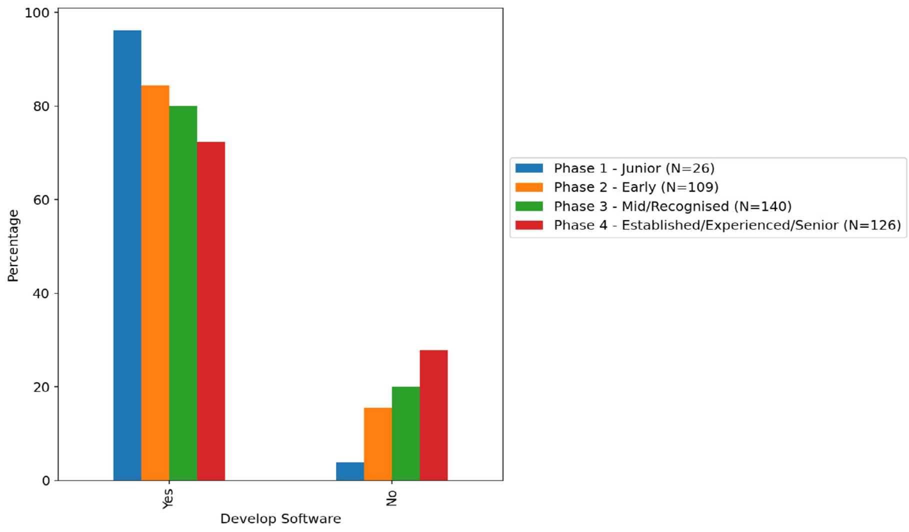
Q10 x Q7 (Are you free to use the software you want, by funder)



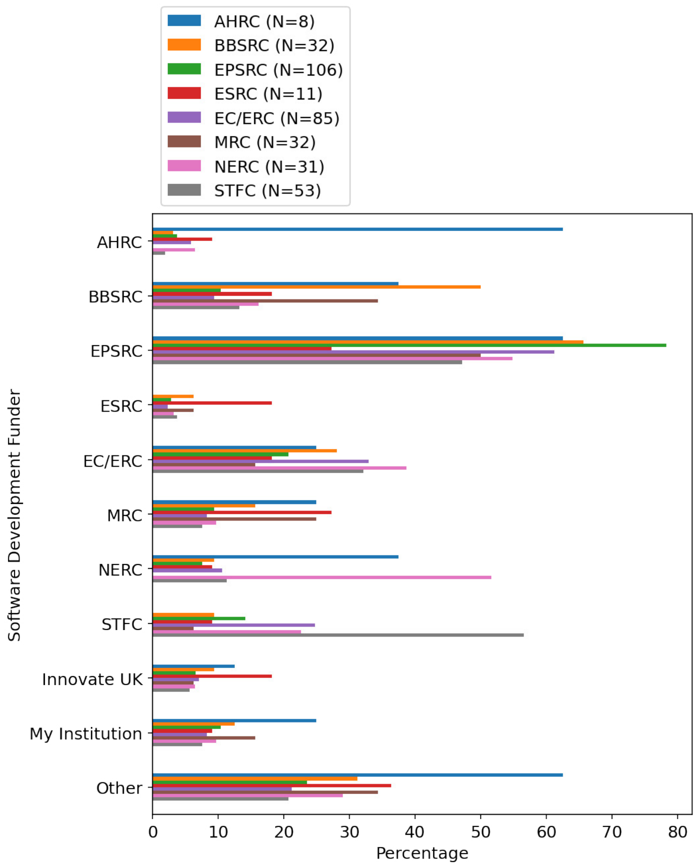
Q10.a x Q7 (Barriers faced by survey participants who were not free to use the software they want, by funder)



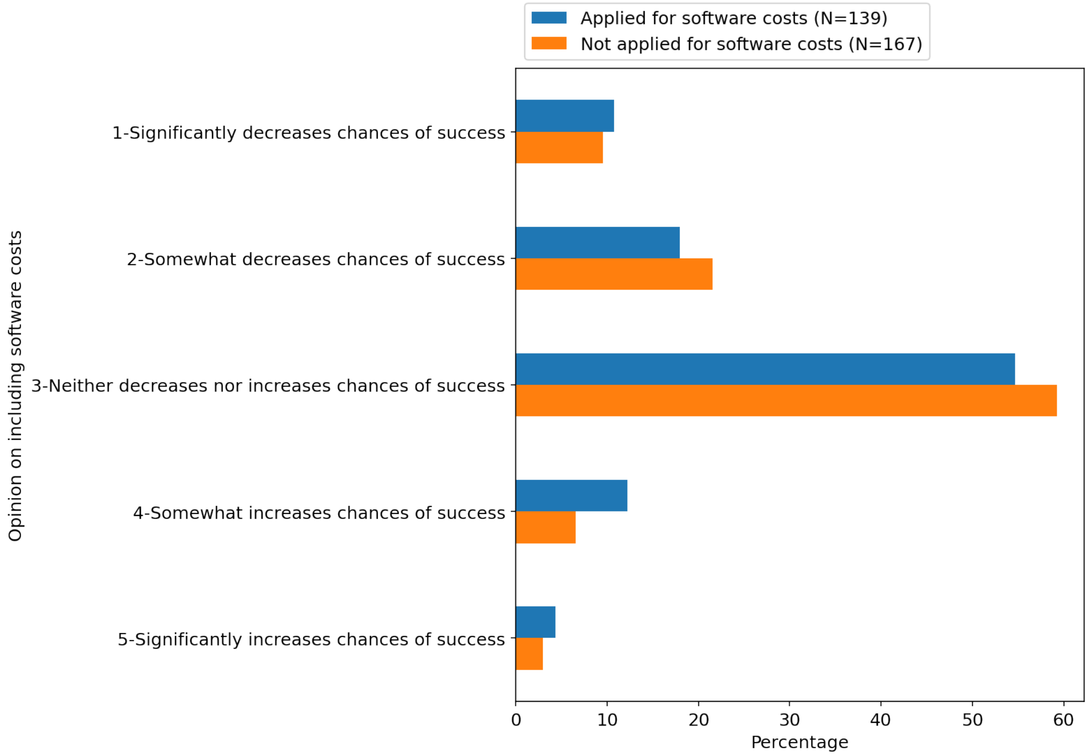
Q12 x Q6 (Whether participants develop software, by career stage)



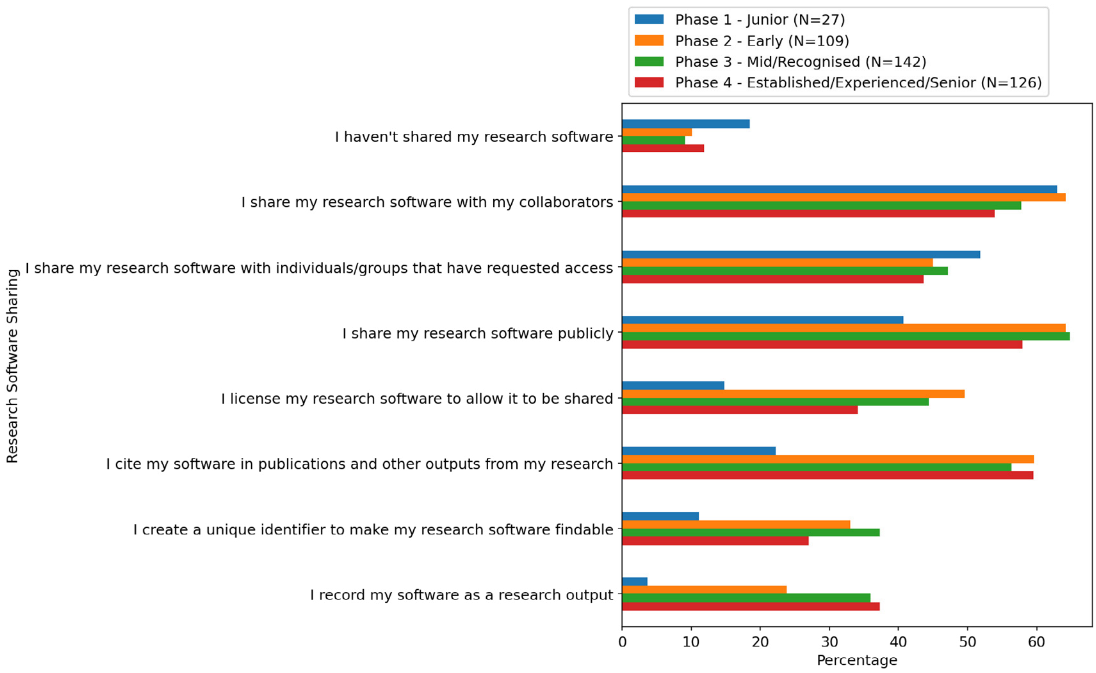
Q12.b x Q7 (Funders that survey participants had applied to for funding that explicitly includes costs for software development [y-axis], compared with funders that they applied for and/or were funded by more generally [legend])



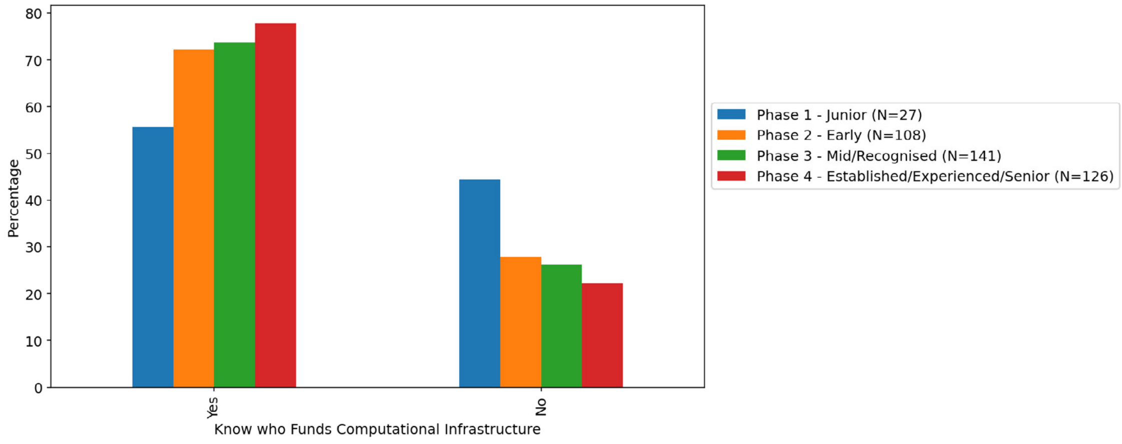
Q13 x Q12.b (Opinions of survey participants on including software development costs in a funding proposal, by whether survey participants had applied for funding explicitly including costs for software development)



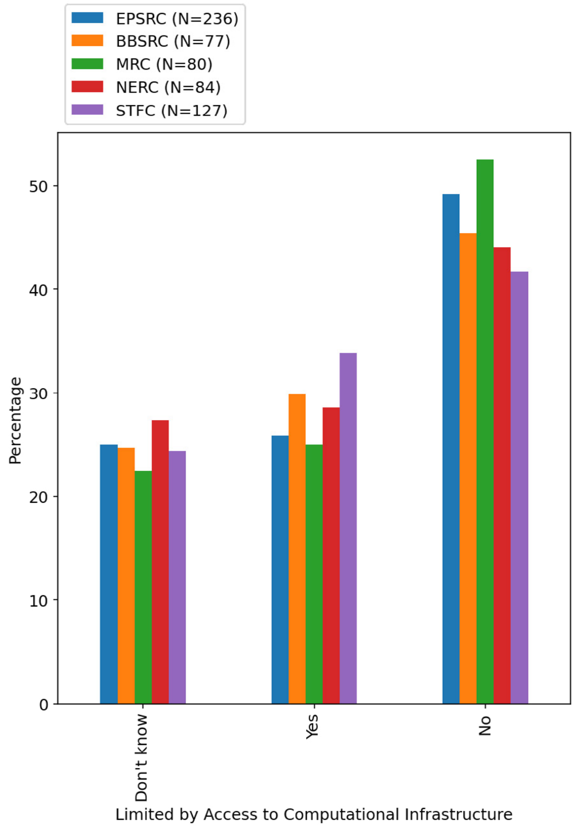
Q14 x Q6 (Sharing software practices, by career stage)



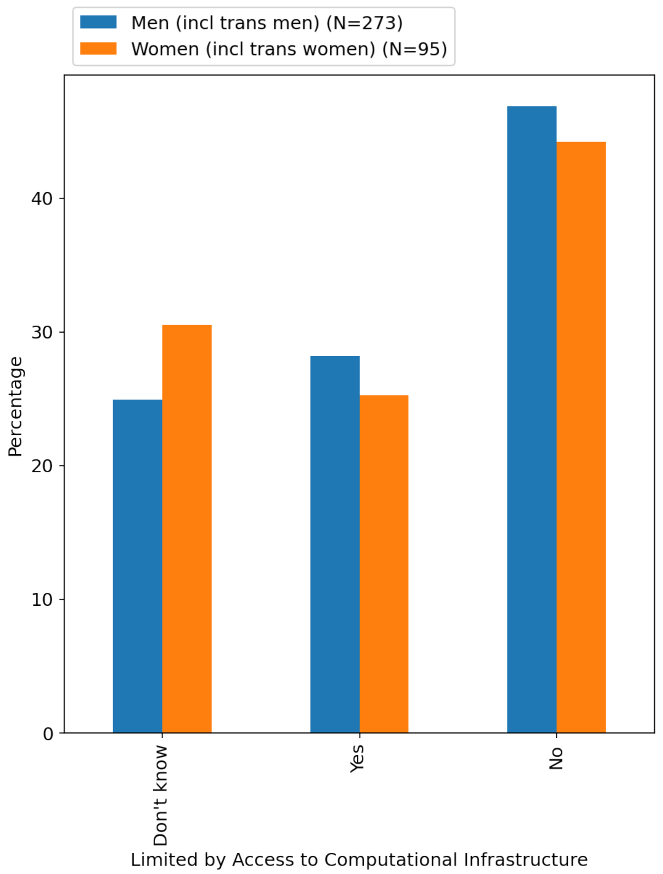
Q16 x Q6 (Know about computational infrastructure funders, by career stage)



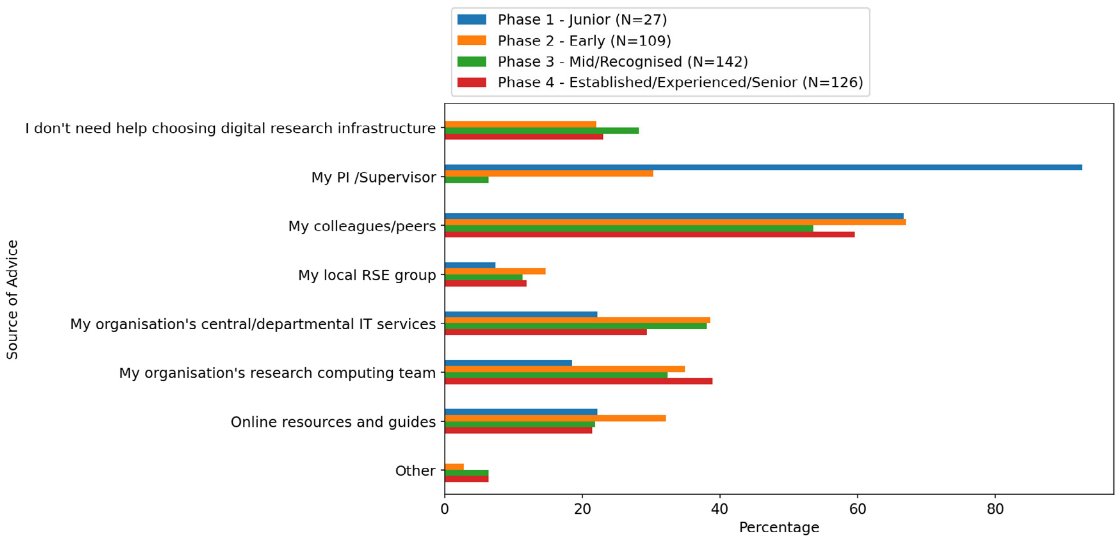
Q19 x Q7 (Whether survey respondents feel that in the next year their research will be limited by access to computational and/or data research infrastructure, by funder)



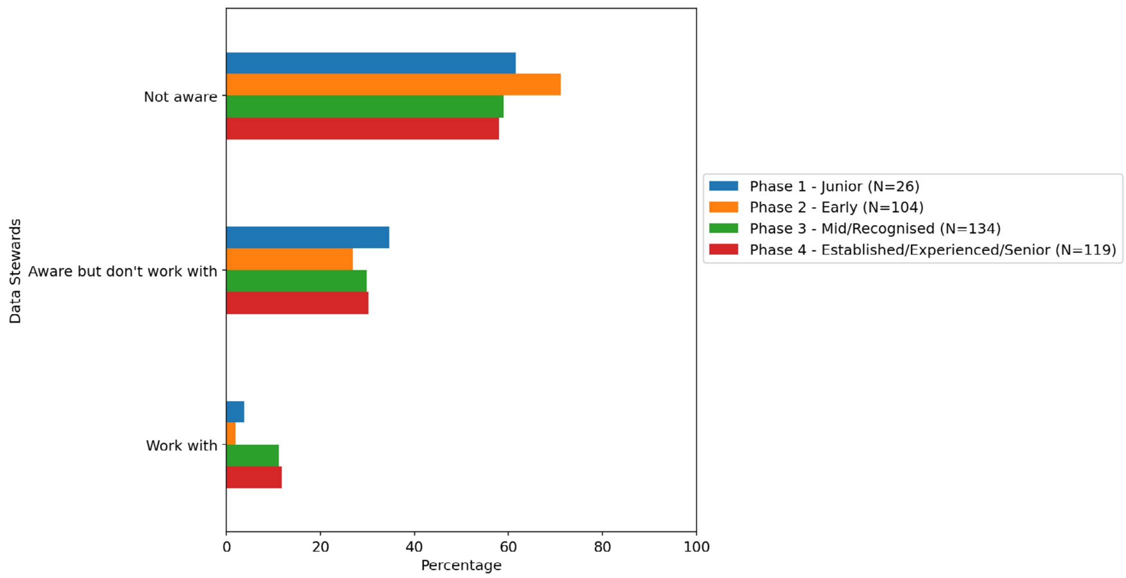
Q19 x Q29 (Whether survey participants felt that their research would be limited by access to computational and/or data research infrastructure, by gender)



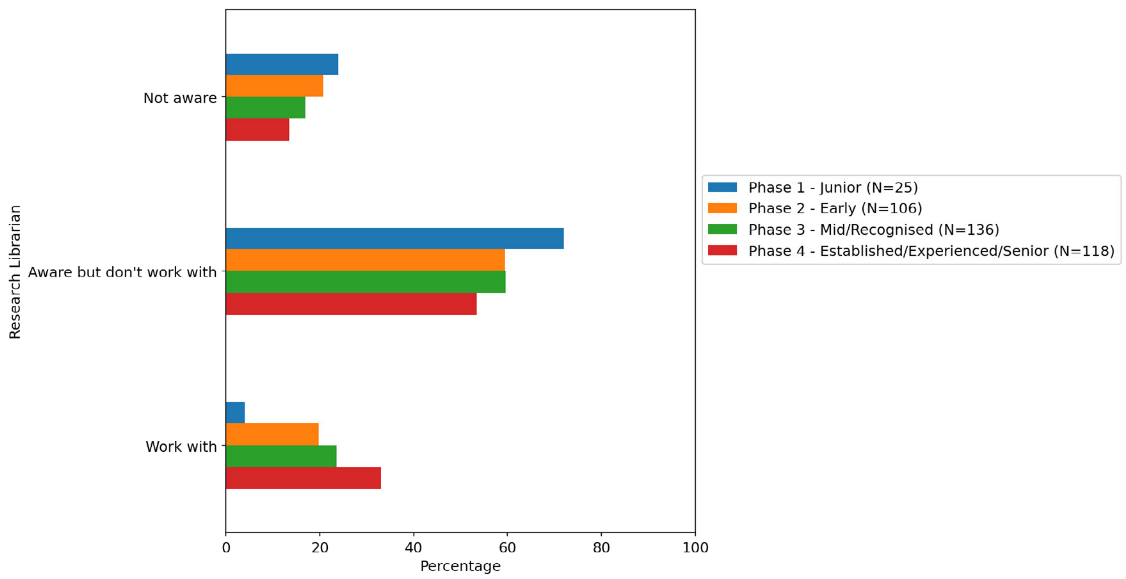
Q20 x Q6 (Who survey participants ask for advice on choosing the computational and/or data research infrastructure they use, by career stage)



Q22.3 x Q6 (Survey participants' awareness of Data Steward role)

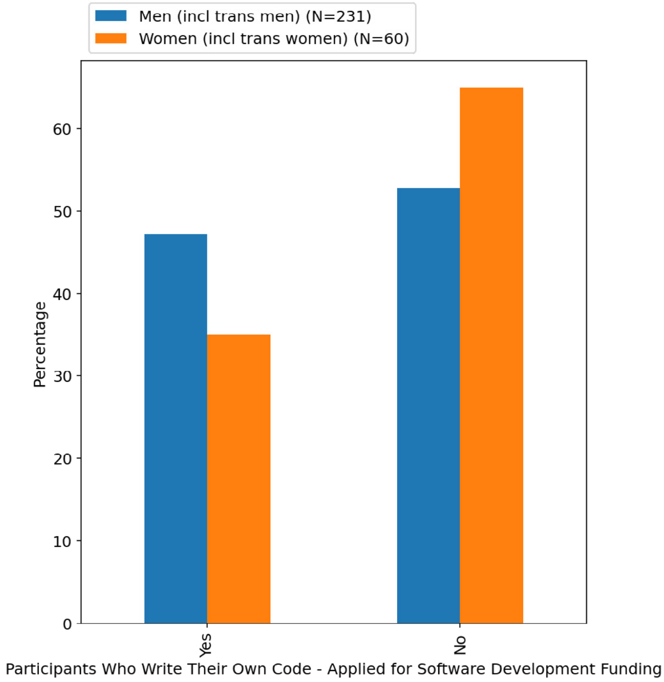


Q22.5 x Q6 (Survey participants' awareness of Research Librarian role, by career stage)

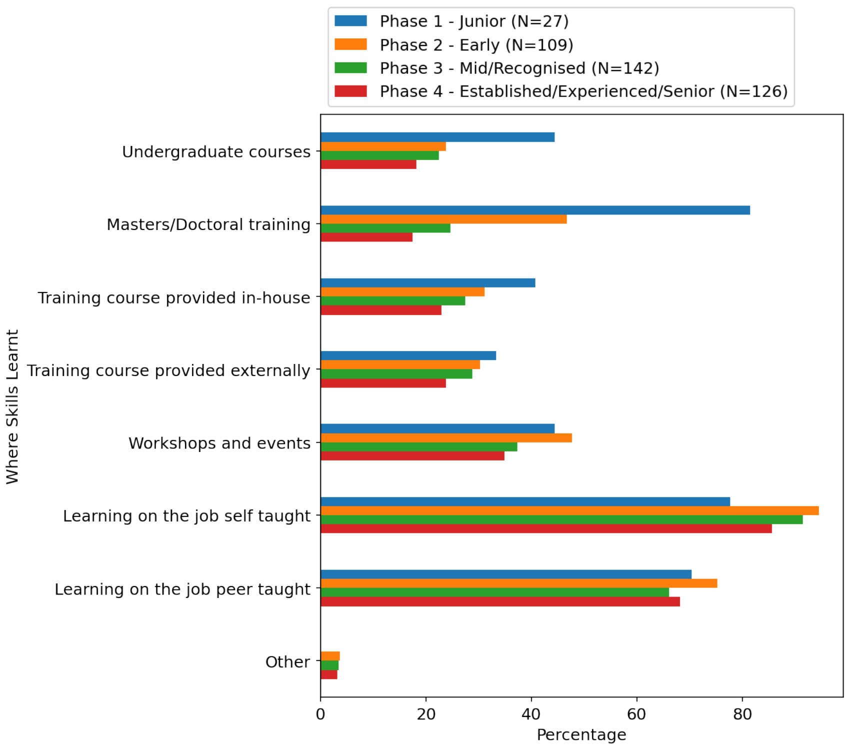


Q23 & 24 x Q6 (New skills or knowledge learned by survey participants over the last five years, by career stage)

Q12.b x Q29 (Percentage of survey participants who have applied for funding that explicitly includes costs for software development, including only participants who write their own code, by gender)



Q25 x Q6 (How participants acquire their skills and knowledge, by career stage)



APPENDIX H: MATCHING OF HESA COST CENTRES AND PRINCIPAL SUBJECT CODES

Cost centre (HESA, n.d.)	Principal subject code (HESA, 2013)
133. Business & management studies	D. Business & administrative studies
134. Catering & hospitality management	D. Business & administrative studies
109. Veterinary science	4. Veterinary science
110. Agriculture, forestry & food science	5. Agriculture & related subjects
123. Architecture, built environment & planning	A. Architecture, building & planning
111. Earth, marine & environmental sciences	6. Physical sciences
112. Biosciences	3. Biological sciences
113. Chemistry	6. Physical sciences
114. Physics	6. Physical sciences
122. Mathematics	7. Mathematical sciences
143. Art & design	H. Creative arts & design
144. Music, dance, drama & performing arts	H. Creative arts & design
108. Sports science & leisure studies	3. Biological sciences
135. Education	I. Education
136. Continuing education	I. Education
115. General engineering	9. Engineering & technology
116. Chemical engineering	9. Engineering & technology
117. Mineral, metallurgy & materials engineering	9. Engineering & technology
118. Civil engineering	9. Engineering & technology
120. Mechanical, aero & production engineering	9. Engineering & technology
121. IT, systems sciences & computer software engineering	8. Computer science
125. Area studies	G. Historical & philosophical studies
126. Archaeology	G. Historical & philosophical studies
137. Modern languages	F. Languages
138. English language & literature	F. Languages
139. History	G. Historical & philosophical studies
140. Classics	F. Languages
141. Philosophy	G. Historical & philosophical studies
142. Theology & religious studies	G.Historical & philosophical studies

Cost centre (HESA, n.d.)	Principal subject code (HESA, 2013)
101. Clinical medicine	1. Medicine & dentistry
102. Clinical dentistry	1. Medicine & dentistry
103. Nursing & allied health professions	2. Subjects allied to medicine
104. Psychology & behavioural sciences	3. Biological sciences
105. Health & community studies	2. Subjects allied to medicine
106. Anatomy & physiology	2. Subjects allied to medicine
107. Pharmacy & pharmacology	2. Subjects allied to medicine
124. Geography & environmental studies	6. Physical sciences
127. Anthropology & development studies	B. Social studies
128. Politics & international studies	B. Social studies
129. Economics & econometrics	B. Social studies
130. Law	C. Law
131. Social work & social policy	B. Social studies
132. Sociology	B. Social studies
145. Media studies	E. Mass communications & documentation

Software
Sustainability
Institute

www.software.ac.uk