Software
Sustainability
Institute

SHAPING DATA AND SOFTWARE POLICY IN THE ARTS AND HUMANITIES RESEARCH COMMUNITY

A Study for the AHRC

Rebecca Taylor, Johanna Walker, Simon Hettrick, Philippa Broadbent (University of Southampton, Software Sustainability Institute), and David de Roure (University of Oxford)

July 2022

CONTENTS

FIGURES AND TABLES

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AHRC	Arts and Humanities Research Council
BBSRC	Biotechnology and Biological Sciences Research Council
DTP	Doctoral Training Partnership
ECR	Early Career Researcher
ESRC	Economic and Social Science Research Council
EPSRC	Engineering and Physical Sciences Research Council
FAIR	Findable, accessible, interoperable, reusable (principles for digital assets)
GDPR	General Data Protection Regulations
GIS	Geographic Information Systems
GLAMs	Galleries, Libraries, Archives and Museums
HEI	Higher Education Institution
HESA	Higher Education Statistics Agency
IRO	Independent Research Organisation
JACS	Joint Academic Coding System
LERU	League of European Research Universities
MRC	Medical Research Council
NERC	Natural Environment Research Council
PGR	Postgraduate
REF	Research Excellence Framework
RSE	Research Software Engineer
STFC	Science and Technology Facilities Council
UKRI	UK Research and Innovation

EXECUTIVE SUMMARY

BACKGROUND

Digital tools and software are revolutionising the nature of data and research across the arts and humanities community; how data is collected and analysed, and how it is managed, shared and sustained for future generations.

> The principles of open research and the necessity of providing effective infrastructure to support the changing shape of research are driving policy and practice across UKRI, higher education and government.

> The rich disciplinary diversity that characterises arts and humanities research poses challenges for providing infrastructure and support for research skills that can meet wide ranging needs and priorities.

> There is a lack of evidence about current research practices and levels of engagement in and use of digital tools in the arts and humanities community and the implications of this for digital skills gaps and needs.

> AHRC commissioned a team at the Universities of Southampton and Oxford to undertake a study into how best to support and build the skills, knowledge and capacity of the research community to utilise digital tools and infrastructure and ensure world class research.

RESEARCH METHODS

Our study involved an iterative three stage, mixed methods research design that entailed:

1. A scoping stage which involved a mapping of the arts and humanities community; a review of literature and UKRI data policies and 12 exploratory interviews with researchers and stakeholders.

2. A large-scale survey of researchers from across the community, including those in a wide range of higher education institutions and independent research organisations; in various roles (academic, professional and technical); at various career stages (junior, early, mid and senior); and from a broad array of arts and humanities disciplines. Responses were collected from 335 researchers (281 complete). The survey measured current research practices including data collection and analysis, data management and sustainability, use of software, institutional support and research skills.

3. In-depth semi structured interviews with 30 survey participants who had opted in to further research and were purposively selected to capture disciplinary and career stage diversity. The interviews explored research practices, experiences, cultures and skills including barriers and facilitators to training and development, forms of institutional support and future policy scenarios.

KEY RESULTS AND FINDINGS

Data and software practices

> Arts and humanities researchers were collecting and analysing an array of digital and non-digital data forms, often more than one. Over 80% were working with textual data, 70% with images, 45% with audio, 42% with physical artefacts and 40% with numerical data.

> 75% of respondents had used software in their research. Of those over 50% used resources provided by their institutions and around 25% were using open-source software. Just 17% were developing their own software or drawing on expertise to develop bespoke software for their research, which is lower than other communities.

> Both survey and interview data suggested a spectrum of engagement with digital tools and software ranging from committed users to steadfastly 'analogue' and theoretical approaches. Most respondents fell somewhere in between.

> Routine sharing of data was low, only 27% said they regularly shared it. 22% said they had never shared their data. Digital humanities scholars were more likely than others in the community to be engaged in routine data sharing practices.

> Around 38% of researchers mainly stored their data at the end of the project in a shared area provided by their institution (i.e., Google Drive or SharePoint) and 17% used a personal shared

area. Together only around 12% were using data repositories (either institutional or external). More of a concern was that 25% of respondents said their data was stored on a laptop or computer, suggesting this was not shared and could potentially be lost. Unsurprisingly this was particularly the case for doctoral students who had not completed their research.

> The emergence of new digital forms of data collection and analysis was related to more collaborative and cross-disciplinary research, that contrasted with the traditional 'lone scholar' model typical in arts and humanities. Research teams also included various professional technical roles: developers, technicians, librarians, data analysts. The interviews found that in projects with a technical element it was more common to bring in digital expertise rather than develop it within the team. Finding domain experts with digital skills was challenging.

Digital skills development

> Almost 70% of respondents identified learning data management and storage skills in the past 5 years. Just over half of the respondents had also gained data analysis skills and just over 40% had acquired data collection and data sharing skills. This was the case for all career stages, except junior researchers who were less likely to have done data sharing training.

> Asked about where researchers acquired skills and knowledge in data and software, the two categories with the highest response rates were 'learning on the job', either by being self-taught (70%) or through peer teaching of skills by colleagues (50%). Fewer indicated formal training mechanisms such as postgraduate study (35%) or external training courses (15%). The most well used formal training was that provided by institutions, possibly because this was locally available and likely to be free of charge. Around 35% of respondents had accessed this over the past 5 years.

> In response to a question about barriers that prevented researchers learning data and software skills, 45% of respondents said they were aware of training opportunities but did not have time to attend, around 27% said they were not aware of training opportunities, and just over 10% felt that their institution did not support the technology or infrastructure necessary to make use of the skill.

> The interview discussions shed light on the barriers to formal training, particularly work intensification in the sector combined with institutional cultures and practices which did not directly reward skills development. Early, mid-career and senior researchers all described being overstretched and unable to prioritise training.

> There was extensive discussion of more informal mechanisms for skills development, including self-teaching and peer teaching that were free of charge and could be undertaken flexibly. The former involved using online lectures and videos, and reading 'how to' guides. Peer teaching involved learning specific skills from colleagues and team members.

> Communities of practice played an important role in supporting peer teaching. Participants had joined online forums and groups focused on a particular methodology or tool that provided virtual communities of practice. Attending and talking to presenters at conferences and seminars highlighted skills gaps, new techniques and potential networks. Others working in teams and on cross-disciplinary projects described how these enabled the transfer of skills and knowledge between team members.

Policies and infrastructure

> Our review of UKRI data policies found that councils which had modified and adapted UKRI guidance were able to set clear policy agendas, in particular around reuse, deposition in a repository/archive, training and skills. The AHRC's data policy was the least tailored and specific, suggesting there may be variation in how researchers interpret guidance.

45%
of respondents were aware of training opportunities but did not have time to attend

27%
were not aware of training opportunities

10%
felt their institution did not support the infrastructure necessary to make use of the skill

> Researchers identified a range of challenges to sustaining data and software, linked to institutional infrastructure. These included:

> > adequacy of and access to repositories and platforms and effective guidance on their use

> > long term usability and compatibility of stored data

> > substantial complexity of legal and technical infrastructure working across multiple institutions

> > lack of stewardship and oversight.

> Many researchers were supportive of sustaining data in spite of the infrastructure challenges.

CONCLUSIONS AND POLICY IMPLICATIONS

> **Diversity of data and software practice** - This is a central feature of the community and should be celebrated. Equally, it requires a tailored and sensitive approach to research policy and guidance that speaks to the groups within the community and provides a clear steer for research practice and skills development. Data policies need to be updated to address all stages of the research process and to include software.

> **Collaboration and interdisciplinarity** - The rise of large team-based studies and interdisciplinary research is shaping and driving digital transformation in the community. There are challenges in providing research infrastructure that can support and promote this.

> **A spectrum of engagement with digital tools and digital skills** - There is considerable digital expertise at one end and low use of digital tools at the other. It is the broad swathe of researchers between the two ends of this spectrum where intervention is likely to have the greatest impact.

> **The use of digital tools and associated knowledge and skills** - Overall, this is low (at least compared to other sectors). Pockets of innovation and engagement exist in and beyond the digital humanities, and among researchers at all career stages. There is scope for a substantial investment in signposting, supporting and promoting training opportunities, providing a creative environment for skills development that mainstreams digital skills alongside others. Promoting digital skills alone may not attract those who do not see the relevance to their research practice.

> **Communities of practice** - These are a rich source of knowledge exchange and informal and flexible learning for arts and humanities researchers. Supporting and encouraging researchers and technical professionals to develop and use those communities and networks through cross-disciplinary research and knowledge exchange, online and at conferences, will be an important way to ensure the development of the community and the cross-fertilisation of innovation in research practice.

> **Formal and informal learning interdependencies** - Formal training (internal and external) and informal forms of learning via self-teaching and peer teaching exist together, have interdependencies and often play different roles in a researcher's development journey. Both need to be encouraged and supported as part of a wider skills development strategy.

> **Formal training** - This has a key role to play in skills development for junior researchers, as entry level training to ensure a baseline of skills across the sector and to offer more advanced skills development in areas of demand. Understanding the time and workload constraints researchers face and their need for flexible courses with low upfront costs will facilitate take-up of training opportunities.

> **Data management and sustainability practices** - These are not yet fully embedded in research processes and there is a lack of knowledge and understanding of both the broad rationale for open research and the institutional infrastructure that supports it. In addition, this infrastructure may not be functioning effectively. There is a need to ensure clear guidance and widely available and up to date training.

> **Supporting research skills development** - Training grant schemes provide firm foundations for developing broad skill sets. However, job insecurity, intensification of workload, and institutional cultures related to the REF, challenge the skills development of both researchers and research technical professionals. There is a need for inclusive and explicit support for new research roles on projects, new forms of research practice and less traditional skills and knowledge (including digital) in funding opportunities and application processes.

01_

INTRODUCTION AND BACKGROUND

The arts and humanities (AH) research community is constituted by a diverse array of disciplines and fields that span archaeology and performance art, corpus linguistics and museum studies. The community's boundaries stretch beyond academics in traditional Higher Education Institutions (HEI), to include staff in Independent Research Organisations (IROs) - and the wider ecosystem, for example Galleries, Libraries, Archives and Museums (GLAMs) and creative industries. The data created and analysed by arts and humanities researchers are extraordinarily rich and varied, encompassing images, interview transcripts, databases, performances, document archives, specimens, texts and material artefacts. Whilst much research practice in the community continues to be primarily 'analogue' and theoretical, digital tools and technologies are revolutionising many aspects of it. The digitisation of existing data and the creation of new (born) digital data means software, programming, web-based platforms and other digital tools are increasingly utilised at every stage in the research process. These tools facilitate and enable the collection of data in its myriad forms and the subsequent analysis (data manipulation). They also provide the digital infrastructure that underpinning the management and sustaining of that data to ensure their use for future generations.

This ongoing digital transformation of research practice is widespread across the research sector. In the arts and humanities community, as elsewhere, it has been a long process of change and development. A digital subgrouping of the humanities has existed since the 1940s and the emergence of humanities computing. The Department of Digital Humanities at King's College London[1] defined the remit of this field as 'researching digital culture and society, and in exploring the use of advanced technology-related methods in humanities research'. Digital humanities techniques involving metadata, text, images, audio and video and software,[2] have reinvigorated certain areas of study, and enabled increased collaboration via digitisation of, for example, large text corpuses or banks of images. Yet, despite the historical embedding of digital tools and techniques, researchers working in dedicated digital humanities roles or departments only constitute a small proportion of the arts and humanities community. The use of digital methods and tools is not confined to the digital humanities, but the picture of research practices across the sector as a whole is less clear. For the Arts and Humanities Research Council (AHRC), as a primary funder of arts and humanities research, it is crucial to understand the effects of this digital transformation on the community if they are to ensure that infrastructure is in place and skills development opportunities are available to support it and drive forward research and innovation.

The wider UK research policy context is important here. A number of key strategies and interventions to support research are already being implemented by UKRI and constituent research councils such as AHRC. The principle of open research - the sharing of research data and the use of open-source software is - for UKRI[3] a key priority and a central driver of digital infrastructure development. UKRI's *Guidance on Best Practice in the Management of Research,*[4] is derived from OECD[5] guidelines on the public funding of research. Drawing on open research principles, UKRI's digital research infrastructure strategy has identified key components of a digital infrastructure system that includes computational capabilities, data storage facilities, software, secure networks and people.[6] As part of this strategy, they have funded the first phase of a project to build **a national infrastructure for digital innovation and curation for arts and humanities (iDAH)** in the form of a digital skills pilots scheme. The upgrading of existing digital infrastructure for arts and humanities research included the **Capability for Collections Fund (CapCo)** in 2020, that enabled research organisations holding extraordinary collections to refurbish their research and development facilities. In July 2022 **the Creative Research Capability (CResCa)** funding opportunity was opened for applications and aimed to support the renewal and upgrade of facilities and equipment for creative and cultural research.

The digital transformation of research practice and infrastructure necessitates a focus on the skills and careers of researchers and others who contribute to research. Digital skills gaps have been identified among the UK's researcher community.[7] UKRI's 2019 *Infrastructure Opportunity Roadmap* report identifies investment in 'digital and analytical skills, as a key priority alongside data science,

---

1   https://www.kcl.ac.uk/ddh/about/about
2   Madsen, Christine and Hurst, Megan (2018). Requirements for a Sustainable Digital Humanities infrastructure at the University of Oxford
3   https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/
4   Guidance on Best Practice in the Management of Research Data
5   The OECD Principles and Guidelines for Access to Research Data from Public Funding
6   https://www.ukri.org/what-we-offer/creating-world-class-research-and-innovation-infrastructure/digital-research-infrastructure/
7   https://www.ukri.org/blog/attracting-and-training-talent-for-the-ai-enabled-economy/

AI, and software engineering'.[8] The DARE programme's 2021 *Data Research Infrastructure Landscape* report also highlights the need to support the career structures of individuals creating or engaging with digital research infrastructure.[9] A number of UK reports highlight the need for software skills and training specifically. The 2022 TALENT Commission suggests that 'the need for computational skills – including enhanced digital design, software engineering and skills in AI and machine learning – is widespread'.[10] The ESRC's data infrastructure strategy has a focus on skills and capacity for data use, enabling researchers to effectively utilise data in their research for public benefit.[11] The AHRCI is focused on supporting 'the training of the highly skilled multidisciplinary research workforce needed across academia and industry'.[12] One recent targeted intervention has been the release of their 'Embed digital skills' funding opportunity in May 2022, which encouraged the design and piloting of digital skills training programmes for the use of digital tools and methods in arts and humanities research. 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.[13]

New digital tools and forms of data and research practice require new forms of research relationship and collaboration. Research practices in the sector have traditionally been structured around a lone researcher model[14] with a single disciplinary focus. Supporting and encouraging interdisciplinary research by UKRI and the AHRC has promoted more diverse working practices and rising use of large cross-disciplinary research teams.[15] Increased digital connection facilitated or accelerated by the COVID-19 pandemic have created the possibility for research focused 'communities of practice', defined as domains 'for exchange and collaboration with others across the world'.[16] Equally, as digital tools become embedded in research, the Research Software Engineer (RSE), and Data Analyst have emerged as new and vital roles in research teams in both HEIs and IROs. RSEs may either operate as a central institutional resource drawn on by multiple departments, or be employed on short contracts to support a specific project. The Software Sustainability Institute was awarded funding by all research councils over the last four years[17] to support RSEs and promote open research as well as better, more sustainable, research software. Importantly, there is commitment from UKRI to actively ensure the inclusion of diverse voices and experiences in the development of research practice. The fundamental principles of equality, diversity and inclusion are guiding UKRI's vision for a national digital research infrastructure that is 'driven by the ambition of UKRI's diverse communities' [18]

Against the backdrop of this dynamic policy landscape, in which stakeholders have sought to manage the digital transformation of research practices across HE and industry more widely, there is an urgent need to understand the experiences, practices and needs of researchers as they adapt and develop their skills in the new terrain. The disciplinary diversity of the arts and humanities community is its strength, but it also poses challenges for providing infrastructure and support for skills development that can meet a wide range of research needs and interests. A long-term ambition of the AHRC, as per its 2019 Delivery Plan has been to 'address the shortage of skills in digital asset management and data manipulation as part of [their] integrated programme of infrastructure development'.[19] Recent developments make this even more pressing. Therefore, the AHRC commissioned researchers at the universities of Southampton and Oxford to undertake a study on the digital and data research practices and skills of the arts and humanities research community. The goal was to understand how the AHRC could best support and intervene to build the capacity of its research community to utilise digital tools and infrastructure.

8   https://www.ukri.org/wp-content/uploads/2020/10/UKRI-201020-UKinfrastructure-opportunities-to-grow-our-
     capacity-FINAL.pdf
9   https://zenodo.org/record/5584696#.YuZrxHbMJro
10   MI Talent, UKRI, & Research England. (2022). The TALENT Commission: Technical skills, roles and careers in UK
     higher education and research.
11   https://www.ukri.org/wp-content/uploads/2022/06/ESRC-090622-DataInfrastructureStrategy2022To2027.pdf
12   https://www.ukri.org/wp-content/uploads/2020/09/AHRC-250920-DeliveryPlan2019.pdf
13   https://www.ukri.org/opportunity/embed-digital-skills-in-arts-and-humanities-research/
14   https://insights.uksg.org/articles/10.1629/uksg.280/print/
15   https://authorservices.taylorandfrancis.com/wp-content/uploads/2017/09/Coauthorship-white-paper.pdf
16   https://www.heacademy.ac.uk/individuals/communities-of-practice
17   https://www.software.ac.uk/about
18   https://www.ukri.org/what-we-offer/creating-world-class-research-and-innovation-infrastructure/digital-research-
     infrastructure/#:~:text=UK%20Research%20and%20Innovation's%20(UKRI)%20vision&text=be%20driven%20by%20
     the%20ambition,catalyse%20breakthroughs%20and%20accelerate%20innovation
19   https://www.ukri.org/wp-content/uploads/2020/09/AHRC-250920-DeliveryPlan2019.pdf

The study aims were:

> To map the diversity of the UK arts and humanities research community and its constituent groups.

> To map existing policies and mechanisms provided by UKRI and other funders and stakeholders to support and promote digital skills in the arts and humanities research.

> To undertake a large-scale survey of the arts and humanities research community to establish practices and skills in the use of digital tools and software for digital manipulation, management and sustainability.

> To undertake in-depth qualitative interviews with arts and humanities researchers to understand the barriers and facilitators, and to identify potential interventions to support the development of digital skills for digital manipulation, management and sustainability.

> To draw on the results to produce recommendations that will help the AHRC design and deliver effective future interventions as part of their Infrastructure and Skills portfolios.

02_

METHODOLOGY

Our study involved an iterative three stage, mixed methods approach. Stage 1 consisted of a mapping of the arts and humanities research community, to establish constituent groups and ensure we were aiming for the broadest possible sample for Stage 2. As part of the mapping, exploratory interviews were conducted with key informants and researchers to facilitate survey design at stage 2. A review of data and archiving policies across UKRI and higher education was conducted to provide context. Stage 2 consisted of a large-scale survey of current research practices in the arts and humanities research community that touched on data collections, data analysis and data management, institutional support and research skills. The survey was focused in scope to maximise response across a broad range of constituencies. Stage 3 consisted of in-depth semi structured interviews with survey participants who had 'opted in' to further research. It focused on understanding research practices, experiences and cultures, including barriers and facilitators to skills development, forms of institutional support and future policy scenarios.

2.1 ETHICS

Ethics approval was sought from Southampton's research ethics committee (ERGO2). A staged application process was adopted, meaning the application was updated and resubmitted at each stage of the data collection process to cover each element: the mapping, the exploratory interviews, the survey and the in-depth qualitative interviews. This supported the iterative study design, enabling flexibility and responsiveness to issues as they arose. Informed consent was embedded at the start of the survey in the form of an information page that took clicking on continue as signalling consent (see Appendix 1, Survey Ethics). A participant information sheet (see Appendix 2, Interview Ethics) and consent forms (see Appendix 2, Interview Ethics) were sent out to interview participants in advance of the interview. As a way to encourage prompt participation and thank participants for their time and insight, we offered £30 shopping vouchers to interviewees, and a prize draw of £100 for survey participants. We recorded our interviews, transcribed them, pseudonymised the transcripts and then destroyed the recording.

As the subject of our study is essentially understanding best practice in using data and software in research, it was vital that we 'walked the walk' within the study itself. We committed to following the principles that would enable our data to be reused where possible. As part of this, we,

> Recorded informed consent for interviews: for data to be used or reused, it is vital that the informed consent can always be located with the interview. At the start of the interviews (see Appendix 4, Interview Topic Guide), we switched on recording and gained verbal consent for each statement which was recorded. The anonymised consent was then embedded in the transcript removing the need for separate secure storage of anonymised personal data. It is also thus permanently co-located with the data.

> Future proofed our data: we included specific consent for the storage and reuse of the data using the following wording (see Appendix 2, Interview Ethics), 'Now we ask you to state that you give permission for the pseudonymised transcript of this interview to be deposited in the researcher's data space (UoS SharePoint) and repository (Pure) as described in the Participant Information Sheet so it can be used for future ethically approved research and learning on the arts and humanities research community'.

> Created a data management plan (DMP): This provided a way to keep track of the data sets we created, clarify how they would be treated and stored according to funder requirements and audit what we had done. Our plan covered: the project overview; the purpose of the DMP; a description of the legal basis for processing personal data; a full list of the datasets collected with their name, type, origin, scale, purpose, documentation and location (see Appendix 5, Data Management Plan).

> Followed a risk mitigation strategy for General Data Protection Regulation (GDPR):
> > Our survey was completely anonymous. Respondents were able to volunteer their contact details if they wished to be considered for follow up interviews or for the prize draw.
> > This data is therefore personally identifiable, and in line with what is stated in our ethics form it will be destroyed at the end of the project.
> > Our interview data was pseudonymised before use, including removing identifiable details from the transcripts.
> > Only fully anonymised transcripts will be retained for future use. Video of the interviews, the pseudonymisation key and any identifiable transcripts will be destroyed.

2.2 STAGE 1: MAPPING

The mapping stage took place over the first 5 months of the study and provided important foundational understanding for designing stages 2 and 3. It consisted of a number of elements.

2.2.1 Mapping the research community

A broad sketch of the community provided by the Higher Education Statistics Agency (HESA) data helped us to define key characteristics and subgroups that would be relevant to the survey such as discipline, job title, project role, career stage, institution type. These are outlined in detail in relation to our sample in the 'Results and findings', section. In addition we consulted with the AHRC and wider research networks, and undertook desk research to compile a database of 122 key contacts in the the arts and humanities research community. The aim of this exercise was partly to suggest potential participants for our exploratory interviews, and partly to provide a listing that we would use to distribute the survey. We were keen to ensure diversity of the survey sample and aimed to include researchers from as wide a possible range of characteristics as possible, particularly those from black and minority ethnic groups. To do this we sought out and listed arts and humanities networks and departments that were focused on these specific communities and disciplinary interests.

2.2.2 Reviewing data policies

The data management policies of the seven UK research councils which have policies that apply to all applicants were reviewed to gain a wider understanding of the landscape (see Table 1 below). Innovate UK and Research England vary in whether data management policies apply to their funding on a grant-by-grant basis. Policies were reviewed and compared to gain a greater understanding of their implications in terms of the skills and knowledge required to implement them.

Organisation	Policy	Guidance
AHRC	UKRI Open Research	DMP Outline UKRI Best Practice of Management of Research Data
BBSRC	Data Sharing Policy	In Policy Statement on Good Scientific Practice Resources for Data Sharing
ESRC	Research Data Policy	UK Data Service UK Data Archive Peer Reviewers Guidance
EPSRC	Policy Framework on Research Data	Clarification of ESPRC Expectations of Research Management
MRC	Data Sharing Policy	Guidance and Standards for data management, preservation etc DMP Guidance
NERC	Data Policy	Guidance for ODMP
STFC	Scientific Data Policy	Data Management Plan

Table 1. UKRI data policies and data management plan guidance overview (correct as at 12/21).

2.2.3 Exploratory interviews

A small number of individual and group interviews were conducted with 12 key informants and researchers. We sought to identify the variety of research practices and skills that exist in the community, understand the different language and terminology used to articulate them and obtain some preliminary policy solutions. This contextual fieldwork was designed to ensure robust design of our survey instrument and interview topic guide. The participants for these interviews were purposively selected and recruited from the mapping database with a view to capturing diversity of the community (see Table 2 below). The interviews were recorded and the transcripts imported into QSR NVIVO software for analysis.

Category	Number (12)
Female	8
Independent Research Organisation	3
Researcher with disability	3
Researchers from ethnic minorities	1
Non-academics	2
Digital humanities	4
Early career researchers	2

Table 2. Exploratory interview sample (n = more than 12 as interviewees fell into more than one category).

The interviews explored research practices, skills acquired, the meaning of data and the use of software. We asked about terminology used in various disciplines, to ensure that we worded our survey and interviews inclusively. For instance, a repository might also be referred to as an archive; what one discipline might understand as computational skills might be referred to by another as digital skills.

These exploratory interviews revealed the diverse understandings of the term 'data' within the community and suggested a spectrum of data and software engagement from digital humanities researchers at one end to theorists and interpretive researchers at the other. Participants raised the distinction between informal vs. formal training, and they talked about communities of practice and interdisciplinarity. A number of insights emerged in relation to research infrastructure and funding including: the patchy and uneven nature of institutional training for doctoral and career development provision; the lack of knowledge about where to access training and the need for digital training to be validated and accessible, perhaps taught by domain experts rather than technicians; the importance of conferences and seminars for methodological dissemination and for closer relations between IROs and HE to ensure dissemination.

The discussion informed the design of the survey instrument, shaping question wording and potential routing for different groups. It also provided insight into the practical questions of how much time people might invest in completing a survey, and how vouchers might work to encourage participation.

The discussion on policy and infrastructure informed the development of four scenarios: a data and software skills hub, knowledge exchange facilitators, data and software fellowships, data and software in funding infrastructure. These were used in the qualitative interviews in stage 3 as an elicitation device to explore different policy solutions.

2.3 STAGE 2: THE SURVEY

The survey involved three elements: development and testing, distribution, and data analysis. Given the risk of a low response rate, a robust testing stage enabled the development of a sensitive survey instrument that would engage researchers and map their current practices. Incentives were used to encourage participation. An opt-in question for further research was included to provide the basis for sampling the qualitative study.

2.3.1 Survey development and testing

The survey instrument (see Appendix 3, Survey Questions) was designed to identify key characteristics and practices of arts and humanities researchers. In broad terms, these were researchers' understanding of and use of data, their use of research software, their institutional context both within HEIs and IROs and in relation to their funders. Questions consisted of multiple choice, scale, and a limited number of free text responses. The aim was for the survey to take 10 minutes to complete in order to facilitate a high response rate, this required a limited number of questions and minimal use of free text responses. Participant demographics were also collected to allow investigation of the effect of different characteristics on software and data practices. The subject area and career stage of the participants are highly subjective fields which needed to be codified. Popular systems in widespread use were chosen for these demographics: for subject area the Joint Academic Coding System (JACS 3.0)[20] was used. These disciplines were chosen for the survey to better reflect the AHRC's discipline list[21] and because they provided more granularity than the HESA disciplines. For career stage we used the League of

---

20 https://www.hesa.ac.uk/support/documentation/jacs/jacs3-principal
21 https://www.ukri.org/wp-content/uploads/2021/08/AHRC-090622-ResearchFundingGuideV5.6June2022.pdf

European Research Universities (LERU)[22] system, as the HESA does not provide a definition for this.

Phase	Name	Description
1	Junior (JCR)	e.g., PhD candidate, Junior RSE
2	Early (ECR)	Research Assistant/Associate/Fellow, first grant holder, Lecturer, Research Software Engineer
3	Mid/Recognised (MCR)	e.g., Senior Lecturer, Reader, Senior Researcher, Senior Research Fellow, Senior RSE,
4	Established/Experienced/Senior (SNR)	e.g., Professor, Director, Professorial Fellow

Table 3. LERU career development categorisation.

The survey questions were subjected to cognitive testing, in which a volunteer from the arts and humanities community worked through the survey with a member of the study team. This allowed the study team to identify areas of the survey that required further revision. Final testing suggested the goal of a 10-minute response time was achieved. The finalised survey was hosted in LimeSurvey, which is a professional quality survey platform that provides flexibility in survey creation alongside robust data security.

2.3.2 Survey distribution

The collection of a statistically relevant sample across the arts and humanities research community was the main challenge for the survey. The sample had to be representative, including those people who do not use data and software in their research. The study team used the UKRI's Gateway to Research website to collect details of researchers based in UK HEIs and IROs who had been awarded funding from the AHRC over the previous five years. This process generated 5400 contacts from the broadest possible range of backgrounds, which formed the backbone of the contact strategy. It was evident that this group would be skewed towards successful grant holders and would miss doctoral students unless they were also employed on a project. A number of other strategies were employed to ensure the survey was widely distributed and captured doctoral students and those who were not necessarily grant holders. These included promotion via:

> the key informants database produced at Stage 1;
> the AHRC's newsletter; the AHRC's and Software Sustainability Institute's Twitter accounts;
> mailing lists for arts and humanities organisations,
> IROs, GLAMs, and individual contacts.

The invitation to complete the survey also asked recipients to send on the invitation to their own contacts in the hope of creating some viral marketing. The invitation emails and survey publicity were worded carefully to attract a response from researchers across the arts and humanities research community. The prominence of terms like 'digital' and 'software' were minimised to ensure the survey would appeal to researchers who were not using software or digital tools and might have excluded themselves from a study they assumed to be aimed at digital humanities.

2.3.3 Sample size, response rate and significance

Sample size for a survey is calculated as follows:

$$Sample\ size\ =\ \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \frac{z^2 \times p(1-p)}{e^2 N}}$$

where: N = population size, e = margin of error, z = z score.

The AHRC judged their community to include 40,000 researchers *(N = 40,000)*. The industry standard for a survey is to aim for a 95% confidence level *(generating a z = 1.96)* and a 5% margin of error *(e = 0.05)*. To achieve this level of significance, the survey had to collect a sample size of 381 responses.

The survey was open between 10 February and 8 March 2022. 335 responses were collected in total, but a reduced number of participants, 281, reached the end of the survey. We were still able to use the data that had been submitted. We take the lower of these two numbers to calculate significance. There is a highly non-linear relationship between significance and

---

22 https://www.leru.org/files/UK-England.pdf

sample size, which means that the 281 participants produce a 95% confidence level and a 6% margin of error for this survey.

2.4 STAGE 3: SKILLS AND POLICY INTERVIEWS

The in-depth qualitative interviews allowed us to explore the issues raised in the survey in more detail and to focus specifically on skills and policy which required more probing and context than was possible in the survey. 30 participants were purposively selected from the 100 survey respondents who completed the opt in details to give permission to contact them for further study. We used the information from the demographic survey responses to ensure we captured as diverse a range of disciplines, institutions etc. as possible in the final sample, although we were also constrained by which participants replied positively to our request for an interview. We were careful not to over-select digital humanities interviewees. As well as the LERU categories listed in the section above, some interviewees self-identified as outside of the academic hierarchy and we designated these interviewees as 'Professionals (PRO)'.

The topic guide (see Appendix 4, Interview Topic Guide) was designed to explore in more depth the central questions covered by the survey on current practices in data and software manipulation and management, and provided the space to focus on the skills and policy questions. The interviews explored how researchers identified and acquired their data and software training and skills. We asked questions about the participants' research experience of doctoral/professional training, post PhD/ professional training skills acquisition, 'missing' skills, perceptions of skills amongst peers, colleagues and research teams, experiences of institutional infrastructure that supported digital tools and skills development.

The interviews used policy scenarios developed during stage 1 as an elicitation device to explore participants' experiences and priorities and tease out their views on different dimensions of the institutional and individual practices around skills. Four broad scenarios were derived from the exploratory interviews in stage 1 that prompted responses on ways the AHRC might support and facilitate skills development in the future. These were: (1) a data and software skills hub, (2) knowledge exchange facilitators, (3) data and software fellowships; (4) data and software in funding infrastructure, including in doctoral training requirements. Each scenario had a skills dimension, which keyed into issues of informal/formal training, communities of practice and interdisciplinarity. They also enabled discussion of the possibilities for timely and scale-appropriate skills interventions and the infrastructure necessary to support them.

The interviews lasted between 45 minutes and 1 hour and 10 minutes. They were recorded on MS Teams, and transcribed primarily via Otter AI online voice to text software, with corrections and anonymisation by the researcher. Transcripts were imported into QSR NVIVO software for thematic analysis.

2.5 ANALYTICAL APPROACH

2.5.1 The survey

The raw survey data was used for the analysis in this report. An anonymised data set was created by removing email addresses and all free text fields from the raw data (see Appendix 6, Survey Results). To prevent deanonymisation, fields on gender, ethnicity and disability were removed from the data set, reordered, and provided as a separate dataset. Analysis was conducted using scripts written in Python.[23] Univariate analysis produced summary data for each question and multivariate analysis was conducted where it was important to understand whether factors, e.g., career stage or subject field, might affect the response. The univariate and multivariate analysis charts are reproduced in Appendix 6, Survey Results.

2.5.2 The interviews

Once the transcripts were cleaned, checked and pseudonymised, they were imported into NVivo. Attributes were applied to each transcript based on the participants' career stage. Thematic analysis was conducted using a broad coding frame devised from the key areas of the topic guide (see Appendix 7, Coding Frame). These top level codes included 'data' (research practice), 'software' (what and how used), 'skills' (all discussion on skills and knowledge used and acquired), current support (institutional context and infrastructure) and 'future scenarios' (material related to policy scenarios or other policy related discussion). Following a coding of all documents at this top level, each code was then coded to more refined subcodes. These included 'current scenarios', 'future scenarios', 'communities of practice', 'interdisciplinary working', 'sustainability', 'formal and informal skills acquisition', 'barriers and facilitators'.

---

23 https://github.com/Southampton-RSG/ahrc-survey-analysis

03_

RESULTS AND FINDINGS

This section presents the results of our analysis across the three stages of our research - mapping of the community and policies, the survey and the interviews. In section 3.1 we look at key characteristics of the research community and examine how our samples compare with the HESA data. We then explore the survey results and interview findings in three sections: section 3.2 describes the data and software practices of our survey respondents and interview participants and highlights the skills and knowledge they are using on a daily basis. In section 3.3 we examine what and how skills and knowledge have been acquired and sustained, and the facilitators and barriers to those processes. Finally, section 3.4 focuses on the respondents' experiences of the institutional context and infrastructure that supported their research and skills development.

3.1 THE ARTS AND HUMANITIES RESEARCH COMMUNITY

Data from the HESA was used to check the extent to which the survey had collected a representative sample of arts and humanities researchers. HESA collects data on HE staff in the UK, which is representative of the research community as a whole. It should be noted that not all arts and humanities researchers and grant awardees are working in HEIs. However, HEIs represent such a large majority of them that this data still provides a useful comparison for our sample. The following HESA data is representative of staff at work during 2019/20, which was the latest data available at the time of performing the analysis (January 2022).

HESA states that the arts and humanities research community consists of 18,385 researchers across ten broad disciplinary groupings of which the largest is Art and Design.
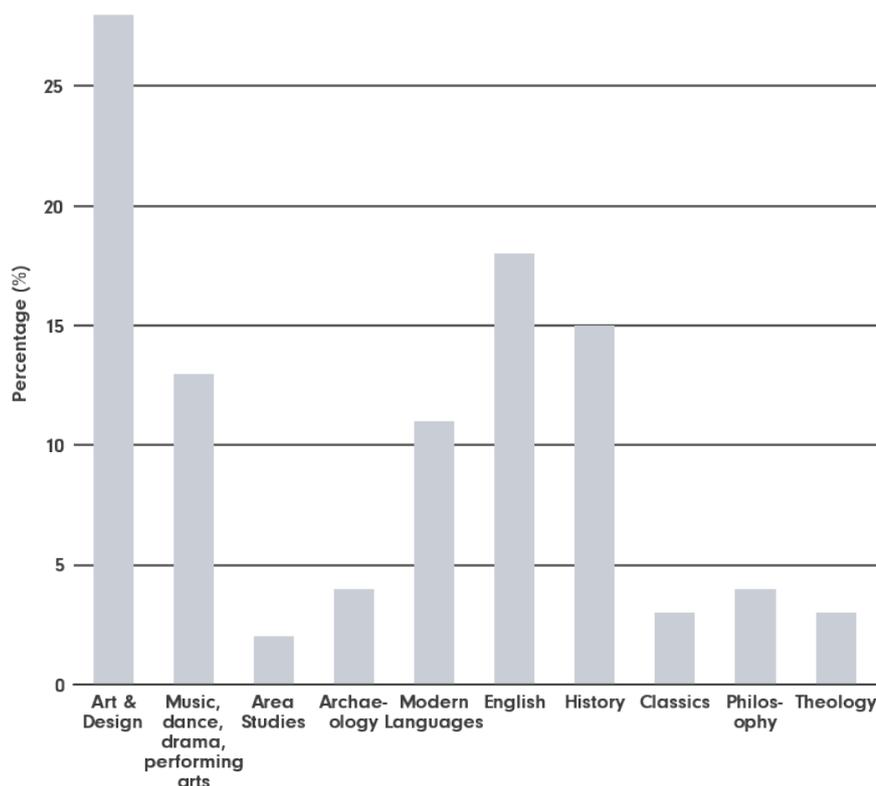
Figure 1. Researchers by discipline in arts and humanities 2019/20 (HESA).

We further examined key demographics across gender, ethnicity, disability and career stage. According to the HESA, 48.5% of arts and humanities researchers were female, 51.4% were male and the remainder identified as other.
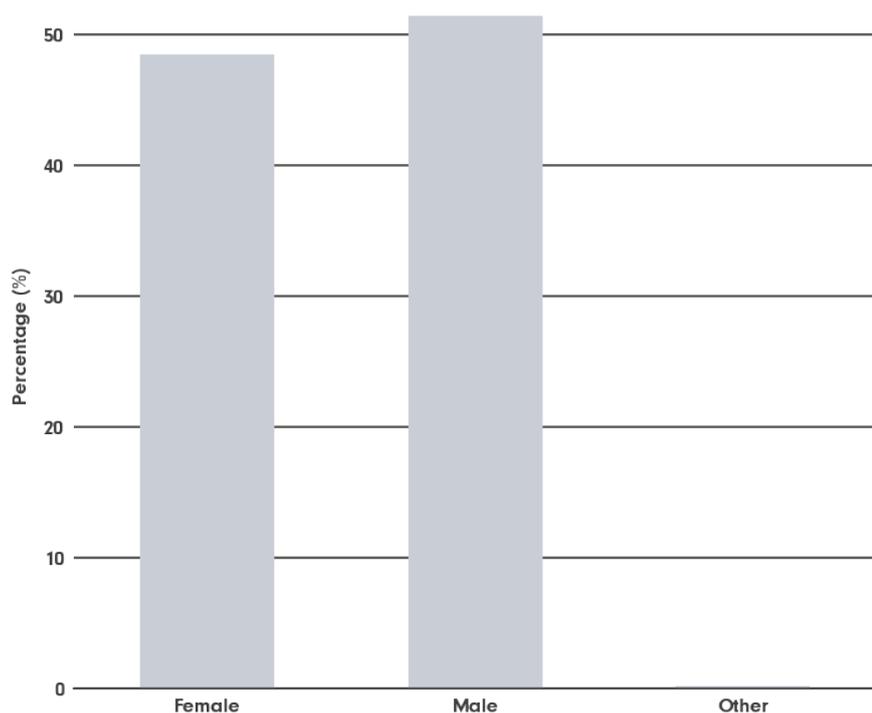
Figure 2. Researchers by gender in arts and humanities 2019/20 (HESA)

The research community is overwhelmingly white, with the largest minority demographic being Asian. There is also a substantial unknown section of the population.
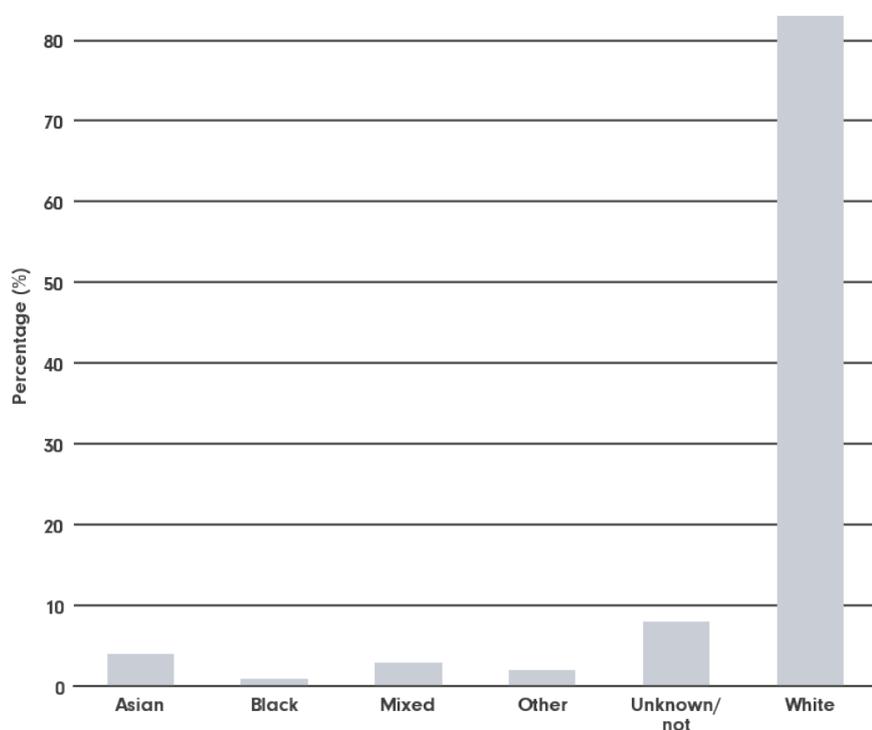
Figure 3. Researchers by ethnicity in arts and humanities 2019/20 (HESA).

The HESA data shows 5% of the community is registered disabled.
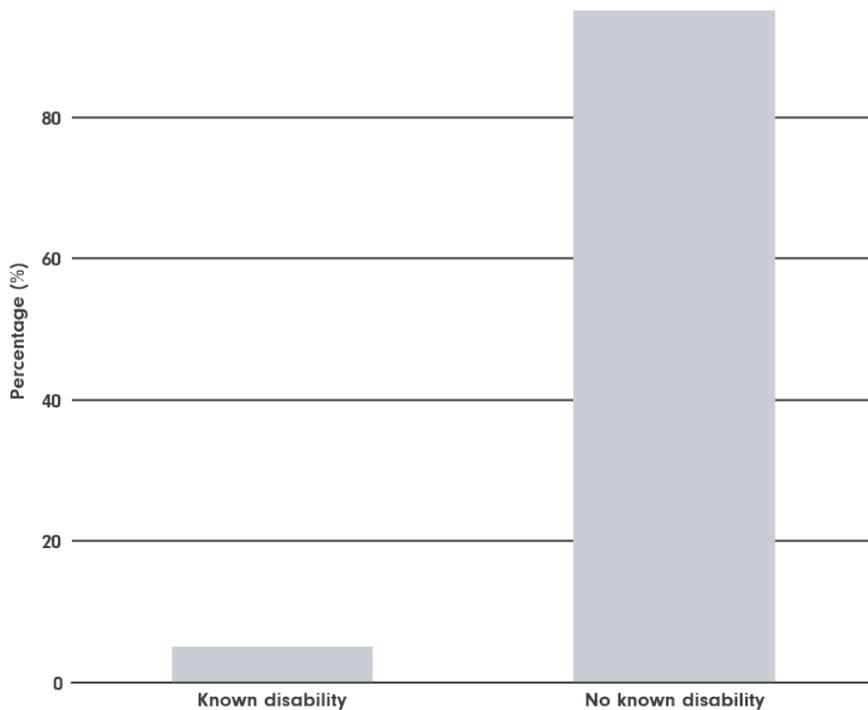
Figure 4. Researchers by disability in arts and humanities 2019/20 (HESA).

We used 'age' as a (limited) proxy for career stage as career stage data is central to our sample but is not captured by the HESA. We can see here that from approximately 36 - 55 there is consistency of numbers in each age group. There are (expectedly) fewer researchers in the younger age groups because HESA's data includes staff only, not post graduate researchers, although we included these in our study.
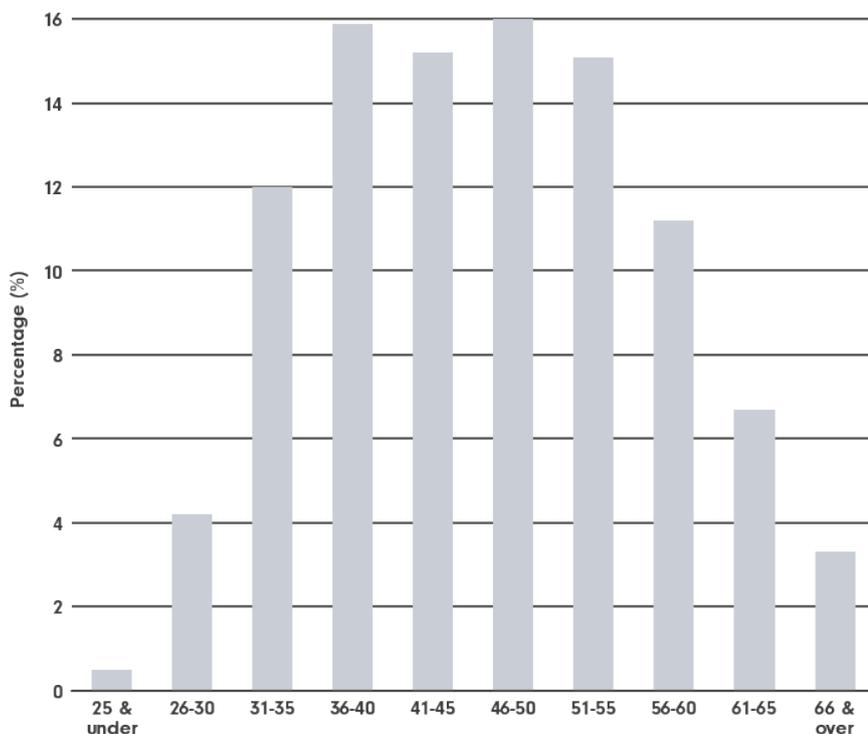
Figure 5. Age of researchers in arts and humanities 2019/20 (HESA).

3.1.1 Survey: sample and characteristics

Our survey sample outlined in Table 4 was able to reach researchers in all the demographic categories identified in the HESA data. Several points are worthy of mention. Firstly, we captured researchers in all 21 disciplines based on JACS categories used in the survey. Secondly, we attracted many postgraduate researchers, they were our largest career stage grouping. This may be because the incentive appealed to them disproportionately, and they had the time and motivation to complete the survey. Thirdly, our sample of people with a disability is slightly higher than that of HESA, as we invited respondents to state if they identified themselves as disabled rather than if they had a known disability. Fourthly, our efforts to include diversity of participants in terms of ethnicity were only partially successful. 91% of our survey sample and 90% of our interview sample identified as white, against 83.3% of the research population according to HESA. Finally, we were able to reach a variety of different institutional contexts from IROs, and professional organisations, to Russell group and post 1992 universities.

Category	Sample 334 participants
Discipline	38 Archaeology; 14 Classics; 35 Cultural and Museum Studies; 5 Dance; 24 Design; 9 Development Studies; 54 Digital Humanities; 9 Divinity and Religion; 12 Drama and Theatre Studies; 94 History; 23 Information and Communication Technologies; 59 Languages and Literature; 5 Law and Legal Studies; 20 Library and Information Studies; 23 Linguistics; 31 Media; 35 Music and Visual Arts; 19 Philosophy; 5 Political Science and International Studies; 9 Theology; 64 Other[24]
Career stage	37.5% Junior; 13% Early; 17% Mid/established; 33% Senior.[25]
Ethnicity	62% White: English / Welsh / Scottish / Northern Irish / British; 24% White: Any other White; 2.5% White: Irish; 2.1% Asian / Asian British; 1.7% Mixed / Multiple Ethnic groups; 4.3% Other ethnic groups including Arab, Chinese, Gypsy, African, Black, Caribbean; 2.8% Do not wish to declare.[26]
Disability (self-identified)	85% No; 11% Yes; 4% Prefer not to answer; <1% Unknown.[27]
Gender	36% Male; 60% Female; 2% Prefer not to answer; 1% Prefer to self-describe.[28]
Institution	94 (91.2%) Academic; 9 (8.8%) IROs.

Table 4. Summary of demographic characteristics of survey participants.

---

24 Respondents were able to select as many disciplines as they felt applied to their particular circumstance, therefore it sums to more than 334.

25 Not all respondents answered this question. Therefore, it does not sum to 334. In the survey we included the LERU categories so only respondents who fell within these categories would have selected them. The descriptions of each of these categories can be found in Table 3. Percentages have been rounded.

26 We had multiple categories for ethnicity. However, the number of interviewees in all the other categories except the White categories were too few to allocate them each their own category without risk of identification, so we have aggregated them. We have also aggregated the White categories for simplicity. Not everyone answered this question.

27 Not all respondents answered this question.

28 Not all respondents answered this question. Percentages are rounded.

Discipline

In response to the question 'In which disciplines do you work?' [Q3, n334] researchers could identify themselves in any number of the 21 categories that were supplied rather than a single discipline. This provided a way to demonstrate disciplinary subgroupings and accommodate interdisciplinarity. 180 respondents chose a single discipline, and 118 respondents chose more than one discipline. Almost 20% identified (one of) their discipline(s) as 'other', i.e., not represented by the JACS list and wrote their other disciplines in a free text box. These self-defined disciplines included computer science, maths, physics, chemistry, anthropology, archival science and heritage science, demonstrating the breadth of the interdisciplinarity both within the broad arts and humanities community and beyond it.
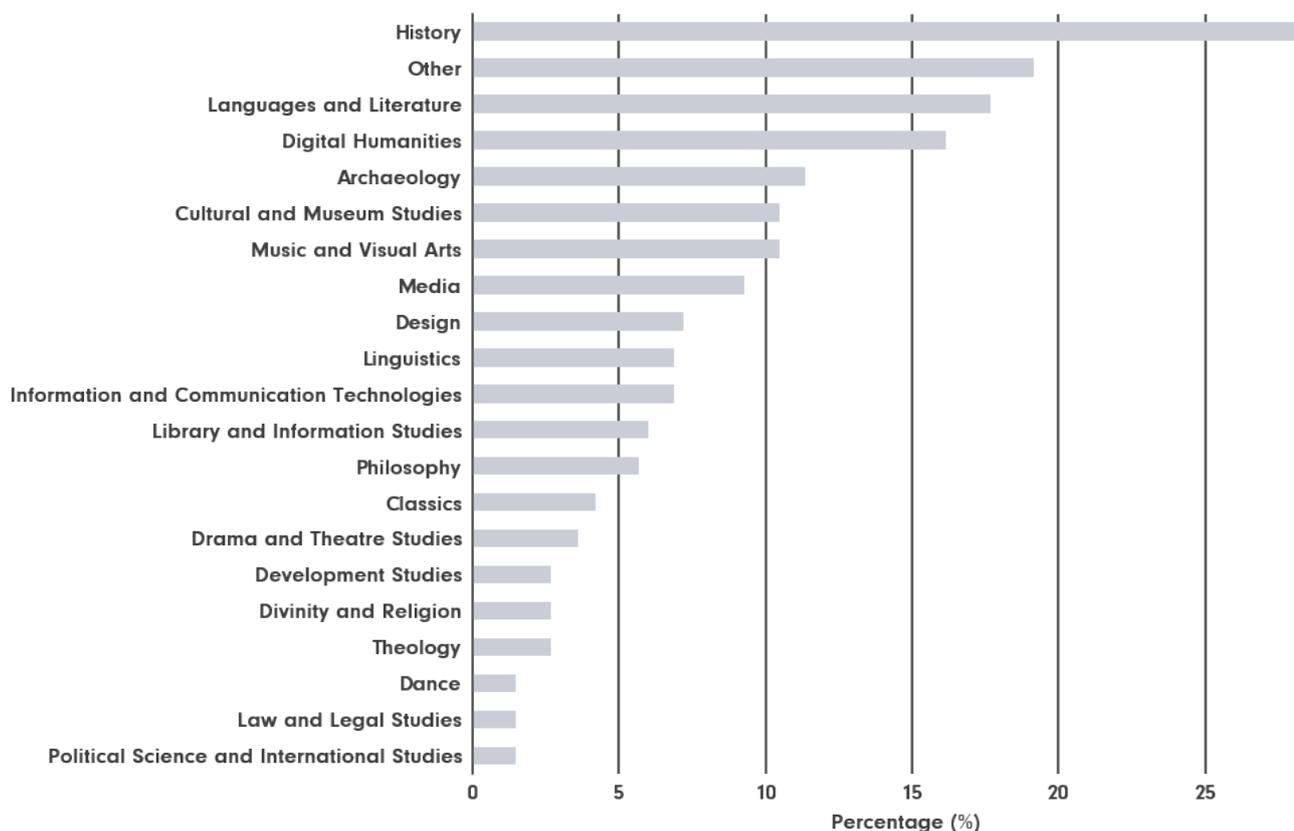
Figure 6. Discipline of survey respondents - JACS categories.

The digital humanities as a disciplinary category was specified by around 16% of the sample. This group is not measured by the HESA criteria so we cannot compare this response against the arts and humanities community. It is possible that there is a slightly higher proportion in our sample than in the community as a whole. In addition, whilst digital humanities constitute a specific sub-group within the arts and humanities community, researchers usually also have a disciplinary home such as history or linguistics. Given their different research practices it was important to be able to separate researchers who identified themselves as digital humanities from those who did not when undertaking analysis of the survey data.

Funder

The following chart [Q4, n334] shows which organisations are funding arts and humanities research. This further illustrates that we have reached the AHRC community, but also that the community is drawing on multiple funding sources and working on cross-disciplinary projects.
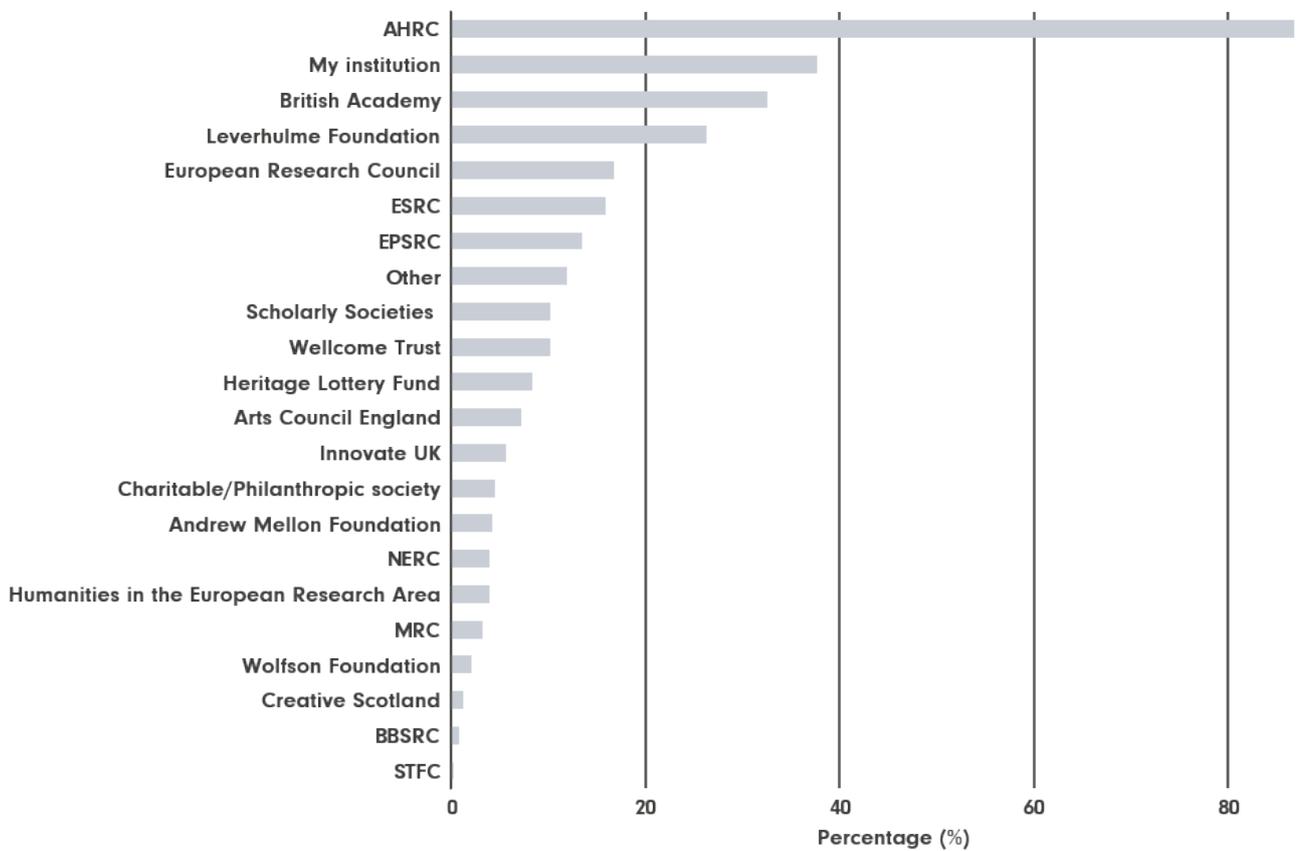
Figure 7. Funding organisations applied to by survey respondents.

3.1.2 In depth interviews: sample and characteristics

Recruitment for the in-depth interviews was based on respondents to the survey opting in. Around 100 responded positively to further contact. From that group we sought to purposively sample participants to achieve a balance of characteristics (demographic, career stage, discipline, institution etc) among our 30 participants (see Table 5 below). Our goal was an inclusive sample reflecting as much as possible the diversity in the survey respondents and wider arts and humanities community defined by the HESA. Overall, this was successful and our selection of interviewees captures key characteristics of the community as described by the HESA. It includes participants from 15 broad disciplines and we endeavoured to ensure digital humanities were not over-represented amongst interview respondents. It also includes a variety of professionals in IROs, and those in professional services/ infrastructure roles such as RSEs.

Category	Sample 30 participants
Discipline	Archaeology; Architecture; Cartography; Chemistry; Classics; Computer Science; Cultural Heritage; Data Science; Digital Humanities; History; Law; Linguistics; Media Studies; Theatre Studies; Theology.[29]
Career stage	30% Junior; 16.5% Early; 20% Mid; 20% Senior; 13.5% Professional.[30]
Ethnicity	90% White; 10% Other.[31]
Disability (self-identified)	7% Yes; 93% No.
Gender	47% Male; 53% Female/Non-binary.[32]
HEI	73% Academic; 27% Professional services and IROs.[33]

Table 5. Summary of demographic characteristics of interviewees.

3.2 DATA AND SOFTWARE PRACTICES

In this section, we look closely at research practices and skills in relation to data and software. We begin by exploring those concerned with data manipulation, i.e., data collection and analysis. We then examine how researchers use and develop software in their research. Finally, we explore data and software management and sustainability, including issues around storage and sharing.

3.2.1 Data manipulation: collection and analysis

In the survey we defined data as 'data of any type generated during the conducting of your research', in order to ensure a broad inclusive understanding. We asked survey respondents what types of data they or their team had worked with over the past five years and allowed them to select more than one [Q6, n=305]. The findings showed over 80% of researchers were working with textual data in some form, but also that a substantial number were using images (70%) and audio (45%). 42% were working with physical artefacts and 40% were working with numerical data.

---

29 We have omitted numbers or percentages to reduce risk of re-identification.

30 The category of 'professional' is for those interviewees who identify as being outside the LERU classification, e.g., library support staff.

31 We had multiple categories for ethnicity. However, the number of interviewees in all the other categories except the white categories were too few to allocate them each their own category without risk of identification, so we have aggregated them. We have also aggregated the white categories for simplicity.

32 We had too few interviewees identifying as non-binary to allocate them their own category without risk of identification. We have therefore combined them with the female category. We have done this as 'female' is the minority category according to the HESA data.

33 By 'professional services' we mean those interviewees who work in HEIs but are not academics. We interviewed too few for them to have their own category without risk of re identification, but their roles are substantially dissimilar to traditional HEI academic research so we have included them alongside IROs.
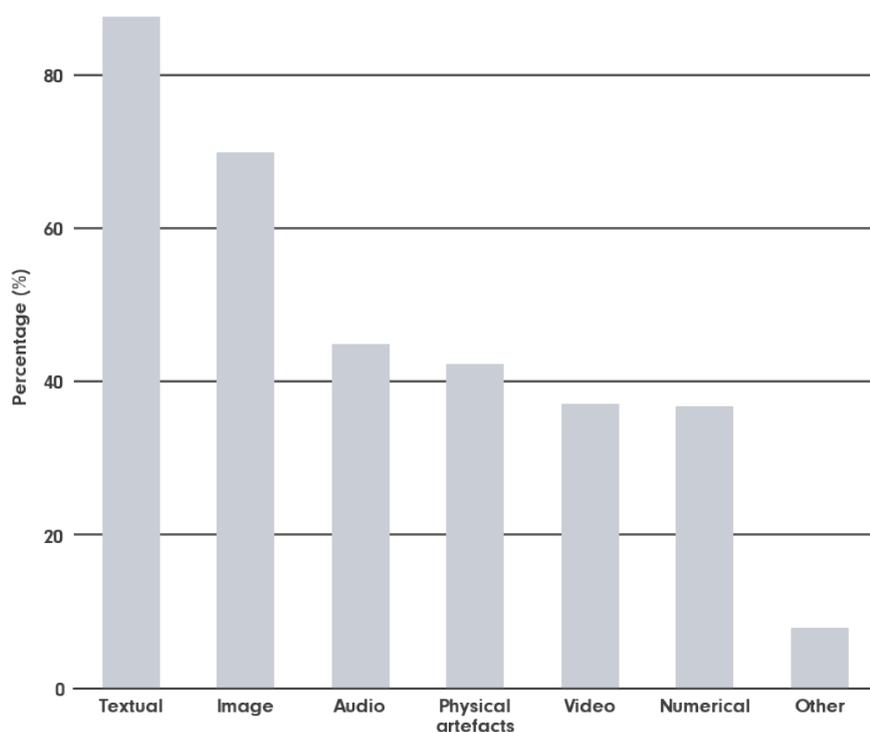
Figure 8. Types of research data used by arts and humanities researchers.

In the interviews, we invited participants to describe their data and software to us in their own terms. When asked broadly what data meant to them, interview participants were wide-ranging in their understanding. Whilst not all used the term 'data', or did not see the concept of data as relevant to their academic work, they were able to describe their research practices. One participant from an IRO noted that they encountered researchers who were working in digital areas, but *'they wouldn't know how to describe that, because it's not the language they can speak, they don't tend to talk in terms of data and software'* [483]. Most, however, were engaged with multiple forms of data and finding ways to innovate what data might look like in their research practice.

Types of data described included cultural heritage archives; 16th century literature and maps; cartoons; XML texts; descriptions of artefacts and people; images; observational and ethnographic data; eye movement tracking and biblical editions. Audio data included oral histories, recorded archives and interviews, and sound recordings. Researchers in disciplines such as archaeology, cultural heritage and design were working with physical artefacts that included labels, photographs and manuscripts, creating new knowledge and insight from digitising non digital data sources. Others were involved in the creation of digital data (born digital) in forms such as data in geographic information systems (GIS), survey sensor and imaging that required appropriate software tools. Some were working with sizable datasets. One project, which aimed to understand the materials used in mediaeval manuscripts, generated about ten gigabytes of data in two days. Another used a semi-structured corpus of about one million words written in the sixth century. Others described research projects involving complex, multimodal data from a range of different sources. Participants from the field of design and the built environment were working with architectural measurements, drawings, plans, building control records, photographs and video. In other cases, researchers were working with a single data type but stored in a vast number of repositories, subject to a wide range of issues regarding indexing, naming and cataloguing.

The emergence of new digital forms of data collection and analysis had led to changing research relationships and collaboration. For example, the digitisation of historical theological texts had enabled wider access to and analysis of content, requiring extensive teamwork in the process. Researchers described how working with extensive corpuses required an integrated team of cross-disciplinary researchers. A single academic would not be able to achieve this alone. *'The lone scholar model is absolutely dead in the water when it comes to much historical theological study. The [project], no lone scholar in the world could do what an integrated and inter communicated team of philologists Latinus, theologians, historians, mathematicians, physicists, psychologist engineers can do'* [183]. The building of cross-disciplinary teams for research projects, often drawing in developers, technicians, librarians, data analysts and social scientists, was challenging and involved project management and communication skills among team members. Many working in this context were attracted to the digital end of the humanities because it facilitated more collaborative work: *'This is really what turned me on to doing digital stuff, it's very collaborative in nature'* [235].

3.2.2 Using and developing software

In the survey we defined research software as 'any software used in your work to produce a result that you intend to appear in a publication or output'. 75% of respondents had used software of some type in their research. We asked respondents how they acquired the software used in their research over the last five years [Q12, n=300]. The most common engagement with software was to use resources provided by institutions (56%). Around 25% of researchers were using freely available (open-source) software. Smaller numbers of researchers (17% in total) were developing their own software or drawing on expertise to develop bespoke software for their research. Again, these numbers are lower than other communities or the research community as a whole. A 2014 UK survey of researchers across all disciplines from Russell Group universities found that 56% developed their own software[34].
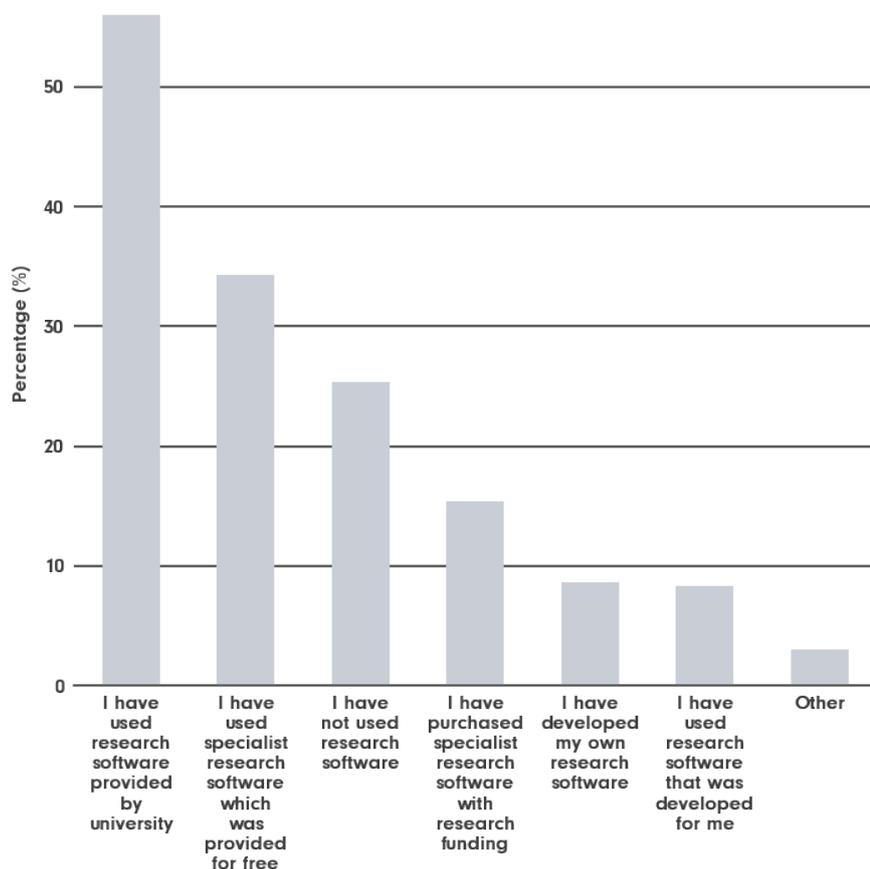
Figure 9. Research software practices across the past five years.

Our interviewees discussions help to shed light on the range of attitudes towards and engagement with software in the research context. At one end of the spectrum some participants articulated a deep engagement with a whole range of software, quite often accompanied by a sense of responsibility and excitement towards disseminating the possibilities of software to others. Many, but by no means all of these, were digital humanities researchers. Software was being used to undertake a range of different research processes in data collection and analysis, and included geographic information software; scanning, data science approaches such as natural language processing; programming languages such as R, Python and Matlab; transcription software; ontologies and linked data processes; XML, tagging, and Oxygen (an XML editor).

At the other end of the spectrum were researchers who engaged with email and word processing programmes, yet used very few other digital technologies. This group largely worked with theoretical or interpretative methods where the researcher's process was resistant to digitisation, and specific tools did not add value. In other cases, such as with thematic analysis, they felt none of the qualitative analysis systems improved on a highlighter pen. *'Text in its verbal forms, text in its printed forms but, equally, the understanding that text is socially constructed and I'm interpreting it...I'll also spend time in the environments that I'm researching and write my own reflections and notes on things and try to develop connections that way'* [711].

Between these extremes lay a swathe of arts and humanities researchers who were using or interested in software to differing degrees. Many of those we interviewed mentioned using or trying out qualitative data analysis programs such as NVivo. Other non-specifically academic software used included Access. *"I'm taking indexes and converting them into a*

Microsoft Access database, so the project then captures not just the transcriptions but lots of data about them." [235] The key barriers that emerged from the discussion appeared to be either a lack of appropriate software, *"I have thought of using machine learning, but most of the digital humanities options I've been given so far have been with words, and what I'm trying to identify is images and, it's a lot harder"* [228] or a lack of familiarity with software, *"They are, of course familiar to my colleague in digital humanities, and they're totally alien to me."* [379]. There were also respondents who were aware that processes could be carried out more effectively but appreciated the more manual process: *"it feels kind of handmade. And I know that really clever AI people can scrape directories and databases whereas we did an awful lot of data cleaning, and manual inputting. And I think it was good to do it that way. It feels that the way work is a little bit artisanal."* [171]

Using software was also made more challenging by the same issues that pertain to software use in all contexts, e.g., those of skills atrophy (if software isn't used for a while then it is easy to forget how to use it, see case study 2); software depredation (the software used may no longer be supported) and software upgrades (continuously having to learn new versions of a software). *"Even if we still use similar software, I would say the libraries in the module of this software are changing quite a lot. And we need to adapt to this relatively quickly."* [377] A researcher with an embedded RSE in their team told us how lucky they felt that they had a team member to deal with that. *"It was incredibly lucky that...we had a person just look after data standards, developed software and upgrades."* [124] There is also the challenge of having the right hardware to run the software on. *"If I had a slightly better computer, basically, there's a lot of stuff which just became, I can't wait a week for this model to render in."* [409]

We asked all our interviewees if they felt that their gender, ethnicity or any disability they considered themselves to have affected their use of data or software. The low numbers of interviewees from ethnicities other than white, or respondents with disabilities, meant that we were unable to establish any insights in this area, however, the interviews suggested that differences in engagement with software may be exacerbated by gender stereotypes in software development. *"As a female, early career researcher working between the divide between the digital and the humanities, I'm always feeling this kind of tension... let's be honest, it's populated with the white male stereotype"* [97]. A female professor also described how they were prevented from going to a technical workshop at a conference because the school manager felt it was *"too technical."* They noted that this seemed to be about sexism, and *"a presumption somehow that it's not in your domain"* [409].

We asked survey respondents who developed their bespoke project software [Q15, n=39]. Respondents could select all options that applied. The total response number for this question was 39, so slightly over 10% of the total respondents. Of those that responded, 66% said they used Research Software Engineers (RSEs), 51% developed their own software, 46% used doctoral and postdoctoral researchers and 28% used industry-based developers.

CASE STUDY 1
Data

Nic is both a postgraduate researcher and a practitioner of the art form that they are researching. They are looking at their art form as part of a broader cultural focus on the capabilities of the brain, with particular reference to additional technologies. Their data collection techniques appear natively digital, with almost a transhumanist element due to the devices employed, but the analysis techniques are deeply based in creative practice.

Nic's research methods are therefore derived from creative practice and resist traditional definitions of data. *'This is me putting this device on my head and observing it as a performer comparing it to theatrical performances that claim to do similar things now'.* Through the physical process of wearing these devices and experimenting with what they do, Nic then writes performance scripts. These scripts are in turn reflected upon by the researcher in a tradition of performance philosophy, where performance is research in itself.

Nic also aims to work with historical archival data from past performers on their art form, that comprises not only documents but other physical objects.

Therefore, there was clearly a tendency to use more than one source of software development.
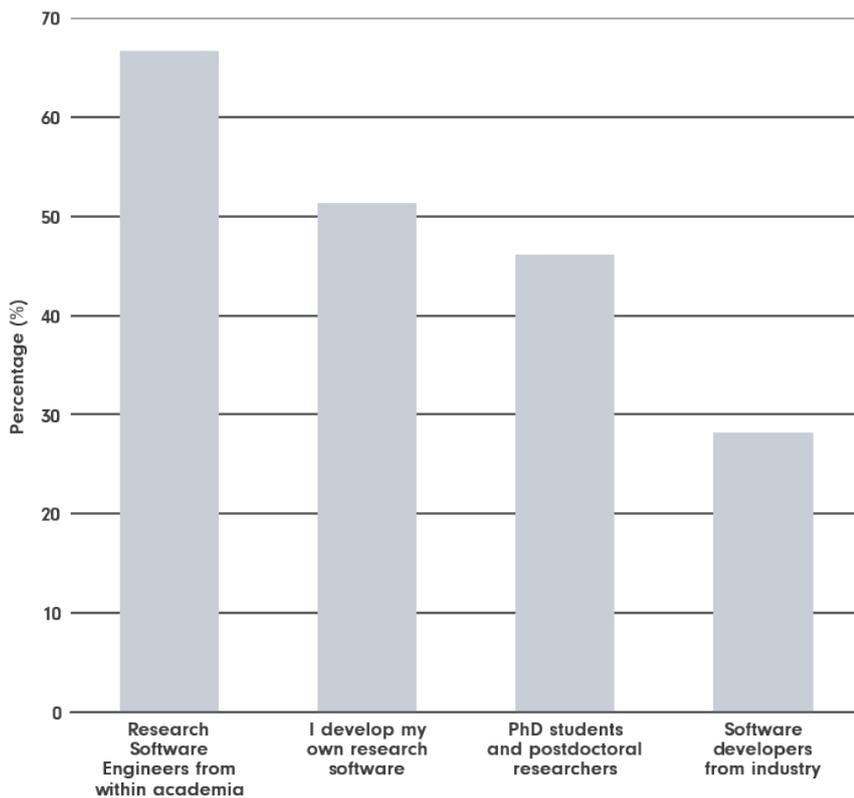
Figure 10. Who develops bespoke research software used in projects.

Interviewees also described how projects tended to use RSEs with software expertise to support their projects and develop bespoke software, rather than arts and humanities researchers using and developing it themselves. It was hard to find domain experts who were also software experts. Assembling teams with the right combination of skills was an ongoing challenge discussed at length in the interviews. One mid-career researcher who had been running a lengthy project described a catalogue of different staffing solutions they had used to bring developer expertise into the project over the past five years. *"Our current developer was somebody who actually did an academic doctorate at the same time as me and then did a conversion course, a masters in software development."* [124]. Relationships with RSEs and software experts were not always straightforward. Some reported confusion over responsibilities for different types of software in institutions, where bespoke software may be developed by RSEs but the IT department might be responsible for licences for proprietary software. *"So often it's not knowing, I'm sure we probably have some incredible programmes and platforms available."* [165].

We heard in the exploratory interviews that it was not easy to recruit RSEs, even when projects had planned for and costed them. The poor salaries and career progression for RSEs in the sector was percieved as an underlying reason for this. An academic who had moved roles to become a RSE described themselves as *"not having a career."* [147] While the scarcity of RSEs was not an issue currently for most participants, as they were not using them, the predominance of RSEs when bespoke software is developed, as is shown in Figure 10 above, means there is a potential longer term infrastructural problem as demand grows. One researcher who worked with an RSE remarked that *"What would potentially be helpful is.... If the AHRC could find some way of promoting the contribution [specifically] of RSEs, and not just call it digital humanities"* [124].

3.2.3 Data and software management and sustainability

In this section we examine how data and software is managed throughout the research process, and how it is sustained, i.e., both stored and shared at the end of a project. Sustainability can be seen as a function of both the infrastructural and technical capacity to share (including storage and methods) and the desire to do so. First, we look at the use of and perspectives on data management plans, which reveals researchers' familiarity with formal data management infrastructure and their compliance. Then we look at how, when and why data and software are stored and shared, and the barriers and facilitators of these practices.

Data Management Plans

Data management Plans (DMPs) are important because they are a formal requirement of UKRI's funding opportunities and involve researchers providing an account of what will happen to their data at each stage of the study. All the research councils, with the exception of the EPSRC, require a submission of a data management plan with a proposal. In our survey we asked about researchers' familiarity with DMPs. At the junior career stage, the majority (60%) of respondents had not engaged with a data management plan [Q10, n=305]. By mid-career, 57% of respondents reported that DMPs were mostly followed. However, as can be seen in the chart below, digital humanities scholars accounted for a larger proportion of those researchers familiar with DMPs from an early stage and following them closely at mid-career. Similarly, non-digital humanities scholars were less likely to have used a DMP.

Figure 11. Experience of DMPs by humanities/digital humanities researchers.

In the interviews DMPs tended to be viewed as a box ticking element of research bids rather than providing useful support, since there was no audit or review or compliance mechanism. Since the AHRC requires a DMP for applications, but does not prescribe what they should specify, researchers described a variety of approaches, *"I would say [uni research support] get a lot of inquiries from people who want to apply to AHRC, who don't think they have data or are not sure what their data is, and certainly wouldn't be sure what software they're using beyond Word and Excel."* [483] Some found them useful, but there were concerns that as plans were developed at the bidding stage in advance of actually acquiring datasets, they were unlikely to accurately reflect the practice and processes of data management used in the project. Some suggested there needed to be a way to ensure they were live documents and representative of what happens throughout a project and at its end. There was good practice in evidence, in terms of researchers managing their own data, but we heard about challenges created in using data from partners. *"With the third party information, we don't have the*

60%

of respondents at junior career stage had not engaged with a data management plan

57%

of respondents at mid-career stage reported that data management plans were mostly followed

copyright and the licensing sorted out, and while there is a recognition of working towards open access and open publications...those terms and conditions are not going to apply." [49] Others were concerned that in a competitive bidding process they could be used to undermine submissions: *"One of my colleagues put in a project where, you know, one of the reviewers basically focused their entire review on her data management plan and tearing that apart...and gave her a low score for the proposal."* [384]

Data and software sustainability

Data and software sustainability poses challenges with regard to unpicking the various practices, tools and attitudes that shape what happens at the end of a project. Several aspects of data sustainability were explored in the survey including data sharing frequency, data sharing methods and data storage locations. Together these responses were revealing of the amount of data and software that is lost and not made available for reuse. It was also interesting to look at these by digital and non-digital humanities, which revealed substantial differences in practice.

Regarding frequency, we presented respondents with a series of statements and asked which of them applied to their sharing of research data over the last five years [Q9, n=305]. The percentages are quite consistent across the statements, with 22% not sharing their research data and 27% saying they regularly share it. However, when we separate out the activities of the digital humanities researchers, 'I regularly share my research data' becomes the top choice by quite some distance (57%). They are more likely than others in the community to be engaged with routine data sharing practices. This may be because they have more experience of the digital infrastructure and their data is in digital form, which facilitates storage and sharing.

CASE STUDY 2
Software

Riley is completing their PhD. They have a vast amount of data, including census data from the UK data service, survey data collected online, audio recorded interviews, drawings, images of houses and photos of outputs from workshops, such as post-its, as well as miscellaneous data such as bylaws and regulations and fragments from social media.

They manage and analyse this data using a variety of software, including spreadsheets, a relational database and NVivo.

They analyse some of their survey data in SPSS, but then export this to Excel as this enables better creation of charts. *"And then some of this is put into Access, so I can combine it with other data formats as well. So even a lot of my [qualitative data] ended up in Excel, so I could quantify the main themes."*

They also want to share data spatially. *"The output that I've got from the database and spreadsheets I've linked into GIS, so I can share it on maps as well. So there's quite a lot of stuff."*

They also work with specialist software that is used in the practice of their discipline. At one point Riley started making a 3D model in Autodesk Revit, but whereas they were previously competent, lack of use has caused their specialist skills to atrophy, and it was taking too much time to use the software. *"I haven't [used Autodesk Revit] for a while now, so I've not completely lost it, but I'm not where I was before with it. And I guess CAD as well, because I'm just not doing it daily."*
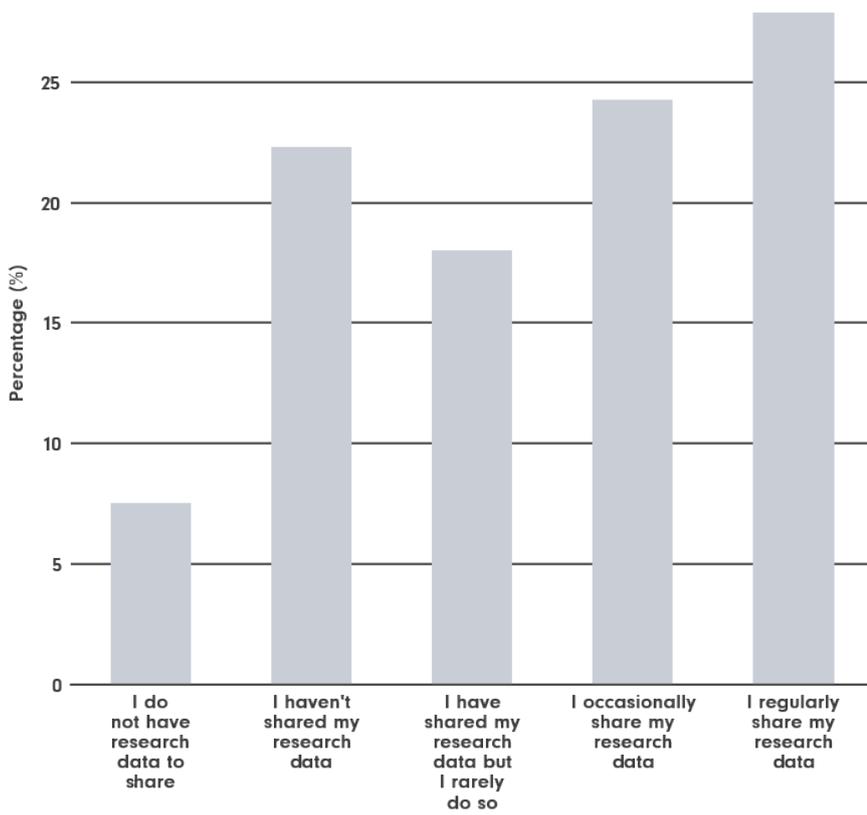
Figure 12. Frequency of data sharing - all respondents.

Digital Humanities
Not Digital Humanities
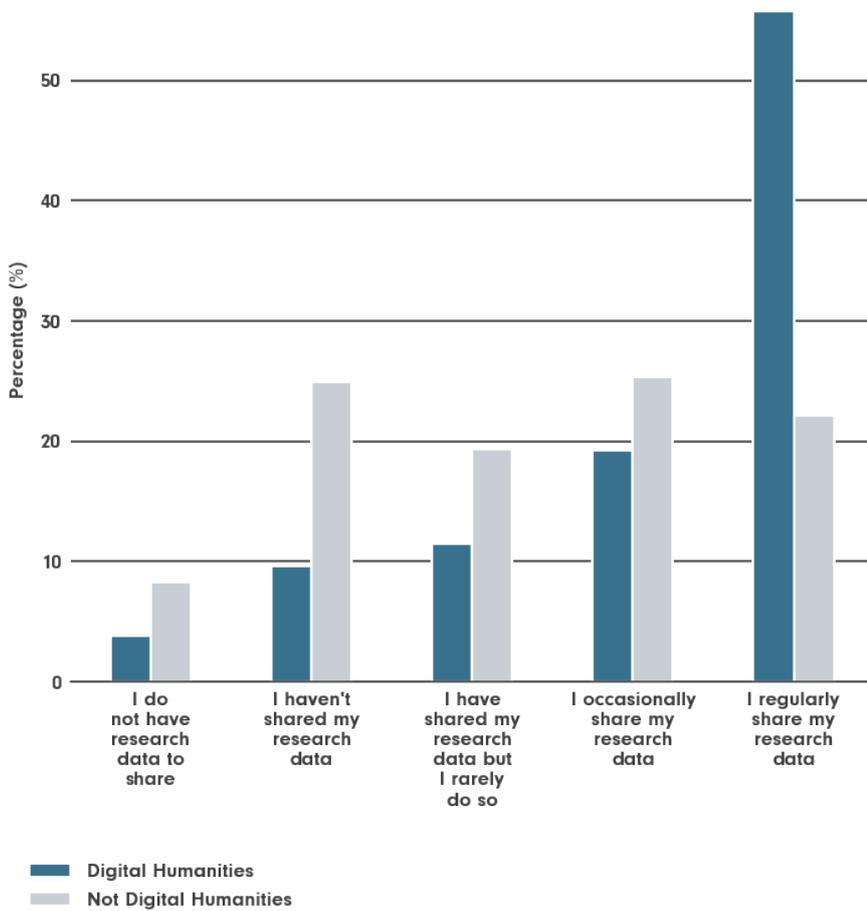
Figure 13. Frequency of data sharing by humanities/digital humanities researchers.

Data sharing methods

Survey respondents were more likely to licence their data to be shared and to create a Digital Object Identifier (DOI) for sharing [Q11, n=305]. The tendency to store data on a laptop or private repository appears to be in part a function of career stage.
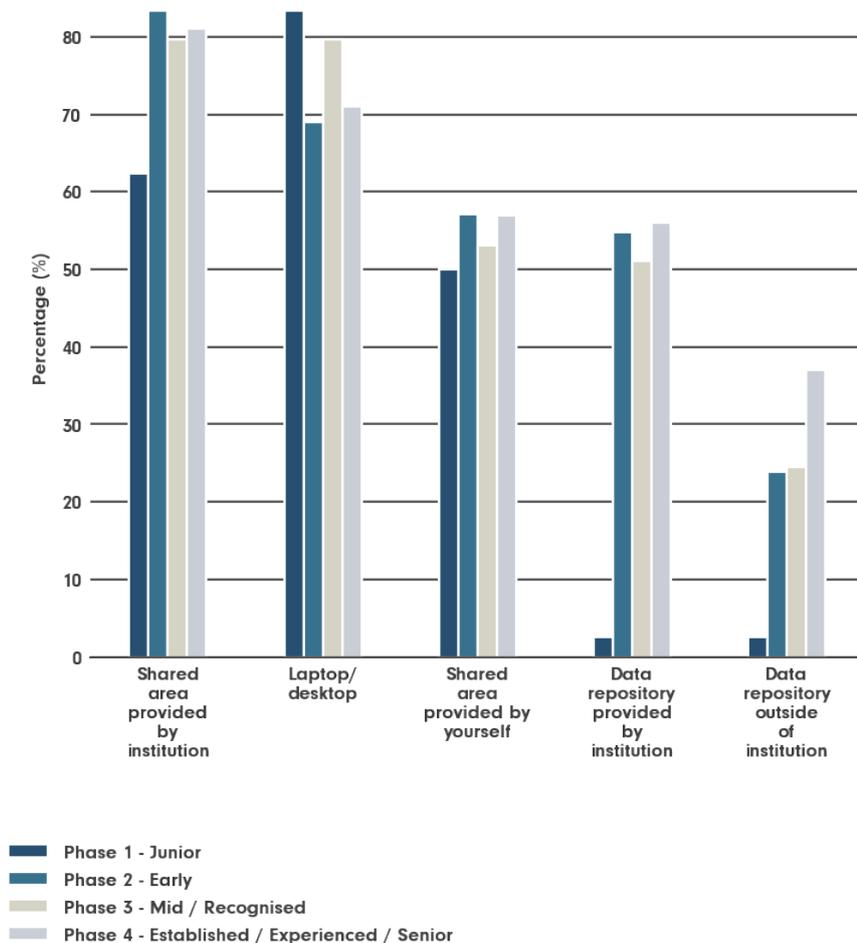
Figure 14. Where data is stored by career stage.

ECRs may be motivated by mobility and insecurity (there is no point in depositing data to a platform they may not have access to in the future), whilst PGRs are frequently not working as part of a larger team, and generally have not been in academia long enough to finish a project (when most depositing takes place). This mobility and potential movement out of academia complicates institutional ten year depositing requirements, and adds complexity to personal data protection legislation requirements such as subject access requests.

When we separate out the digital humanities researchers, we can see that the less 'technical' approaches of citing data in a publication and sharing data when requested are similar across both groups. When it comes to licensing and creating DOIs, digital humanities researchers are around 40% more likely to apply these to their data.
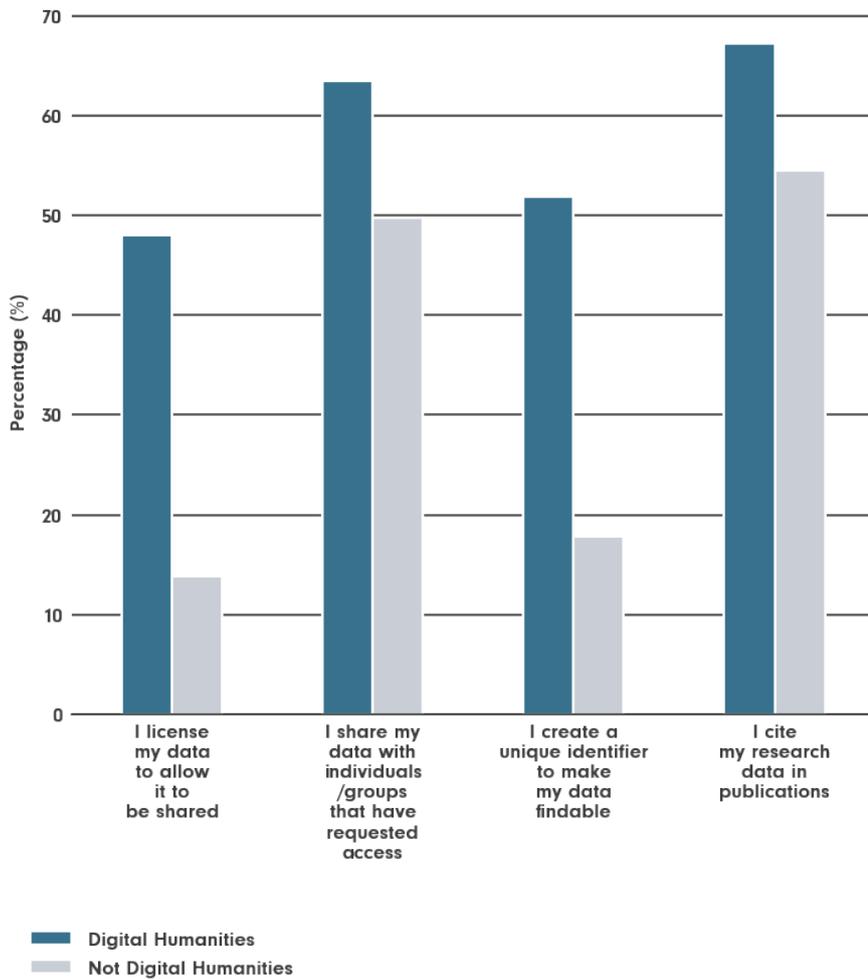
Figure 15. Aspects of data sharing by humanities/digital humanities researchers.

Data storage locations

Regarding data storage locations, we asked which best described the typical approach to storing data generated by a project [Q8, n=305]. Most used shared file space. Around 38% specified a shared area provided by their institution (Google Drive or SharePoint) and 17% specified a personal shared area. Together only around 12% were using data repositories (either institutional or external). More of a concern was 25% stored data on a laptop or computer, which suggested that this was not shared and could potentially be lost.
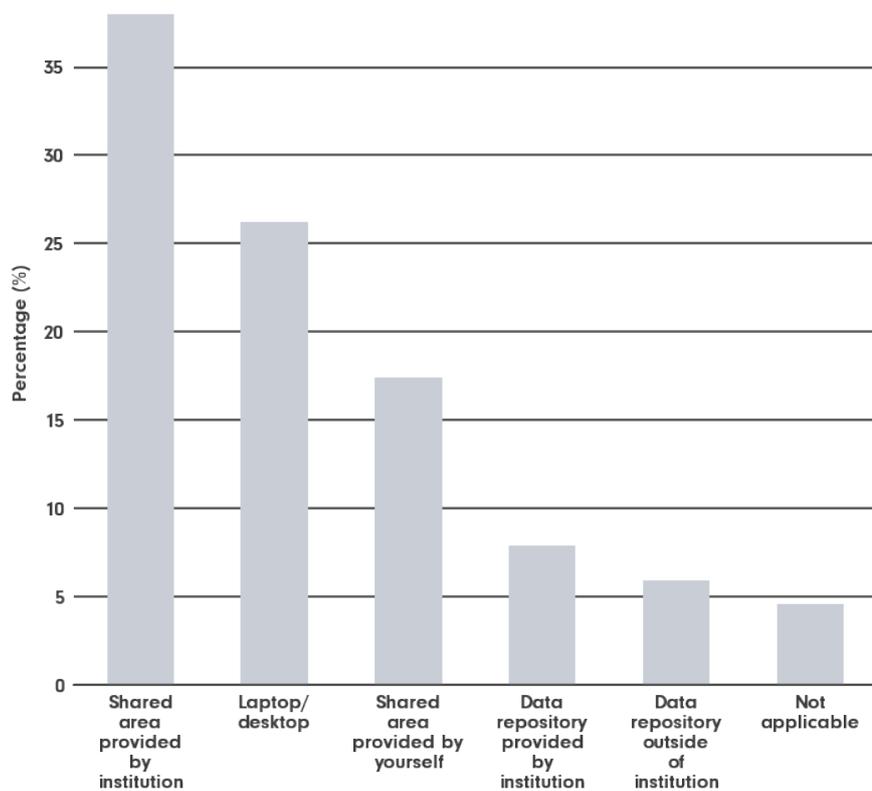
Figure 16. Data storage locations - all respondents.

Software sharing methods

We also asked which software sharing activities had been part of standard research practices over the past five years [Q25, n39]. Responses were low here since they only applied to those using bespoke software in their research. When examined by career stage, licensing software for sharing, sharing directly and citing in research publications were all practices undertaken by the majority of respondents apart from junior researchers. Junior researchers only shared with those who requested access which might be assumed to be supervisors and examiners. Less than 50% of the respondents in each career stage group created a DOI for their software.
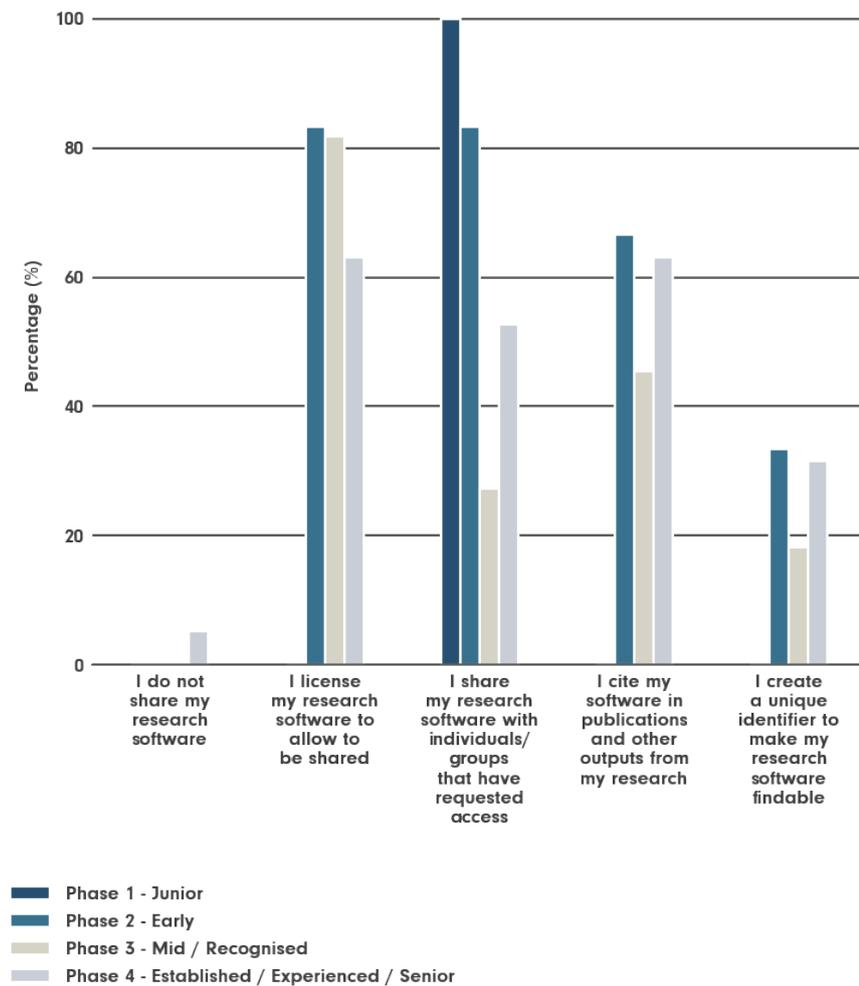
Figure 17. Methods of sharing bespoke software by career stage.

Discussion in the interviews covered practices and attitudes towards various aspects of data and software sustainability. Various official and less official storage locations were referred to, including Google Drive, SharePoint, Drop Box and GitHub. These discussions revealed something of the nature of the barriers and facilitators for researchers. Participants gave a number of reasons why they believed sustaining data was important. A key reason was that unintentional loss should be prevented, *"I mean, we have a really big problem with legacy information just being lost."* [403] Projects are being funded to recover poorly curated and unlinked data, with the implication there is value in linked and discoverable data. This value can sometimes be calculated: *"Another sort of back of the envelope calculation, there's about 17 million pounds worth of radiocarbon data that aren't being curated or archived."* [403] Some linguistic or archaeological data can only be collected at a specific point in time, so if it is lost, it is lost forever, *"because historically, people have produced these data. And they have been completely in lots of cases made rogue because they haven't been curated in any way. And in some cases are undiscoverable, completely undiscoverable."* [403]

There are also ethical reasons for reusing data once collected, especially where the data is expensive or complicated to collect. *"I think I live in a data graveyard. We produce, we produce, so much stuff we can't use. There's a huge ethical consideration about that, why do we keep producing so much data?"* [268]

However, participants also identified important reasons why they had not shared their data. Some felt that shared data was not used or useful, and the substantial work required to make it available was consequently a poor use of resources. *"99.99% of the data we acquire is not used, or at least it's not reported."* [268] Storage options were also seen as inappropriate for specific types of data. *"[university repository] is actually not very user friendly, and it's not very much set up for [arts discipline] type outputs, I struggled to figure out where I'm supposed to put those things in that system."* [165]. Funding was key to supporting the work to prepare data for a repository, including meeting standards. *"People put in data, if you don't control the CAD image, people, the amount of misspellings are unbelievable. Even a word like photographer can be put in so many different ways, badly. So again, this is the need to standardise."* [49] We also heard how multi-partner projects with large data outputs could become politicised where an important dataset becomes associated with one of the partners to the exclusion of the others. *"It becomes a bit of political issues, because if it's maintained by [university] there'll be [university] logos all over it."* [401]

Interviewees who were developing software for research noted the lack of guidance for how that software should be managed or shared. One respondent noted that they would find it useful to write a software management plan using a similar format to DMPs. Importantly, these would need to recognise where an algorithm is not intended for sharing or sustaining, as some are intended to degrade rather than be sustained indefinitely.

3.3 DIGITAL SKILLS DEVELOPMENT

In this section we look at how researchers acquired and developed their skills and knowledge over the course of their career, the way in which they gained new digital skills and engaged with software tools in particular. We included every stage in the research process from data collection and data analysis to data management and sustainability. The survey questions focused on types of data and software skills acquired in the past five years, the form their learning took, infrastructure, support and perceived barriers. The qualitative data then explored these questions in detail to understand the interplay of institutional and individual contexts that facilitated learning and created barriers.

3.3.1 Types of data and software skills learnt

In order to gain insight into the data processes that were being most used, we asked survey respondents whether over the past five years they had learnt new skills or knowledge in data collection, analysis, management/ storage or sharing [Q16, n=282]. A very large proportion, almost 70% of respondents, had acquired skills in data management and/ or data storage. This suggested that across the sector as a whole these skills were an increasingly important element of research practice for academics, and likely reflects the

focus on these aspects in personal data protection legislation such as GDPR and frameworks such as the REF. Just over 50% of the respondents had also learnt data analysis skills and just over 40% had acquired data collection and data sharing skills.
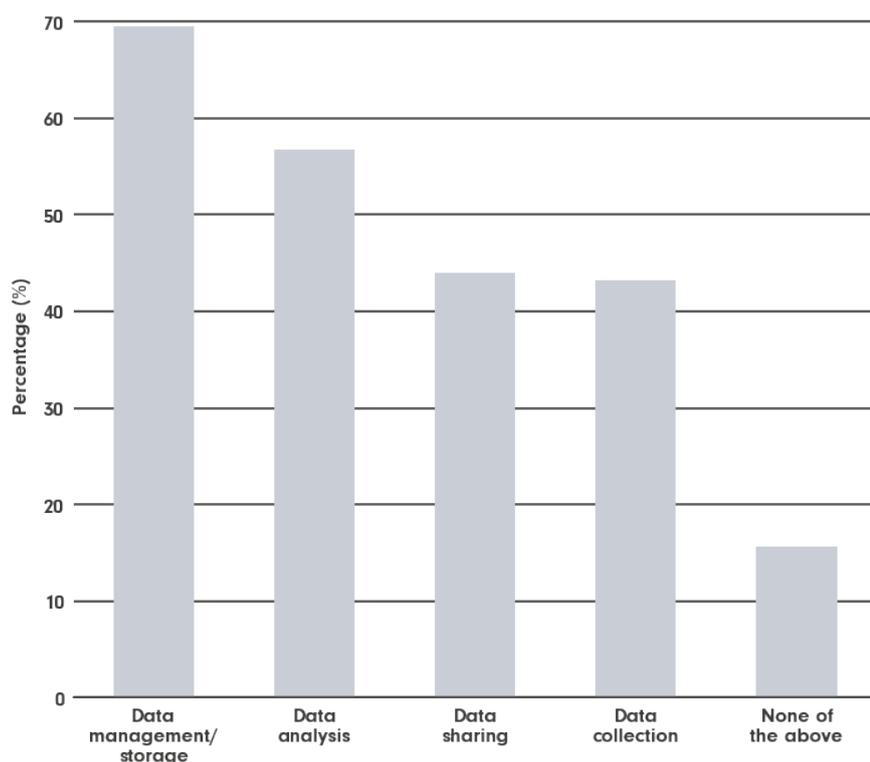
Figure 18. Data skills learnt over the past five years.

The qualitative discussions highlighted a range of skills our interviewees had acquired over the course of their career. Data manipulation (collection and analysis) techniques and skills mentioned included qualitative analysis software packages (NVivo and Atlas TI), corpus linguistic tools, imaging tools such as Adobe Lightroom, geographic information systems such as ArcGIS, Access databases, quantitative analysis software such as SSPS, and programming languages like Python. To a lesser extent, they also talked about acquiring data management skills using packages like Zotero and Mendeley, and techniques for project management such as Agile. In contrast to the high numbers in the survey response, few of the interviewees discussed learning about data sharing and data storage in any detail.

Returning to the survey and looking at the type of skills learnt by career phase (Figure 19 below), we can see that recent focus on developing skills and knowledge in data management and storage and sharing has been universal for most academics. Only junior researchers had acquired less data sharing skills, perhaps because they had not yet reached the 'end of project' stage that could encourage them to learn to share their research data. There may also be structural reasons related to the precarious and mobile nature of ECRs, who might not see value in making use of institutional repositories because they will have to move to a new institution when they find a job. If they continue to work on short term contracts, data sharing may not be a viable option. However, overall, for junior researchers the chart shows a more balanced pattern of skills acquisition across data management, data collection and analysis. This might be expected given the context provided by doctoral training grants where skills and knowledge development across the breadth of the research process are a key part of the structure.

Phase 1 - Junior
Phase 2 - Early
Phase 3 - Mid / Recognised
Phase 4 - Established / Experienced / Senior
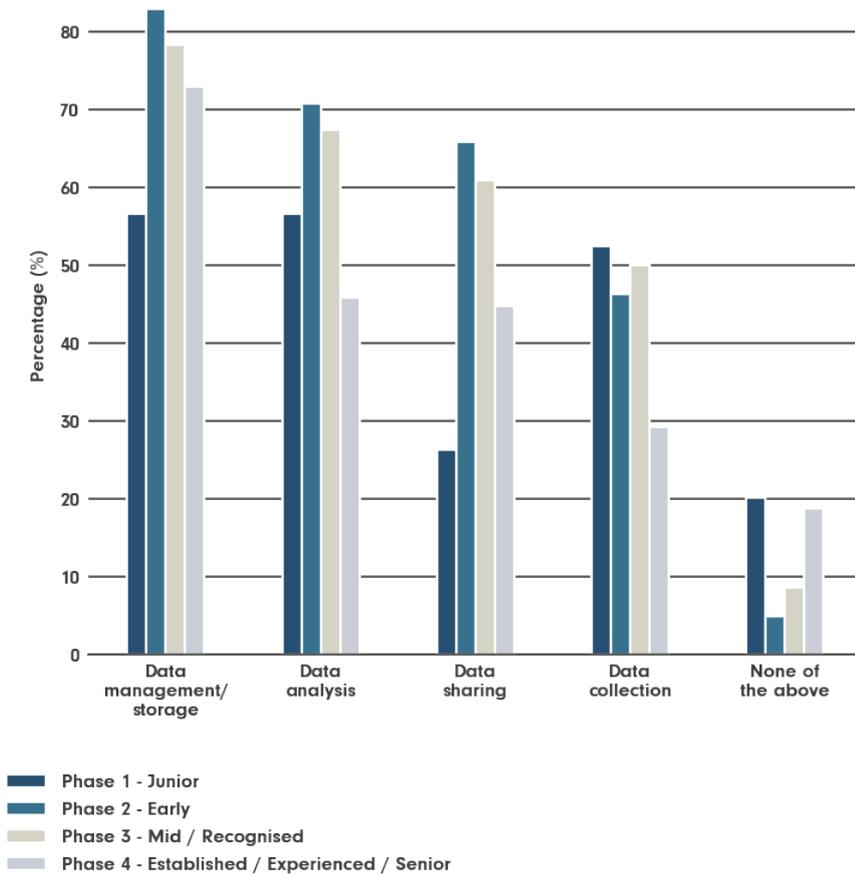
Figure 19. Skills acquired in the past five years by career stage.

The survey also asked respondents to think about their current skill set and idenitfy the main data or software skill that they would like to obtain to improve their research [Q19, n=272]. This was a free text question designed to elicit more individual insights into what researchers were looking for in developing their research skills in the future. Around 22 % of respondents suggested data analysis was their priority. Just under 10% said they were not interested in obtaining new skills. Small numbers of around 8% each identified variously software development/coding, data collection, data storage, data visualisation, bibliographic software and web design as skills that would improve their research. Around 26% identified a specific piece of software, including R, Python, NVivo and GIS, and a very long tail of software and programming languages that included TEI XML, Excel, SPSS, SQL, MATLAB, JavaScript.

3.3.2 How are data and software skills learnt

As important as what researchers learn, is how and where they learn it. Knowing how skills were acquired over the course of a career, the ways in which researchers develop their knowledge and expertise clarifies where intervention is possible and what policies can do to enhance and support existing mechanisms for skills development. We asked respondents where they acquired skills and knowledge in data and/or software [Q18, n=282]. They could select all options that applied. The chart shows that they had gained their research skills in a variety of ways, of which formal training courses were only one element. The two categories with the highest response rates involved 'learning on the job', either by being self-taught (70%) or through peer teaching of skills by colleagues (50%). Far fewer indicated formal training mechanisms such as postgraduate study (35%) or external training courses (15%) as the source of their skills and knowledge. The most well used formal training was that provided by institutions. Around 35% had accessed this over the past five years.
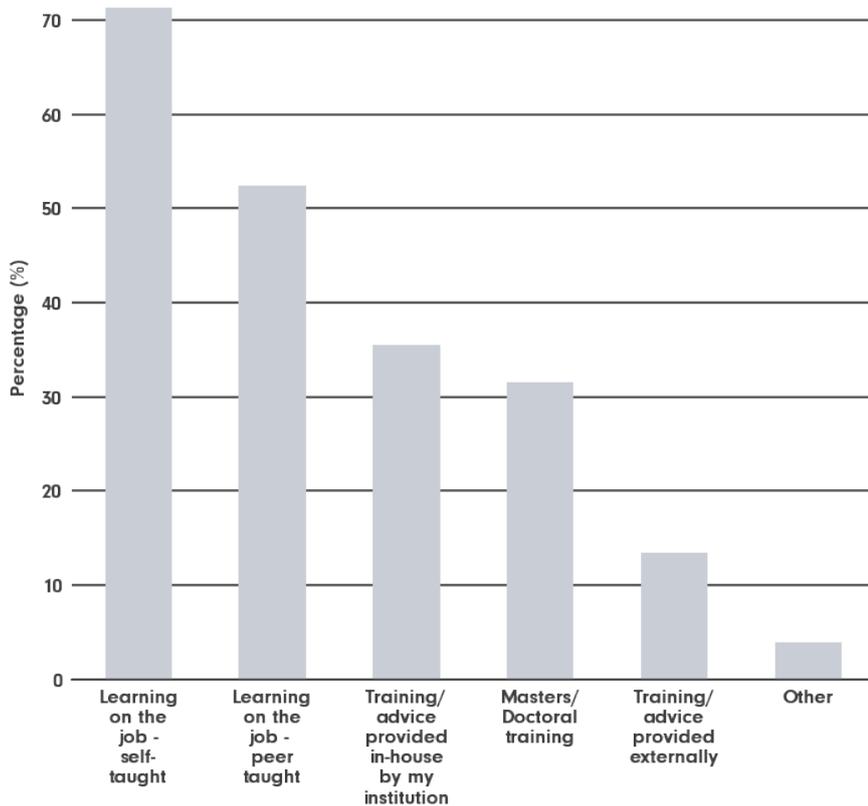
Figure 20. Where respondents acquired data and software skills.

The qualitative interviews provided more detail on what these different forms of skills acquisition involved. Given the arts and humanities research community is constituted by highly diverse disciplines, it is not surprising that the strategies for gaining and maintaining skills varied considerably. They also varied by career stage and were shaped by the nature of the local and virtual communities of practice.

Interviewees described different stages of the development process. The start of this process involved researchers gaining an understanding of what they didn't know, and what tools and practices were 'out there' in their field or aligned disciplines that might be useful to them. In other words, apprehending what they needed to know. It was often via communities of practice linked to their research area that they described navigating this initial learning curve. The trigger was sometimes problem solving around a particular research or methodological issue, or attending conferences and reading research articles which highlighted new or unfamiliar methods for data collection or analysis. One researcher remembered meeting someone at a conference who introduced them to a new methodological area and a research community based in Europe that ran regular symposiums. *'I went to 3 or 4 of those and then continued to go on to their conferences and contribute'* [409]. As members of those communities, they worked with colleagues on bids or projects using specific techniques and gained experiences of particular elements of the research process which stimulated an interest in gaining new skills. In addition, working on research projects and bids brought them into contact with new regulatory requirements and university systems that implied updating their skills, particularly in data management ethics and sharing. Communities of practice linked together researchers working in specific domains or datasets, with specific methodologies or technologies for collecting and analysing data by facilitating discussion of techniques. Case study 5 illustrates an interviewee who set up their own community of practice, and they were not the only such researcher in our group. *"It's something that I've been working on personally in the [faculty]. I've put together a workshop called [digital work in the discipline]... where I've been bringing together researchers across the faculty to try to talk about what they've been doing and how other people might think about doing that."* [132]

However, such communities sometimes have barriers associated with them such as cost. *"The [national forum] brings together a lot of us...it includes university partners...But there's a hefty cost associated with being there. So not everybody is there...Those people are in the Independent Research Organisation Committee."* [384]. This again illustrates how HEIs and IROs attitudes to conferences are not homogenous. We also heard how people used social media and other channels to acquire

CASE STUDY 4
Developing digital skills

Laurie is an ECR with a background in English and Media studies, more recently working in cultural sociology. After a string of insecure positions, Laurie had acquired a permanent job and was writing bids and applying for research funding to AHRC and others. They found it interesting to identify the various types of data in their research which included traditional interview transcripts, social media data, mostly images and screenshots and annotated hand drawn images — a physical object which required converting into digital form for analysis and storage. *'Our project had drawings girls had done of themselves and were scanned and sent to us by parents and came in different formats, so they came in as PDFs and Word documents and JPGs. And you know different image files. We tried to rationalise all of them into the same type of file, so we could store them.'*

Although Laurie understood data storage in broad terms, they had questions about the rationale for this in relation to their own research, *'who is even going to access it, and what use would it be?'* and felt overwhelmed by the specific processes required for their particular data. *'I don't know how to manage it long term. I don't even know what is available to do that or what capacity there is or where I should be sending it as it were'* .

Laurie was already engaged with using various digital tools but keen to better understand the software that was available to help manage and analyse data. They described how difficult it was to find out what was available and useful. *'All of these projects I've sat down with the people on it and we've all gone. Do you know of any program that does this? No, and then a few times I've looked at our like whichever university I was at, library software to see if there is anything that leaps out and there never is'.*

Laurie was in touch with their PhD network who provided a useful sounding board for research skills and knowledge, but they were struggling with the new institutional context and how to find information about training and research support. Laurie described the process of trying to get a question about their eligibility for funding answered. *'Research services people said "speak to your research officer". So I contacted Maggie, and Maggie said "talk to research services". [...]*

Laurie's main issue with regard to digital skills development was their lack of time. In this context, generic training was not useful, they wanted tailored support instead, which would help answer specific research questions and problems. *'The useful thing needs to be a lot more tailored to an individual. It needs to be a sit down meeting for 20 minutes, half an hour where you actually get to ask the questions which are pertinent to you and are not answered by the handbook.'*

information from their communities of practice, including Twitter and Discord. *"It's been such a brilliant resource, but it's not formal, and it's not official."* [419]

When researchers understood what they needed to know (or which specific tool they wanted to use) there were multiple ways to acquire those skills and knowledge. Formal and informal ways of skills development coexisted and served different purposes in the process. Formal training courses run by their institution played a role and were well used, but the choice of courses was limited to what the university would support. Formal mechanisms were most visible in the narratives of PGRs and ECRs (and some mid-careers), recalling or describing their doctoral training. Whilst there appeared to be significant variation in what was offered (and the COVID-19 pandemic had substantially reshaped that offer for recent cohorts), there was evidence of the availability of formal training courses and modules specifically for students with doctoral training grants, but also for students auditing masters courses, attending wider institutional training and less frequently accessing a doctoral training budget to take external courses with upfront costs. There was evidence that doctoral training sometimes included

skills such as research ethics or data management. One PGR described an array of training they had undertaken at the start of their doctoral programme *'I did the compulsory ethics and all of that data management, GDPR, that kind of training, and then in terms of software, I learned a few, I went to a few training sessions offered by the library on certain packages, which I can't remember the name of them to save my life.'* [132] Another described how they had *'incredible support'* with writing a DMP at their institution as a PGR: *'I got a kind of data management person, so I described the project to him, he wrote the data plan, and then he just kind of...I signed it off'* . [826]

Formal external training was not a key mechanism for skills acquisition for many of the participants, in particular mid-career and senior researchers. Many of them would describe themselves in the interviews as self-taught, although this involved a range of different strategies. They describe *'digging around'* and trying to *'figure it out'* using Google, watching YouTube tutorials, webinars and taking more involved online courses. One ECR, using a number of digital methodologies, explained *'I've had no formal training on any platforms, any data, software, [...] I taught myself Dreamweaver, I taught myself to blog, because I did a maker's blog alongside my PhD so that I could capture everything that I was doing'* [165]. Being self-taught meant using the 'help' facilities and support services and websites of the relevant software, as well as reading manuals and 'how to' books. One mid-career researcher referred to their learning of software as the 'gentleman amateur' approach, *'I was entirely self-taught off the website, and I learnt Excel the same way I'd learned TCL, [...] we had some templates and I fiddled with those and found stuff on the internet to do that'* [124]. A PGR described how they *'started looking for lectures online from other universities, I mean America seems to put a lot of their lectures online, and since they are very fixated with this period of history in England, it's a good resource'* [294]. Self-teaching was particularly prominent for those who had learned software skills and this echoes findings from other studies of software skills for research. One UK survey reported that of the researchers who develop their own software, 21% stated that they had no training in software development.[35]

Much of this informal learning could be a solitary activity and some PhD students noted that they were quite isolated in their department using different methods and tools to other students and often to their own supervisors. An ECR explained that although they had a colleague with lots of experience, they were so over stretched they did not feel comfortable asking them for help and that was true of other colleagues too *'with my own peers, yeah it's very difficult'* [165]. Whilst informal skills development could be lonely, participants had joined online forums and groups that provided virtual communities of practice focused on a particular methodology or tool. One senior researcher who couldn't *'remember when I was ever on a training course for a piece of software'* , explained *'[I] pick it up, pick it up with the community I'm working with and muddle along through'* [183]. These communities of practice played a key role in knowledge exchange and skills development. *"I can't comment enough on how valuable communities of practice were in my training and gaining my skills repertoire. Whenever I have a software skill that I want to gain, I'm referring to existing communities of practice or the Programming Historian."* [97]. An ECR working with databases explained how they had solved a particular research problem by joining the online community where solutions were being discussed. *'It's just a lot of googling and going onto forums, you find yourself on Stack Overflow or something similar, so I was aware of the different forums and how you approach it, sort of a technical question, and read the blogs about it.'* [128]

Research teams were also an important community of practice and source of informal skills development through the opportunities they offered for peer teaching. A professor working on cross-disciplinary project with digital humanities scholars noted they were learning a huge amount from their colleagues through the process of bringing disciplines together. *'I'm learning about good web design as well and about the protocols of textual encoding, what text becomes when it is in digital form [...] so yeah I'm being stretched all the time with this project it's really exciting'* [379] Another professor, reflecting on working with cross-disciplinary teams, noted the importance of what they called 'shadow skills', a basic knowledge that enabled collaboration on an interdisciplinary team, *'I make sure I read in that world [mediaeval scientific Latin], they don't make me an expert in that discipline'* [183] The key was to ensure team members were training *'we would train each other up all the time throughout the duration of the project and that would include a subset learning to engage with pieces of software'*. [183]

3.3.3 Barriers and facilitators to acquiring data and software skills

When asked in the survey if any of the issues listed prevented researchers learning data and software skills, around 45% of respondents said they were aware of training opportunities but did not have time to enrol [Q20, n=281]. Around 27% said they were not aware of training opportunities. Just over 10%

35   https://www.zotero.org/google-docs/?Fdo9WC

felt that their institution didn't not support the technology or infrastructure necessary to make use of the skill. Around 20% selected 'none of the above', but their responses in the free text box overwhelmingly referred to a lack of time, suggesting the options were not clearly worded enough. Other text answers suggested that some found potential training options confusing or noted that they didn't know where to access information about training courses, or were reluctant to do training if they couldn't then embed the skill in their current research practice.

When we look at the same chart by career stage, we can see roughly similar patterns for all career stages but with some differences. The time issue was notably higher for early, mid- and senior researchers compared to PGRs, which perhaps reflects the prioritisation of skills development in the doctoral training context. Equally, a much higher number of ECRs identified a lack of funding for training. Again, this might reflect their greater interest in training than later stage researchers combined with their lack of access to resources.
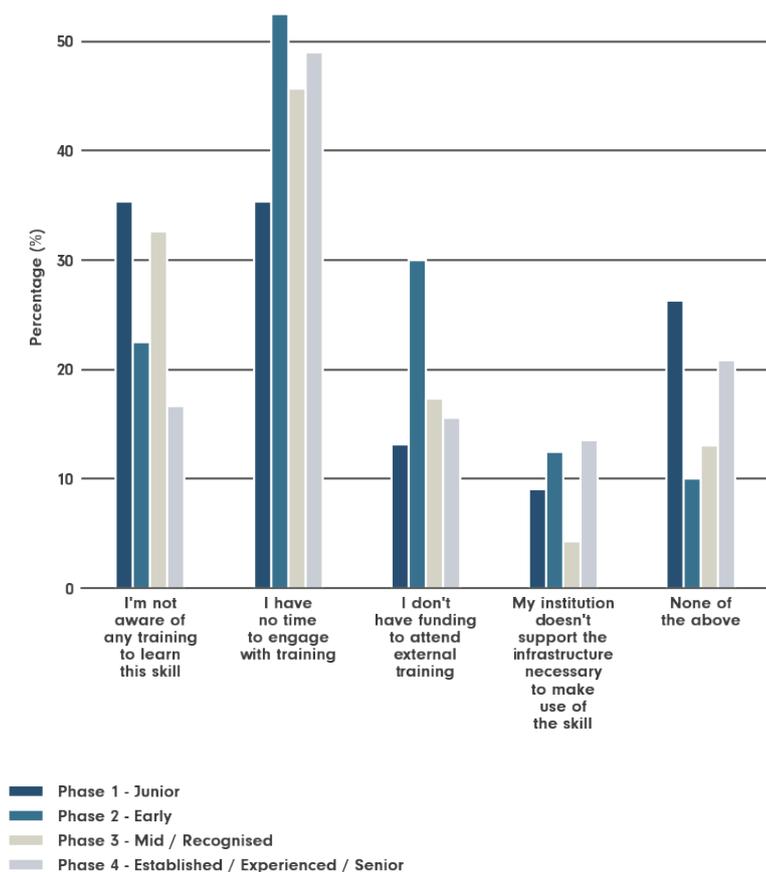
Figure 21. Barriers to learning data and software skills by career stage.

The interviews provided rich insight into the intersection of barriers and facilitators that operated for researchers in relation to their skills development. Issues of cost were visible for both PGRs and ECRS. One PGR for example noted that officially *'you can apply for things'* (via their doctoral training partnership consortium), but in reality there were caveats about need, priorities, and budgets were limited... *'It's the same pot that gives things for like internships, and I want to do an internship...'* [325]. Another PGR similarly noted that they found a really good course but it was expensive, *'nearly 500 pounds'; 'I don't actually have that money in my research budget to do that'* [254]. Ultimately, they had not applied for the training. ECRs who were not working as part of a research project struggled to access funding to support the cost of external training and were limited to what their organisation could offer for free or what they could learn themselves. One ECR mentioned that they had done some research on a particular area they were interested in and identified a course at another university, but they could not attend because they had to pay for it. *'If it comes with upfront costs then it's quite hard to front these costs if you don't know if it's going to directly help you'* [132]. Another had been encouraged to apply for postdoctoral funding and 'cost in' the various training courses they wanted to take in order to ensure they were developing their skills. However, their postdoctoral applications had been unsuccessful and this strategy had not helped them gain the skills they wanted.

Current trends towards work intensification in the sector, combined with institutional cultures and practices which do not directly reward training, meant researchers were overstretched and could not

prioritise formal training.[36] ECRs, mid-career researchers and senior researchers all talked about time pressures in relation to academic work that prevented them undertaking training and knowledge development A mid-career researcher who was really keen to learn programming, specifically Python because a colleague who had been informally supporting a project no longer had capacity, noted that they simply didn't have a big enough block of time available for learning something new whilst doing research and teaching. They struggled with how *'to justify the time when it doesn't necessarily lead to an output straight away'* [171]. An ECR pointed out how it should in theory be feasible, if the time to attend a training course was formally set aside each week, ... *'each Monday you go to your coding course and you do your homework, you do a project'*. However, they added that it could only happen if there was formal institutional support *'that means that it has to be something that your employer recognizes as part of your work. It's CPD [Continuing Professional Development], you're not doing it for yourself, you're doing it for your work and if that is not recognised I don't see how it could be done.'* [128] A professor, keen to learn new quantitative social science skills around data visualisation and application programme interfaces that would enable them to contribute more to projects, admitted it was unlikely they would be able to find time: *"...how much of my job can I commit to this?"* . Like others, they found formal training was not sufficiently flexible to fit around their multiple other commitments.

One researcher described how the courses at their research institution were not very up to date, and the training unit wouldn't offer training until there was some *'loud clamouring'*. They were also seen to be mainly *'basic introductions'* and often *'quite generic'*. A PGR described searching regularly for relevant courses in their institutional training site but couldn't find anything they needed. *'There was an intro to coding. But it was so basic, it didn't look like it wasn't going to take me anywhere. And that's all, that's the most that's offered'* [235]. More senior academics were unwilling to spend several days on a generic training course that might not serve the very specific needs of a project or bid. Lack of time then was also related to scheduling (how the training could fit with other tasks, and the direct relevance and specificity of the training). Being able to understand how relevant the training was going to be helped with whether the time could be justified. Several interviewees suggested that rather than signing up to a week-long or even two days-long workshops, it would be more useful if *'you could get someone to sit down with you and train you'* [344] in the specific thing you needed to know. Again, more informal approaches could be tailored to fit the time available and be more targeted in terms of relevance. Many of the barriers to accessing formal training were discussed by those in HEIs.

For those in IROs, formal training was embedded in organisational practice and participants talked about training courses they had undertaken or courses their colleagues were in. Equally, an ECR in a research position on a large funded project talked more positively about skills development and formal training opportunities as part of the project (they had chosen to do training in Agile to help them work with their team): *'My training needs I think that they're kind of evolving alongside the project... I have, let's say, a number of skills or, like, you know, training goals, let's say, a list of training that I would like to do'* [97].

However, far more of the researchers we interviewed were using a range of informal mechanisms to gain and sustain their skills and knowledge. These were more flexible in terms of time and could be tailored to specific needs, providing a way around some of the time, relevance and scheduling issues of formal training. Other researchers leading large cross-disciplinary research teams talked about the importance of getting the right people on the team. Accessing skilled people for a research project did not have to necessitate training up researchers, but rather meant ensuring the project had key specialists such as RSEs and librarians on the team. A mid-career researcher explained *'I now have better people to do those roles than I could ever acquire the skill set for. [...] I think we need a real recognition of the value of specialists, data scientists within heritage'.* [403]

3.4 RESEARCH SUPPORT AND INFRASTRUCTURE

In this section we focus on the institutional context in which research was taking place, participants' understanding of that context and their views on what might be done to better support their research

36 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.

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.

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

and their development of skills and knowledge. We begin with a summary of the key insights from our policy mapping. We then look at some wider policy issues pertaining to data and software, including facilitators and barriers. We go on to examine researchers' confidence in their own knowledge of, and adherence to, their institutions and funders' policies related to data and software use and management, and issues of legal and ethical compliance. We finish by exploring the responses to presenting some potential policy interventions discussed with interviewees.

3.4.1. Key considerations from UKRI policy mapping

The data management policies of the seven UK research councils were reviewed to gain a wider understanding of the landscape (see Section 2 for methodology). The purpose, principles and structure of each relevant policy were mapped and the content regarding access, use and availability of data, training and skills guidance was then analysed. The table below summarises the content and the focus for the guidance which often link to other UKRI guidance.

Organisation and policy	Guidance	Tools	Policy purpose
AHRC - UKRI Open Research	DMP Outline UKRI Best Practice of Management of Research Data	Research Data Management Toolkit (JISC) Suggests Digital Curation Centre (DCC) and Data Protection Commission (DPC)	Replicates UKRI's open research policy - supports open research
BBSRC - Data Sharing Policy	In Policy Statement on Good Scientific Practice Resources for Data Sharing	No reference found	Focuses on opening research data to stimulate new investigations in order to get best value for the funds invested
ESRC - Research Data Policy	UK Data Service UK Data Archive Peer Reviewers Guidance	No reference found	Policy explicitly supports grant holders who *'collect, produce and re-use data'* with an emphasis on roles and responsibilities
EPSRC - Policy Framework on Research Data	Clarification of EPSRC expectations of research management	No reference found	Directly about the management and provision of access to data: publicly funded research data should generally be made as widely and freely available without damaging the research process
MRC - Data Sharing Policy	Guidance and Standards for data management, preservation etc DMP Guidance	No reference found	Maximising research opportunities
NERC - Data Policy	Guidance for ODMP	Data Service assists in creating DMP	Essentially supports open access and open research, but particularly talks about *'enabling the tracking of [dataset] usage through citation and data licences'*. They also have a legislative requirement to guidance on the manage and distribute environmental information
STFC - Scientific Data Policy	Data Management Plan	Suggests DCC	A stewardship requirement to ensure data is carefully managed and optimally exploited for economic impact. Covers both researchers and facilities

Table 6. UKRI Council data management policies.

The review found that the policies and related guidance were challenging to locate, there was a lack of clarity on what constituted data within them that may result in less compliance with those policies. For example, the STFC document specifies it covers raw, derived and published data, although the latter is not often understood as data. The BBSRC discussed a wide variety of types of data, including physical biological specimens. As the AHRC uses the generic UKRI policy, it does not specifically address issues of arts and humanities data.

Most policy guidance around data use focuses on downstream re-use by third party non-applicants, subsequent to collection/generation. However, the ESRC is explicit in stating that all efforts must be made for grant applicants to first attempt to reuse existing data (upstream reuse). The MRC puts the onus on researchers to ensure opportunities for reuse are maximised through preparation and promotion.

Data access is the area with most commonality between policies. In general, councils do not prescribe timelines for sharing, but require 'timely' sharing of data and metadata, with some providing suggested timelines. Exclusivity periods apply. Some councils, e.g., BBSRC, are more specific about accompanying documentation requirements, and the NERC requires a data access statement (also known as a data availability statement). The AHRC does not prescribe but requires *'data preparation'* including metadata, timely publication, exclusivity (which should not impede or delay other research outside the original scope, neither should patents/IP).

Some research councils have a data repository function for making data available, and are therefore quite specific regarding where data should be deposited. On the other hand, other councils are more laissez-faire, with EPSRC only requiring data to be 'available on the internet', which could include mailing the author with a request. The ESRC and the NERC require data to be deposited in one of their services. The AHRC discusses discoverability but does not otherwise mention repositories.

Few councils consider training and guidance as part of their data policy. The ESRC is the most detailed in this area. Some councils do provide more guidance regarding data management in related documents. Whilst the ESRC and BBSRC specify what should be included in a DMP in their policies, the AHRC requires a DMP but does with no specification of what should be included in it.

Although the key time to consider ethical and legal compliance is during data collection or generation, most councils focus on downstream (sharing) activities rather than upstream (collection) activities, perhaps due to the fact that institutions normally have ethics approval infrastructure applicable to any research, UKRI funded or otherwise. The AHRC provides extensive links to guidance on GDPR, confidentiality, and so on.

There are skills and knowledge that are either explicit in the data policies (such as the ESRC encouraging the use of secondary data) or implicit, based on standards or requirements of the research councils (see Table 6 above). The policies are directed at digital asset management (the curation and archiving of data to ensure their useability

CASE STUDY 5
Communities of practice

Kiran works for a major heritage and conservation organisation. Before the COVID-19 pandemic, they became increasingly interested in the possibilities of data science for their work. They attended some regular meetups in their local town, which consisted of a speaker followed by some networking. *'You just meet other people and realise you can feed off each other quite quickly, and you can find out what things are useful and what's not useful and where the training opportunities are'.* While the coding practices discussed were not direct examples of Kiran's work, *'that coding practice could be applied to [my] industry, because it is quite generic'.*

During lockdown, the meet ups had to cease and Kiran found themselves trying to locate other people in heritage and conservation who were interested in coding. They did so by sending *'a few emails and saw a few messages and a few intranet posts and stuff on a few message boards'.* The interest quickly grew, and Kiran sought sponsorship to support the online community, which has now reached 600 people, which they believe, *'represents a good chunk of the heritage workers and conservation workers that are using coding'.*

Not all of them are experts or day to day coders. *'There's probably about 20 people really using data in that sense. Most interest is in learning how to use data and how to code, which is really nice for the group...but for more experienced coders, maybe there's a better way'.*

Questions to the group range from very technical, specific questions to more general. Community members also share information about upcoming training and events.

for future researchers) and digital data manipulation (the collection and analysis of data), although they primarily focus on digital asset management, in line with the wider focus on open research data and effective reuse. Digital asset management covered in the policies includes familiarity with relevant legislation, the ability to create a data access statement, familiarity with institutional or funder repositories/archives, knowledge of metadata and the ability to identify risks of sharing – and balance that with potential benefits, in some cases. It also includes the ability to use standards such as the Data Documentation Initiative (DDI), Statistical Data and Metadata Exchange (SDMX), or INSPIRE (for spatial data), copyright/IP knowledge and data licencing. As the UKRI policy which the AHRC uses, it focuses on publishing open access data and skills related to this (such as appropriate licensing, and the ability to create a data access statement), which are therefore implicitly the most important for the AHRC.

As noted, the AHRC policy links directly to the UKRI Open Research Policy whereas, the other six councils have their own customised data and research policies. Councils which have modified and adapted UKRI guidance have been able to use their policies to drive their own agendas, in particular around reuse, deposition in a repository/archive, training or specific values. By following UKRI rather than providing tailored guidance, the AHRC's data policy is the least specific policy amongst the councils, which suggests there could be variation in how researchers are interpreting guidance.

3.4.2 Confidence in policies

To understand researchers' engagement with wider policy and guidance, the survey asked about their familiarity with and confidence in their understanding of policies on data and software, as well as the correlative responsibilities. In response to questions about confidence during research (i.e., data manipulation), 50% of early, mid- and senior respondents reported largely being confident or very confident [Q21, Q22, n=272]. Junior researchers were much less confident, although this seems to be rectified by the time they reach an early career role. Interestingly, senior researchers were slightly less confident than mid-career researchers, perhaps reflecting a less 'hands on' role in the research process.

Phase 1 - Junior
Phase 2 - Early
Phase 3 - Mid / Recognised
Phase 4 - Established / Experienced / Senior
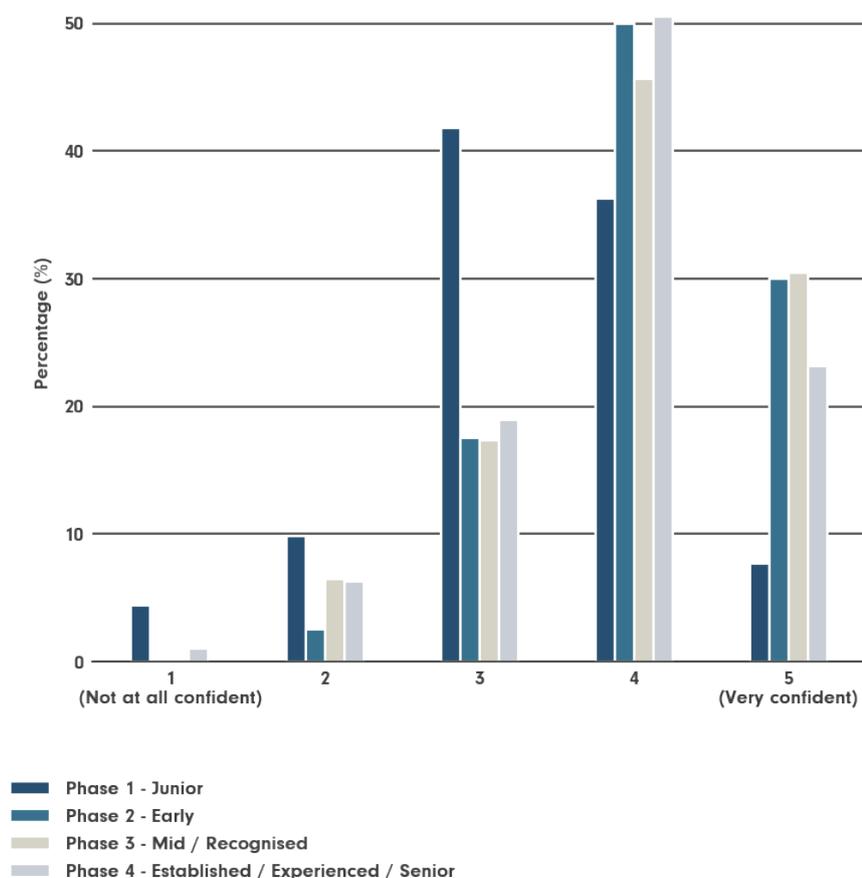
Figure 22. Confidence in understanding relevant institutional/funder data policies during study by career stage.

Confidence in understanding data policies at post-study (i.e., the data sustainability stage of the research) was slightly higher, with over 70% of ECRs saying they felt they were confident or very confident in understanding what they were required to do with data post study, versus 50% of ECRs reporting being confident or very confident about their understanding of applicable data policies during the study.

Phase 1 - Junior
Phase 2 - Early
Phase 3 - Mid / Recognised
Phase 4 - Established / Experienced / Senior
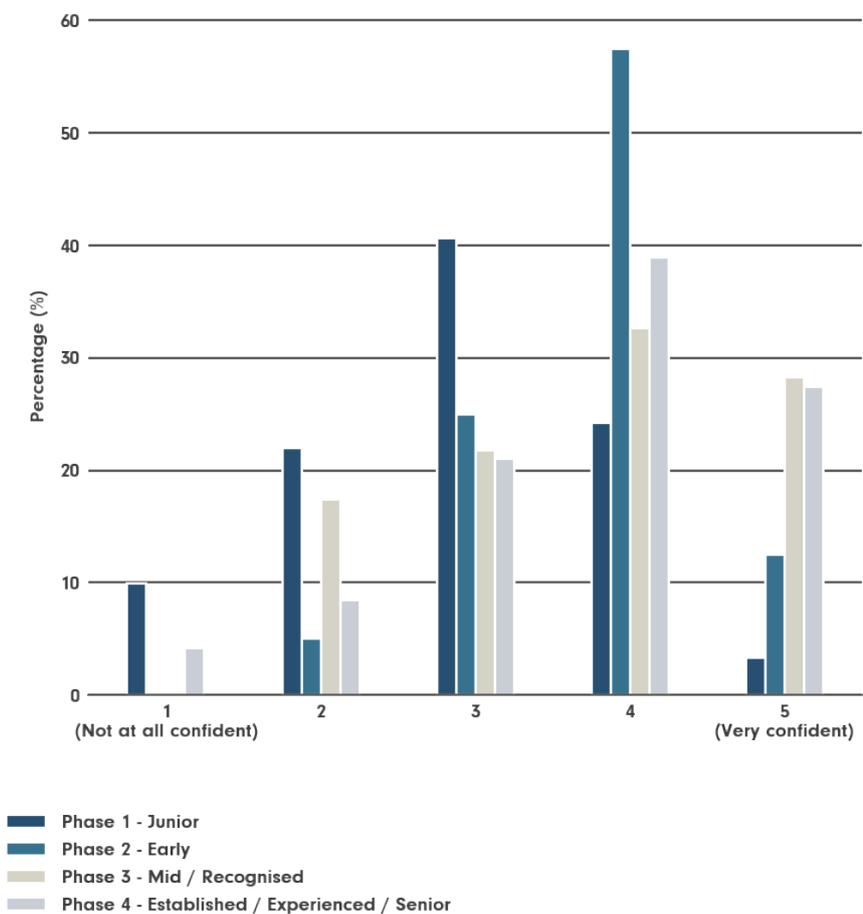
Figure 23. Confidence in understanding relevant institutional/funder data policies post study by career stage.

When asked about their familiarity with their institution or funders' software policies, respondents demonstrated less assurance, with the most selected category being 'neither confident or unconfident'. [Q23, 24,n=264]. Again, junior researchers are the least confident career stage, although ECRs are increasingly confident. Again, the confidence levels of ECRs increases post study.
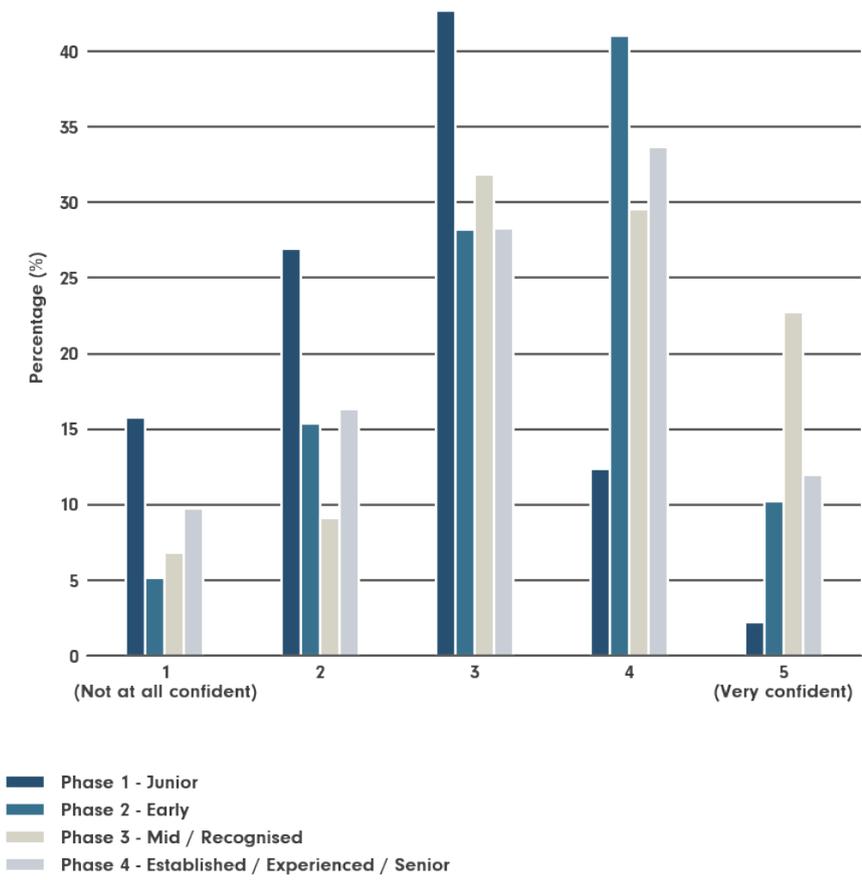
Figure 24. Confidence in understanding relevant institutional/funder software policies during study by career stage.
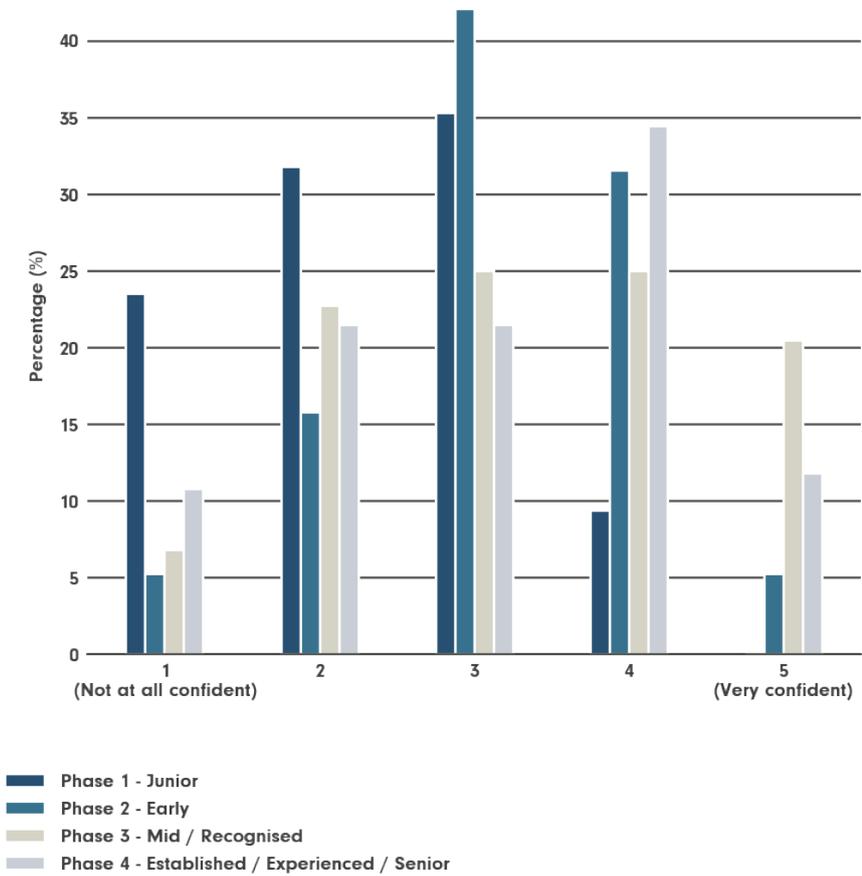
Figure 25. Confidence in understanding relevant institutional/funder software policies post study by career stage.

3.4.3 Infrastructure sustaining data and software

The interviews explored participants' experience of the various forms of infrastructure and guidance that supported and circumscribed their research practice. One of the challenges they identified was the variety in the types of platform on which data could be deposited and access to them, whether because access is not sustained, or access is not available externally. *'Not sustaining that website, whatever that involves, ...[means] you lose all that data you've gathered together about which of these buildings has these particular artefacts.'* [409]. This was a particular issue where journals attempted to create links to the underlying data in a research article. Whilst this was helpful in rewarding the sharing of data, the researcher found the necessary infrastructure wasn't in place for them to do it. *'The journal was like "If you have associated data, where is that? Where can we access that data?" And I was like, "You can't because I don't really have any infrastructure for doing that at the moment." [235]* Another noted that the university repository was not fully accessible in practice. *'In principle, I can say, I let [the journal] have access to the data..[but] in order to access our archive, you have to be a university user. So you have to have an account with the university.'* [140]

When projects operated across partners in multiple institutions, this introduced further complexity. For instance, one partner may be responsible for the data storage, but struggled to formalise access for the other partners. *'We've a good relationship with the [partner University] over the linked data. But that's not formalised in any way, shape or form'.* [344]

A lack of funding for sustaining data was a key issue. A common situation is one where funding has been made available to develop a database or software, but not to sustain it beyond the initial period. *"So let's say once you've developed one of these databases, you've then got to maintain it. And that's not in the funding....it's an unsustainable business model."* [49] Linked to this was a concern with the lack of human management and stewardship of the data: *"Data's for life, not just for Christmas... where does the data go in five years time, two years time, when the people have been involved in the project have moved on, retired, and somebody switched the server off."* [49] Researchers were concerned with difficulties in fulfilling future compliance activities such as subject access requests. *"In ten years in the future do the [partner institution] want an email from somebody saying 'I've just read [respondents] thesis, they say that in 2019, you did a questionnaire with a group of people, can I read that?' "* [419]

While some of the same issues applied to software, there were also some specific challenges to use and sustainability. Proprietary tools *"limit transparency between different users...it's a nuisance"* [140] and also limit sharing, given that if data was associated with a specific software, a researcher would not have access to it without the software, *"the problem is the instrument manufacturers don't make their data accessible, unless you have their software - that's got to change"* [140] There were also issues of having an appropriate level of guidance. This may be a lack of standard guidance, *"They said, "what, what software? Did you use Transcribe or use Word?" And I just assumed that anyone would find like a bit of digital software for transcription, open it and use that."* [401] In some contexts (more often IROs than HEIs) the barrier may be restrictive guidance, *"I use a variety of tools but officially at work I'm only supposed to use Excel to analyse data."* [112]

There are also upstream (suitable hardware for the software to run on) and downstream (suitable data) limiting factors. *"In America, they have, they've been way ahead on digitising historic documents...in England, you would have had to write 16 letters and then finally get a picture with a massive word across."* [409] In terms of the software itself, the limiting factor may be that the specialist knowledge needed is no longer available or funded. *"There's only one person, the person who designed it who is based in [university] and I just have not had the funding or the capacity to get him back up to [university] to pull that data. So at the moment is just sat in these kiosks and I cannot extract it."* [165] In this case, funding had been explicitly available to collect the data, but not to analyse it.

A key final infrastructural issue was fitness for purpose, the lack of appropriate or suitable software to carry out some of the more complicated tasks beyond the ability of current machine learning/artificial intelligence, such as identifying obscure people from drawings, *"I think, you know, crossing traditional quantitative data with qualitative data, we're pretty good at that. I don't think we always really find a way to cross with the data that we produce through the artwork."* [268]. Software is infrequently, specifically created for a particular problem and this means that its not always exactly what is needed. *"We've got great data analysis tools, but they tend to still be in their own areas, rather than a way that I think could be more productive."* [268]

Our interviewees also provided more positive narratives around the use and management of data and software. Often these reflected understandings of data best practicewithin the community, such as the

FAIR principles.[37] One of these was the importance of data stewards, people whose role was to have oversight on what data is archived and how archival content can be exploited. They were described to us as *"research catalysts"* who enabled effective data reuse [579]. Data stewardship of this nature is more visible in other UKRI councils' data policies, which have an explicit focus on the roles and responsibilities of researchers regarding data.

Data publication and promotion, when successful, was considered to be a facilitator in disseminating and sharing. One ECR discussed writing a 'data paper', as opposed to a 'research paper' at the end of the project. *"It just draws attention to the database, advertises that it exists, it explains what it is good for, expands the reuse potential, and indicates what I was going to use it for. This was published in a peer reviewed journal. And because of this, I think some people who are interested in it, the subject, they found the database itself."* [128] Publications that explained the purpose and content of data, and how it was intended for reuse, were regarded as important, although this seemed to be largely idealistic rather than widespread practice. *"I think you need a bit more, you need incentive to do the documentation and, and publish it in a form that is useful for others"* [128].

3.4.5 Concerns and compliance with legal and ethical policies

Ethics, GDPR and licensing are all issues of information and research governance. The interviews generated some important discussion about research and data ethics. We spoke to a number of researchers who were deeply concerned with ethical issues and saw questions about data practice as intrinsically connected to research ethics. *"I think ethics is a process where you think back and forth between your data, the people that are represented by the stories they've shared with you, and how you treat [data] throughout the project."* [254]. There was a suggestion that other councils had more expertise in relation to data and ethics and this was an area that AHRC could develop. One professor, whose work spanned several disciplines, felt there was more *"I feel that [AHRC are] not so aware of these questions, ethical questions about data within the AHRC as I am in my work in the other funding councils MRC and ESRC. Just feel that they've been at that game for longer."* [268]. They were keen to see the AHRC open up debates among its community around data and sharing and its production.

Some researchers felt that much of the ethics process was a tick box exercise rather than a fully realised, accountable process. *"We felt that ethics actually isn't provided anywhere, and where it is, it's very much here's a form, here's the guidance."* [254] We also heard similar views about GDPR. *"I want to make sure that all of that data is being handled in a way that is appropriate, legally, and ethically. And of course, I do my GDPR training. But it's all very general. And actually, that's just for the university to check to say that I've done it, there's no actual infrastructure or support for ensuring that in practice, those things happen in a robust way."* [165] Researchers felt that more accountability needed more skills for delivering the research while staying committed to the ethics: *"It would have been great to have had more enabling for that sort of thing in terms of skills, as the ethical side of it was quite compromised, because you're just trying to make sure that you can deliver on what you said you were going to do."* [409].

Licensing was seen as a challenging area due to a number of interlocking aspects. For instance, there is a lack of clarity on having the right to share data if it had been collected with another organisation. *"We're good at managing our own information. That's the copyright and licences care. But with the third party information, we don't have the copyright and the licencing sorted out"* [49]. This can frustrate moving towards open access and open publications. Previous poor experience made researchers wary. *"We had some hiccups, that some of our collaborators experienced on the way in terms of who was exerting copyright over the work they'd done...So we adopted Creative Commons very early on. And hopefully, you know, that that will have made it all worthwhile."* [147]

A number of interviewees discussed how they were not sure who to talk to regarding licensing, on an institutional basis. *"Finding out about things like who to talk to about contracts, about licences, particularly licences, that's a real problem."* [401] In some cases, there was a perception of insufficient accessible support. *"We probably really need to have a research data manager or, or someone who's actually a specialist, because a lot of [licensing guidance] is technical."* [344] The challenges of navigating an opaque and poorly managed institutional system combined with overloaded academics, provided a subtext in these discussions. *"The university's policies, you must hold things for five years... But I legally have to hold on to these things. And I don't know, you know, I haven't got the space to do it. And I don't know who to and I asked a few different people who just say, I don't know, and pass me to different other people."* [165]. Another researcher told us there was *"no coherent strategy being passed down with support to the coalface."* [140] However, there were also examples where

---

researchers were confident they knew where to get their question answered. *"I would turn to the library for a lot of ...licencing queries...and our institutional archive type stuff."* [147]

3.4.6 Future policy scenarios

In the interviews we presented four policy and practice scenarios to our interviewees, based on the early exploratory interviews (see Section 2.2.3). These acted as an elicitation device providing concrete policy interventions that researchers could engage with and potentially critique. They were presented as ideas for discussion that had emerged from previous interviews, rather than AHRC endorsed policy plans. In practice they also provided a way for participants to talk about their experience of current policy and practice in a neutral way.

1. Improving the prominence of software and data in AHRC funding requirements

The first proposition was to improve the prominence of software and data training (for instance, data literacy, data management, ethics, documentation, data sharing and workflow) in doctoral training requirements. This was popular with participants at all career levels and across disciplines. It was seen as beneficial in providing baseline knowledge of digital research infrastructure, addressing the 'signposting' problem where researchers are not sure what they need to know, and issues of vocabulary and understanding that would encourage engagement in a wider range of research. *"Because I think, to kind of fulfil what the council wants, in terms of engaging with other research, they need to have that baseline knowledge, so you know to what extent you can participate in it. And so you need to be able to engage, discuss on that level."* [344] Respondents welcomed the broad range of data skills, and improved consistency between institutional provision, and individual access. Overall, this was perceived as a well-focused, practical and implementable solution of value to PGRs going on to careers both within and beyond academia.

The second part of this scenario was to improve the prominence of software and data in research grants, with a specific aim to reduce a perceived process bias against cross-disciplinary applications. Whilst there was agreement that this was an important aim, participants pointed out some of the challenges, such as a lack of suitable reviewers able to assess cross-disciplinary research leading to good applications being rejected on technical grounds, or acting as a demotivator for researchers not working with software or data. *"Unless you work with these methods, this sounds like something extra that you even as a doctoral researcher you might not take on...It might reward people who are already doing these things but it might not reach those whom the AHRC want, because the obstacle, the barrier is just too big".* [128] This suggests more research is needed to shed light on the language around software and data in applications and guidance.

2. Knowledge exchange and building communities of practice

In this scenario, the AHRC would support conference attendance at recognised arts and humanities conferences, to enable researchers to present papers or facilitate training sessions that focus specifically on data and software practice, disseminating the technical knowledge rather than or in addition to the research results or findings. This was a popular idea across all participants from PGRs to professors. Most were able to identify a conference in their discipline where this could be implemented and suggested types of sessions they might find useful. They were positive about the opportunities to meet and discuss research methods, particularly at focused sessions, and the interactive and personal nature of conferences was felt to be important as a context. There was a perception that more promotion of methods data and software strategies at conferences could fill a current gap in dissemination of arts and humanities knowledge, which was felt to be too focused on results rather than process. For example, some participants argued it was a missed opportunity by AHRC not to support a CapCo conference to disseminate learning from that programme. Attendance at academic conferences is not usual practice for IROs. However, they were generally supportive of external training and funded attendance at conferences was seen to make them more attractive to staff, but it may not be sufficient in itself. *"If it was, like collaborative research projects, and there was income generation, then we're probably more interested."* [112]

3. Data and software fellowships

This scenario proposed an academic fellowship scheme, through which an arts and humanities researcher developed specific technical skills as part of their research, bringing together domain knowledge with digital expertise. This might be software-oriented or more generally about data collection, data manipulation, data management and ethics. Importantly, the fellowship would require that they then would support colleagues in their department or faculty in developing digital skills. A second type of fellowship that emerged during the discussion was for a RSE fellowship. It would be embedded within a specific school, department or faculty, working with groups of researchers to support their research and disseminate software skills and knowledge about techniques and methodologies relevant to their field. More widely, it was noted that existing fellowships could be strengthened in terms of their requirements for data and software skills. *"I do think there's a way in which perhaps some of the, some of the rigour in terms of data and digital skills training that is built into some of the AHRC strategic calls could be incorporated in other strands like fellowships"* [379]

4. AHRC Skills Hub

We presented two versions of this scenario to our interviewees. The first was an AHRC- managed site, which detailed all the data and software skills that the council is keen to promote, linking to preferred training suppliers and groups, so researchers knew where to go for training. Again, this was received favourably, in particular the idea that all training could be found in one place, that suppliers were approved or validated by the AHRC and that skills provided were considered relevant by the AHRC. Concern focused on how and whether the AHRC would effectively sustain it over a period long enough for it to be effective and generate some momentum around training, rather than simply be set up and left to become increasingly out of date. *"What such an approach needs also to consider is that because of speaking about technical skills, you frequently have updates, which mean, roughly, every two years, you might have to refresh your skills."* [377]

In the second version, the Skills Hub would be run by a person or team outside the AHRC who would be available to give more detailed guidance, and ensure the content was up to date. One respondent suggested this could look like the Arts Council's Digital Culture Network[38], run by nine Tech Champions, but for academia. Benefits of this model that respondents identified were that it would form the basis of a community of practice; it could assist in developing the prominence of training in funding bids by supporting the writing of such sections; it would promote relevant skills and provide a space for discussion, and by extension, assist in developing a culture of digital skills. *"It's about creating a culture, about how you prioritise skills or how you think those should happen...there's something about making it clear that these skills are valued as well".* [140]

04_

CONCLUSIONS AND POLICY IMPLICATIONS

Our study took an iterative three stage, mixed methods approach to understanding the data and software skills of the arts and humanities community. We set out to:

1. Map the diversity of the UK arts and humanities research community and its constituent groups and establish the basis for our survey sample. A review of data policies across UKRI and the higher education sector in the UK was conducted to provide context. Exploratory interviews with key informants and researchers supported the design of the survey questions, the interview topic guide and the policy scenarios that would facilitate the design of recommendations. The mapping and interviews clarified the important subgroupings of discipline and career stage and highlighted the relevance of reflecting diverse views and experiences beyond those of researchers already using digital tools.

2. Undertake a large-scale survey of the UK arts and humanities research community to establish practices and skills in the use of digital tools and software. The short, 10-minutes survey was designed to maximise response across a broad range of groups and asked about research practices, including data collection, analysis, data management, institutional support and research skills. We achieved 335 responses and a broadly representative sample when compared to HESA data on the arts and humanities research community.

3. In-depth semi structured interviews were conducted with 30 survey participants who had 'opted in' to further research. A purposive sample aimed to reflect the diversity of the community as far as possible. The interviews focused on understanding research practices, experiences and cultures in more depth, including how skills were developed, forms of institutional support, the barriers and facilitators to developing digital skills. Policy scenarios that emerged from analysis of stage 1 interviews were used as an elicitation device to explore best practice and policy solutions.

4. Undertake analysis and synthesis of the various data sources (policy mapping, survey and interviews) in order to provide a rich and detailed account of the research practices relating to data and software found in the arts and humanities research community. The account aimed to highlight the diversity of research undertaken, the various data and software skills being utilised, the barrier and facilitators to digital skills development and the ways in which the community is supported by infrastructure, as well as potential interventions through which the AHRC might further develop its support.

Below we draw out nine key conclusions and policy implications from the study. These provide the foundations for our recommendations to the AHRC for potential policy interventions to consider, as part of their Infrastructure and Skills portfolios.

1. **Diversity of practice** - The arts and humanities research community is constituted by a diverse array of disciplines, within these there is significant variation in research practices around data and software. This is an important strength, creating a vibrant field with potential for both domain depth and expertise, cross-disciplinary collaboration and innovation.

   > The AHRC data and software policy need to be flexible and sensitive to disciplinary and sub disciplinary differences that shape researchers' understanding of data and their engagement with digital tools. To be effective, it needs to include software and practices at all stages of the research process, including data manipulation and data management rather than focusing specifically on open data (data sharing).

2. **Interdisciplinarity and collaboration** - Arts and humanities research (like other fields) has undergone a long period of digital transformation that is reshaping its cultures and practices, in particular emerging areas of expertise and increasing cross-disciplinary and collaborative research relationships and projects.

   > The AHRC data policy and funding should recognise and support the diverse practices, needs and priorities of the research community, including a range of research roles such as RSEs, and other technical professionals, ensuring funding opportunities to encourage cross-disciplinary research.

3. **A spectrum of engagement** - The arts and humanities research community is characterised by a wide spectrum of engagement in digital tools and practices, although overall it may be lower than in other research communities. At one end there is a full and enthusiastic engagement with digital research tools and infrastructure, dominated (but not exclusively constituted by) the digital humanities. At the other end, arts and humanities researchers continue to use traditional non-digital methods of data manipulation or focus on theoretical techniques that do not use data.

   > The AHRC interventions designed to facilitate researcher engagement with digital tools and infrastructure should be inclusive, respecting and being sensitive to disciplinary differences. They

may be most strategically focused on the large swathe of the community positioned between the two ends of the spectrum. A focus on developing all research practices (and normalising digital skills within that), as well as promoting the use of digital tools, would broaden the appeal of interventions.

4. **Communities of practice** - Research in the arts and humanities (and elsewhere) thrives where there are communities of practice to support innovation and development of research practice and skills. These communities might be small or large, virtual or geographically located, and focused on a domain or a specific form of data or digital tool or software. They are facilitated by opportunities for collaborative and cross-disciplinary research and knowledge exchange events such as conferences and seminars.

   > Policies and interventions can promote inclusivity and reach larger swathes of the community by focusing on supporting and building communities of practice across the community. Strategies to do that would include encouraging and rewarding collaboration and cross-disciplinary research in funding opportunities, and, by implication, funding fewer lone scholar studies, and supporting knowledge exchange and networking at conferences.

5. **Interdependencies between formal and informal training and development** - Skills and knowledge in data manipulation and data management are acquired through a variety of formal and informal mechanisms over the course of a career for those in HEIs and in IROs. Formal training (internal and external) and informal ways of learning via self-teaching and peer teaching exist together, have interdependencies and often play different roles in a researcher's development journey. There are limitations and benefits to each.

   > The AHRC policies and interventions focused on skills development in the community should value and work to support both formal and informal development mechanisms, i.e., the provision of high-quality training courses, as well as research practices and communities of practice that facilitate informal learning.

6. **Barriers to formal training** - Formal training is predominantly experienced through AHRC's training grants, as well as introductory courses provided by institutions (indeed these may intersect). It would appear to be the optimal mechanism for delivering data management and sustainability skills and knowledge. Beyond the doctoral research context, researcher capacity for undertaking formal courses is increasingly constrained by a range of structural barriers including funding, workload, and by the challenges of identifying a specifically focused course that will address specific research problems.

   > The AHRC policies relating to doctoral training should facilitate the provision of a baseline of research skills, knowledge and understanding of digital data management and sustainability infrastructure and its wider relevance for doctoral students. What that baseline is and how it might be delivered would require a focused study to explore existing training grant provision and the needs of a range of stakeholders in the community, including prospective employers. Importantly, any baseline would need to be flexible enough to be tailored to the research needs of particular disciplines and subgroupings.

   > The AHRC policies and interventions relating to formal training should find ways to promote, inform and identify quality and expertise, as well as reward and explicitly value formal training at all career stages.

7. **Informal skills and knowledge development** - Informal training takes place via self- and peer teaching. It is partly supported by communities of practice, which facilitate the spread of new ideas and in turn create needs for further training and knowledge. Informal mechanisms also provide flexible and focused ways to acquire particular skills and knowledge in the context of overburdened workloads amongst researchers.

   > In addition to supporting communities of practice (see point 4), the AHRC policies should acknowledge, encourage and reward informal skills development through recognising the value of such skills in research careers and supporting and promoting cross-disciplinary work, dissemination, contribution and dialogue that enables learning and knowledge exchange.

8. **Data management and sustainability** - These elements are not yet fully embedded in research processes, and there is a lack of knowledge and understanding of both the broad rationale for open research and the institutional infrastructures that support it. In addition, those infrastructures may not be functioning effectively at institutional levels.

   > The AHRC should promote sustainability strategies, build these into application and review processes, audit and monitor funded research whilst also understanding the challenges researchers face at an institutional level. These processes should ensure flexibility for researchers

to decide if and how their data should be sustained and provide a rationale for their approach.

> There is a need to raise researcher awareness of existing external infrastructure to support sustainable and open data, and encourage knowledge exchange between infrastructure providers and researchers.

9. **Supporting research careers and development** - Training grant schemes provide firm foundations for developing skills. However, intensification of academic workload and institutional and research cultures which continue to value lone scholar models, publication metrics and standard academic roles and career trajectories also fail to support (or may even exclude) ECRs, research technical professionals and indigenous or non-mainstream skills and knowledge (including digital).

> The AHRC policies and interventions should ensure support for an inclusive range of skills and knowledge and research roles in funding opportunities, application guidance and review processes. Encouraging the inclusion of RSEs, librarians and other research technical professionals in a research team is one element of this and is already supported by the AHRC policies such as the Technician Commitment[39] which UKRI is a signatory of.

---

39 Hello | Technicians and UKRI champions technical careers in research and innovation – UKRI

APPENDICES

APPENDIX 1: SURVEY ETHICS (CONSENT AND PIS)

Appendix 1A

UNIVERSITY OF
Southampton

REQUEST FOR CONSENT

Study title: Shaping AHRC Data & Software Policy [AHRC Software and Data Loss in the Arts & Humanities].
Researcher name: Johanna Walker, Rebecca Taylor, Simon Hettrick
ERGO number: 65039

Participant Identification Number (if applicable):

The participant will read the following statements in the introduction to the online survey. They will indicate consent by clicking NEXT and proceeding to the survey. By leaving the survey at this or any point before completion they will rescind consent.

"By clicking NEXT, you consent to the following:

I have read and understood the information and have had the opportunity to ask questions about the study (by emailing Rebecca Taylor).	
I agree to take part in this research project and agree for my data to be used for the purpose of this study.	
I agree that my anonymised data can be made available in an appropriate repository for the purpose of future research on the arts and humanities research community.	
I understand my participation is voluntary and I may withdraw at any time for any reason without my participation rights being affected by providing my Participant ID which I will receive at the end of the survey	
I agree with the above statements and wish to continue	

[Date] [Version Number]   [Ethics/IRAS reference *(if applicable)*]

UNIVERSITY OF Southampton

Participant Information Sheet

Study Title: Shaping AHRC Data and Software Policy

Researcher: Dr Rebecca Taylor, Dr Johanna Walker & Professor Simon Hettrick
ERGO number: 65039

What is this research about?

As a member of the Arts and Humanities community we would like to invite you to participate in a survey to share your perspective on data and software practices.   To help shape policies and practices on data and software and ensure these are fit for purpose, the AHRC have asked us to investigate views on data and software knowledge and skill amongst researchers and others involved in arts and humanities research. A link to our webpage is available here. (https://engagementhub.ukri.org/community/ahrcsoftwareanddata/)

We would like to invite you to complete the attached survey, which we estimate will take 10 minutes. We will ask you about data and software practices, skills and support. It's crucial to our study to gather input from the widest community possible, in order to understand the broadest range of views and experiences, so we are keen to hear from you even if you feel you do not use much data or software in your research. We would also encourage you to forward this survey to as many of your colleagues as you wish.

Do I have to take part?
To help you decide whether you would like to take part, or not, in the above research study, it is important that you understand why the research is being performed and what it will involve. Please read the information below carefully and ask questions about anything that is not clear, or get in touch if you would like more information before you decide to take part in this research. The survey is not compulsory and, even if you decide to participate, you can withdraw at any time. You will be asked to confirm that you are happy to participate by clicking the "next" button below. We hope you will be keen to participate and help shape the way AHRC understands, supports and funds research practice in the future.

Why have I been asked to participate?
You have received this link because you were identified as a researcher in the arts and humanities community or because you have participated in a research project funded by an AHRC grant.

What will happen to me if I take part?
You will be asked a series of questions about you, your research and your experience of using data and, where relevant, software, in your research.

Are there any benefits in my taking part?

The data collected by this survey will be used to help the AHRC to develop policies that support key data and software infrastructure for research and support the building of skills and capabilities within the community. As a thank you, you can choose to be entered into a prize draw for a £100 Love2Shop voucher, just enter your email address at the end of the survey (it will not be stored with the rest of your data).

[Date] [Version number]   [Ethics/IRAS number *(if applicable)*]

UNIVERSITY OF Southampton

Are there any risks involved?
We know of no risks involved in taking part in this survey.

What data will be collected?
The survey will take around 10 minutes and will ask you about :
- your area of research, your institution and some demographic details,
- the research you do, the data you collect and how you analyse it, manage it and store it
- Any software you use or develop in relation to your research
- the skills necessary to support your research

Will my participation be confidential?
Your participation and the information we collect about you during the course of the research will be kept strictly confidential. Only members of the research team and responsible members of the University of Southampton may be given access to data about you for monitoring purposes and/or to carry out an audit of the study to ensure that 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 only collect your name or contact details for the purpose of administering the prize draw or if you wish to be included in future phases of the research and this is removed and stored separately to the data. The anonymised data will be made available under an appropriate licence (TBC) via the University of Southampton's Pure repository.

What happens if I change my mind?
You can change your mind and withdraw at any time by just closing the survey window. If you wish to withdraw from the survey entirely, you can select "Exit and clear survey" at the top right of the survey window, and all the information you have provided will be deleted. To remove your data from our records after completing the survey you must make a note of your participant ID as we need that to identify which record to remove.

What will happen to the results of the research?
The results will be published in a report for the AHRC, parts of which will be made publicly available. The anonymised data will be made available under an appropriate licence (TBC) via the University of Southampton's Pure repository.

Where can I get more information?
If you want more information, you can contact the lead researcher, Rebecca Taylor. (r.taylor@soton.ac.uk). You can learn more about the research at https://engagementhub.ukri.org/community/ahrcsoftwareanddata/.

What happens if there is a problem?
If you have a concern about any aspect of this study, please contact Rebecca Taylor (r.taylor@soton.ac.uk).
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 (023 8059 5058, 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.

[Date] [Version number]   [Ethics/IRAS number *(if applicable)*]

UNIVERSITY OF
Southampton

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 at http://www.southampton.ac.uk/assets/sharepoint/intranet/ls/Public/Research%20and%20Integrity%20Privacy%20Notice/Privacy%20Notice%20for%20Research%20Participants.pdf

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 supporting our research.

[Date] [Version number]   [Ethics/IRAS number *(if applicable)*]

APPENDIX 2: INTERVIEW ETHICS (CONSENT AND PIS)

Appendix 2A

UNIVERSITY OF Southampton

CONSENT FORM

Study title: Shaping AHRC Data & Software Policy [AHRC Software and Data Loss in the Arts & Humanities].
Researcher name: Johanna Walker, Rebecca Taylor, Simon Hettrick
ERGO number: 65039
Participant Identification Number (if applicable):

The researcher will ask the participant to verbally confirm their agreement at the beginning of the focus group or interview. This will be recorded and then transcribed.

We will read the following statements to you. Please indicate your agreement or otherwise verbally.

I have read and understood the information sheet (*18/01/22*) 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 give permission for the pseudonymised transcript of this focus group or interview to be deposited in the researcher's data space (UoS SharePoint) and repository (Pure) as described in the Participant Information Sheet so it can be used for future research and learning on the AHRC community.	
I understand my participation is voluntary and I may withdraw (at any time) for any reason without my participation rights being affected.	
I understand that I can withdraw from this study up to a week after the date of the focus group or interview. I understand that should I withdraw from the study after this date, 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.	
I understand that I may be quoted directly in reports of the research but that I will not be directly identified (e.g. that my name will not be used and identifiable details will be changed).	

[Date] [Version Number]   [Ethics/IRAS reference (*if applicable*)]

Southampton

UNIVERSITY OF

Participant Information Sheet

Study Title: Shaping the AHRC's Policy on Data & Software [AHRC Software and Data Loss in Arts and Humanities Research]
Researcher: Rebecca Taylor, Simon Hettrick, Johanna Walker
ERGO number: 65039

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 will be asked to sign a consent form.

What is the research about?

Collecting, storing and reusing data are essential practices for researchers across disciplines. The Arts and Humanities Research Council wants to make sure that its policies and practices on data and the software used to extract, manipulate and store it, are fit for purpose now and in the future. To help shape these policies, the Research Council have asked us to investigate views on data and software knowledge and skill amongst researchers and others involved in arts and humanities research. In order to do this, we would like to talk to you to help us ==discover more about skills gaps, ways to address those gaps and the pros and cons of possible policy scenarios.==
These questions have been informed by a large scale community wide survey held in early 2022.

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 hope you will be keen to help shape the way AHRC understands, supports and funds research practice in the future.

What will happen to me if I take part?

If, after reading this information sheet and the attached consent form, you are happy to take part in the study, we will invite you to ==an interview of 60 minutes==. You will be asked about your ==views on the skills required to sustainably use data and software in arts and humanities research, and how those skills can be developed and supported by== institutions and funders including AHRC, and your institution. See the section below on what will be collected for more detail.

We will record the interview (including verbal consent), transcribe it pseudonymously and destroy the audio recording. The focus group data will be analysed using Qualitative data analysis software and used to inform the design of the survey. This will be stored during the research period on the University of Southampton SharePoint. The findings will be disseminated in a report to the AHRC. The AHRC will not know who we have talked to and any quotes used in the reporting will be anonymised. In line with best practice, we'll ask you for consent for your pseudonymised contribution to be stored in a repository (currently Pure) at the University of Southampton and available for future Arts and Humanities research.

You will also receive a £30 ==Love2Shop== voucher as thanks for your participation.

[Date] [Version number]   [Ethics/IRAS number *(if applicable)*]

UNIVERSITY OF Southampton

Are there any benefits in my taking part?

By taking part:
- You'll help us validate research data we have collected on skills and knowledge gaps in data and software in the arts and humanities
- You'll help us identify how the gaps in training, skills and knowledge in data and software manipulation and management can be addressed, so that appropriate provision can be made;
- You'll help us make sure that AHRC are able to devise data policies and training provision that meet the needs and priorities of everyone who is eligible for and a beneficiary of their funding, not just the more digitally focused.

Are there any risks involved?

There are no risks involved.

What data will be collected?

We will ask semi-structured questions on the following topics:

- About you: your area of research expertise; recent experience of AHRC funding (if any)
- Your personal experience of doctoral/professional training
- Your post phd/ professional training skills acquisition
- 'Missing' skills that you have not had the opportunity to acquire
- Your perceptions of skills amongst peers, colleagues and research teams
- Your perceptions of institutional infrastructure that supports skills development
-
- Policy/practice questions
- We will present some options for closing the skills gap and ask for your responses to them. These are in two forms:
- AHRC as a 'skills hub' - responses to 5 possible policy scenarios
- Supporting Institutions – responses to 2 possible policy scenarios

Will my participation be confidential?

Your participation and the information we collect about you during the course of the research will be kept strictly confidential. Transcripts and outputs will be pseudonymised. No-one will be identifiable in the final research outputs.

Only members of the research team and responsible members of the University of Southampton may be given access to data for monitoring purposes and/or to carry out an audit of the study to ensure that 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.

Do I have to take part?

[Date] [Version number]   [Ethics/IRAS number *(if applicable)*]

Participation is entirely voluntary, and you are free to withdraw at any time –we greatly value your participation and hope you will agree to take part and contribute your views and shape the future of AHRC funding.

What happens if I change my mind?

You have the right to change your mind and withdraw at any time without giving a reason and without your participant rights being affected.

If you withdraw from the study within a week of the focus group or interview being conducted, we will be able to remove your data. If you withdraw from the study more than a week after the focus group has been conducted, we will not be able to remove your data and it will remain in the study for the purposes of achieving the objectives of the study only. However, your pseudonymised transcript will not be deposited for future research.

What will happen to the results of the research?

The results of the research will be presented to the funder (AHRC), and used to inform future policy on data and software practices. They may also be subsequently written up for conference or journal publication.

Your personal details will remain strictly confidential. Research findings made available in any reports or publications will not include information that can directly identify you.

Where can I get more information?

You can email me at johanna.walker@soton.ac.uk.

What happens if there is a problem?

If you have a concern about any aspect of this study, in the first instance you should speak to the researchers named at the top of the page, and we 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 (023 8059 5058, 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.

[Date] [Version number]   [Ethics/IRAS number *(if applicable)*]

UNIVERSITY OF
Southampton

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 at
http://www.southampton.ac.uk/assets/sharepoint/intranet/ls/Public/Research%20and%20Integrity%20Privacy%20Notice/Privacy%20Notice%20for%20Research%20Participants.pdf

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).

Data will be pseudonymised - it will only be possible to tell from the transcript that a person was speaking but not whom. Individuals will not be re-identifiable.

Thank you.

[Date] [Version number]   [Ethics/IRAS number *(if applicable)*]

APPENDIX 3: SURVEY QUESTIONS

Shaping AHRC policy on data and software

- **Study Title**: Shaping AHRC Data and Software Policy
- **Researcher**: Dr Rebecca Taylor, Dr Johanna Walker & Professor Simon Hettrick
- **ERGO number: 65039**

Overview

Who commissioned this research?

The Arts and Humanities Research Council (AHRC) commissioned this research. The anonymised results from this survey will be added to an evidence base which will help direct AHRC policy on software and data in the future.

How long does the survey take?

The survey takes around 10 minutes to complete.

What is this research about?

As a member of the Arts and Humanities community we would like to invite you to participate in a survey to share your perspective on data and software practices. To help shape policies and practices on data and software and ensure these are fit for purpose, the AHRC have asked us to investigate views on data and software knowledge and skill amongst researchers and others involved in arts and humanities research. A link to our webpage is available here (https://engagementhub.ukri.org/community/ahrcsoftwareanddata).

We would like to invite you to complete the attached survey, which we estimate will take 10 minutes. We will ask you about data and software practices, skills and support. It's crucial to our study to gather input from the widest community possible, in order to understand the broadest range of views and experiences, so we are keen to hear from you even if you feel you do not use much data or software in your research. We would also encourage you to forward this survey to as many of your colleagues as you wish.

Participant information

Do I have to take part?

To help you decide whether you would like to take part, or not, in the above research study, it is important that you understand why the research is being performed and what it will involve. Please read the information below carefully and ask questions about anything that is not clear, or get in touch if you would like more information before you decide to take part in this research. The survey is not compulsory and, even if you decide to participate, you can withdraw at any time. You will be asked to confirm that you are happy to participate by clicking the "next" button below. We hope you will be keen to participate and help shape the way AHRC understands, supports and funds research practice in the future.

Why have I been asked to participate?

You have received this link because you were identified as a researcher in the arts and humanities community or because you have participated in a research project funded by an AHRC grant.

What will happen to me if I take part?

You will be asked a series of questions about you, your research and your experience of using data and, where relevant, software, in your research.

Are there any benefits in my taking part?

The data collected by this survey will be used to help the AHRC to develop policies that support key data and software infrastructure for research and support the building of skills and capabilities within the community. As a thank you, you can choose to be entered into a prize draw for a £100 Love2Shop voucher, just enter your email address at the end of the survey (it will not be stored with the rest of your data).

Are there any risks involved?

We know of no risks involved in taking part in this survey.

What data will be collected?

The survey will take around 10 minutes and will ask you about :

- your area of research, your institution and some demographic details,
- the research you do, the data you collect and how you analyse it, manage it and store it
- Any software you use or develop in relation to your research
- the skills necessary to support your research

Will my participation be confidential?

Your participation and the information we collect about you during the course of the research will be kept strictly confidential. Only members of the research team and responsible members of the University of Southampton may be given access to data about you for monitoring purposes and/or to carry out an audit of the study to ensure that 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 only collect your name or contact details for the purpose of administering the prize draw or if you wish to be included in future phases of the research and this is removed and stored separately to the data. The anonymised data will be made available under an appropriate licence (TBC) via the University of Southampton's Pure repository.

What happens if I change my mind?

You can change your mind and withdraw at any time by just closing the survey window. If you wish to withdraw from the survey entirely, you can select "Exit and clear survey" at the top right of the survey window, and all the information you have provided will be deleted. To remove your data from our records after completing the survey you must make a note of your participant ID as we need that to identify which record to remove.

What will happen to the results of the research?

The results will be published in a report for the AHRC, parts of which will be made publicly available. The anonymised data will be made available under an appropriate licence (TBC) via the University of Southampton's Pure repository.

Where can I get more information?

If you want more information, you can contact the lead researcher, Rebecca Taylor. (r.taylor@soton.ac.uk). You can learn more about the research at https://engagementhub.ukri.org/community/ahrcsoftwareanddata/.

What happens if there is a problem?

If you have a concern about any aspect of this study, please contact Rebecca Taylor (r.taylor@soton.ac.uk).

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 (023 8059 5058, rgoinfo@soton.ac.uk (mailto: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 (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 (http://www.southampton.ac.uk/assets/sharepoint/intranet/ls/Public/Research%20and%20Integrity%20Privacy%20Notice/Privacy%20Notice%20for%20Research

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 (mailto:data.protection@soton.ac.uk)).

Thank you for supporting our research.

- I have read and understood the information and have had the opportunity to ask questions about the study (by emailing Rebecca Taylor).
- I agree to take part in this research project and agree for my data to be used for the purpose of this study.
- I agree that my anonymised data can be made available in an appropriate repository for the purpose of future research on the arts and humanities research community.
- I understand my participation is voluntary and I may withdraw at any time for any reason without my participation rights being affected by providing my Participant ID which I will receive at the end of the survey.

Please indicate that you agree with the above statements by clicking the "Next" button in the bottom right of the page.

There are 28 questions in this survey.

About you

The following information helps us to understand whether our survey is representative of the Arts and Humanities community.

The data from this survey will be anonymised before it is shared.

> ### Which institution/organisation do you work for? *
>
> Please write your answer here:
>
> [   ]

In which disciplines do you work? *

❶ Check all that apply
Please choose **all** that apply:

- ☐ Archaeology
- ☐ Classics
- ☐ Cultural and Museum Studies
- ☐ Dance
- ☐ Design
- ☐ Development Studies
- ☐ Digital Humanities
- ☐ Divinity and Religion
- ☐ Drama and Theatre Studies
- ☐ History
- ☐ Information and Communication Technologies
- ☐ Languages and Literature
- ☐ Law and Legal Studies
- ☐ Library and Information Studies
- ☐ Linguistics
- ☐ Media
- ☐ Music and Visual Arts
- ☐ Philosophy
- ☐ Political Science and International Studies
- ☐ Theology
- ☐ Other: _____

These are taken from the Primary Research Areas covered by AHRC: https://ahrc.ukri.org/funding/research/subjectcoverage/ahrc-disciplines/. Use the 'Other' option if your discipline is not listed. Please tick all that apply.

To which of the following organisations have you applied for funding?

❶ Check all that apply
Please choose **all** that apply:

- ☐ AHRC
- ☐ Andrew Mellon Foundation
- ☐ Arts Council England
- ☐ BBSRC
- ☐ British Academy
- ☐ Charitable/Philanthropic society
- ☐ Creative Scotland
- ☐ EPSRC
- ☐ ESRC
- ☐ European Research Council
- ☐ Heritage Lottery Fund
- ☐ Humanities in the European Research Area
- ☐ Innovate UK
- ☐ Leverhulme Foundation
- ☐ MRC
- ☐ My institution(s)
- ☐ NERC
- ☐ Scholarly Societies (e.g. Royal Historical Society, Past & Present Society)
- ☐ STFC
- ☐ Wellcome Trust
- ☐ Wolfson Foundation
- ☐ Other: _____

74

What is your career stage? *

❶ Choose one of the following answers
Please choose **only one** of the following:

○ Phase 1 - Junior (e.g. PhD candidate)

○ Phase 2 - Early (e.g. Research Assistant/Associate, first grant holder, Lecturer)

○ Phase 3 - Mid / Recognised (e.g. Senior Lecturer, Reader, Senior Researcher, Associate Professor)

○ Phase 4 - Established / Experienced / Senior (e.g. Professor, Director of Research)

If the choices above are not clear, please take a look at https://www.leru.org/files/UK-England.pdf and choose the appropriate phase that matches your career stage.

Your research data

For the purposes of this survey, we define "research data" as "data of any type generated during the conduct of your research".

We would like you to answer the following questions in relation to your research conducted over the last five years.

The data from this survey will be anonymised before it is shared.

Over the last five years, what types of research data have you or your team worked with? *

❶ Check all that apply
Please choose **all** that apply:

☐ Audio
☐ Image
☐ Numerical
☐ Physical artefacts
☐ Textual
☐ Video

☐ Other: [   ]

Over the last five years - once you **have completed** a research project/study - which of the following options have you used to store the data that were generated? *

❶ Check all that apply
Please choose **all** that apply:

☐ On a laptop or desktop computer
☐ In a shared area provided by yourself (dropbox, Google Drive, etc.)
☐ In a shared area provided by your institution (sharepoint, shared drive, institutional Google Drive, etc.)
☐ In a data repository provided by your instution (Pure, ePrints, etc.)
☐ In a data repository outside of your institution (UK Data Service, Zenodo, etc.)

☐ Other: [   ]

If you selected more than one choice above, which of the following options describes your **typical approach** to storing the data that were generated? *

❶ Choose one of the following answers
Please choose **only one** of the following:

○ On a laptop or desktop computer

○ In a shared area provided by yourself (Dropbox, Google Drive, etc.)

○ In a shared area provided by your institution (sharepoint, shared drive, institutional Google Drive, etc.)

○ In a data repository provided by your institution (Pure, ePrints, etc.)

○ In a data repository outside of your institution (UK Data Service, Zenodo, etc.)

○ Not applicable

Over the last five years, which of the following statements best applies to your research data? *

❶ Choose one of the following answers
Please choose **only one** of the following:

○ I haven't shared my research data

○ I have shared my research data, but I rarely do so

○ I occasionally share my research data

○ I regularly share my research data

○ I do not have research data to share

Over the last five years, what best describes your experience of data management plans on your projects? *

❶ Choose one of the following answers
Please choose **only one** of the following:

○ I haven't been involved in research that included a data management plan

○ In my experience, data management plans are not followed

○ In my experience, data management plans are followed in part

○ In my experience, data management plans are mostly followed

○ In my experience, data management plans are followed exactly

---

Over the last five years, which of the following practices have been part of your standard research process?

❶ Check all that apply
Please choose **all** that apply:

☐ I license my data to allow it to be shared

☐ I share my data with individuals/groups that have requested access

☐ I create a unique identifier (e.g. a DOI) to make my data findable

☐ I cite my research data in publications

Your research software

"Research software" is any software that you have used to collect data, assist in analysis or produce a result that you intend to appear in a publication or other output from your research.

The data from this survey will be anonymised before it is shared.

Over the last five years, which of the following have applied to your research? *

❶ Check all that apply
Please choose **all** that apply:

☐ I have not used research software

☐ I have used research software provided by university

☐ I have purchased specialist research software with research funding

☐ I have used specialist research software which was provided for free

☐ I have used research software that was developed for me

☐ I have developed my own research software

☐ Other: _____

Over the last five years, **at the end** of a research project/study, where have you typically stored the research software that was created?

*

Only answer this question if the following conditions are met:
((soft5_SQ002.NAOK (/admin/questions/sa/view/surveyid/438226/gid/1572/qid/102947) == "Y") or (soft5_SQ005.NAOK (/admin/questions/sa/view/surveyid/438226/gid/1572/qid/102947) == "Y"))

❶ Choose one of the following answers
Please choose **only one** of the following:

◯ On a laptop/desktop computer

◯ In a shared area provided by your institution (sharepoint, shared drive, Institutional Google Drive, etc)

◯ In an institutional software repository (e.g. Gitlab, SVN, etc.)

◯ In a repository outside my institution (e.g. Github, Mercurial, etc.)

◯ Other [   ]

Over the last five years, which of the following practices have been part of your standard research process?

Only answer this question if the following conditions are met:
((soft5_SQ002.NAOK (/admin/questions/sa/view/surveyid/438226/gid/1572/qid/102947) == "Y") or (soft5_SQ005.NAOK (/admin/questions/sa/view/surveyid/438226/gid/1572/qid/102947) == "Y"))

❶ Check all that apply
Please choose **all** that apply:

☐ I do not share my research software

☐ I license my research software to allow to be shared

☐ I share my research software with individuals/groups that have requested access

☐ 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

---

Who develops your research software? *

Only answer this question if the following conditions are met:
((soft5_SQ002.NAOK (/admin/questions/sa/view/surveyid/438226/gid/1572/qid/102947) == "Y") or (soft5_SQ005.NAOK (/admin/questions/sa/view/surveyid/438226/gid/1572/qid/102947) == "Y"))

❶ Check all that apply
Please choose **all** that apply:

☐ I develop my own research software

☐ PhD students and postdoctoral researchers

☐ Software developers from industry

☐ Research Software Engineers from within academia

☐ Other: [   ]

Your skills

We would like to understand your skills with data and software and investigate whether you feel there are gaps relating to the collection, analysis, management and sharing of data and software.

The data from this survey will be anonymised before it is shared.

Over the last five years, have you learned new skills or knowledge in relation to any of the following? *

❶ Check all that apply
Please choose **all** that apply:

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

Over the last five years, have you learned new skills or knowledge in relation to any of the following? *

❶ Check all that apply
Please choose **all** that apply:

- [ ] Software requirements collection (e.g. landscape review, usage study)
- [ ] Software development (e.g. new languages, design, testing)
- [ ] Software sharing (e.g. licensing, repositories, distribution, deployment)
- [ ] None of the above

Where did you acquire your skills and knowledge about data and/or software? *

❶ Check all that apply
Please choose **all** that apply:

- [ ] Masters/Doctoral training (completed during PGR and PGT courses)
- [ ] Training/advice provided in-house by my institution (while employed)
- [ ] Training/advice provided externally (academic and commercial providers e.g. NCRM, Turing, etc.)
- [ ] 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)
- [ ] Other: _____

Thinking about your current skill set, what is the main data or software skill that you would like to obtain to improve your research? *

Please write your answer here:

_____

Do any of the following issues prevent you from gaining the skill you entered above? *

Only answer this question if the following conditions are met:
Answer was NOT at question '18 [skills5]' (Thinking about your current skill set, what is the main data or software skill that you would like to obtain to improve your research?)

❶ Check all that apply
Please choose **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 with them
- [ ] 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
- [ ] None of the above
- [ ] Other: _____

78

Are there any further skills that you would like to obtain to improve your research? (Please separate your answers with semicolons.)

Only answer this question if the following conditions are met:
Answer was NOT at question '18 [skills5]' (Thinking about your current skill set, what is the main data or software skill that you would like to obtain to improve your research?)

Please write your answer here:

Research policies and responsibilities

Please choose the appropriate response for each item:

	1 (Not at all confident)	2	3	4	5 (Very confident)
Are you confident that you understand the policies and responsibilities related to research data that are placed on you during a project/study?	○	○	○	○	○
Are you confident that you understand the policies and responsibilities related to research data that are placed on you data after a project/study has ended?	○	○	○	○	○
Are you confident that you understand the policies and responsibilities related to research software that are placed on you during a project/study?	○	○	○	○	○
Are you confident that you understand the policies and responsibilities related to research software that are placed on you data after a project/study has ended?	○	○	○	○	○

More about you

Please help us to understand more about the AHRC's research community and our respondents. All of the following questions are non-mandatory.

The data from this survey will be anonymised before it is shared.

What best describes your gender?

❶ Choose one of the following answers
Please choose **only one** of the following:

○ Female
○ Male
○ Prefer to self describe
○ Prefer not to answer

How do you describe your gender?

Only answer this question if the following conditions are met:
Answer was 'Prefer to self describe' at question '22 [about2]' (What best describes your gender?)

Please write your answer here:

How would you describe your ethnic origin? This refers to people who share the same cultural background and identity, not country of birth or nationality.

❶ Choose one of the following answers
Please choose **only one** of the following:

○ White: English / Welsh / Scottish / Northern Irish / British
○ White: Irish
○ White: Gypsy or Irish Traveller
○ White: Any other White background
○ Mixed / Multiple Ethnic groups: White and Black Caribbean
○ Mixed / Multiple Ethnic groups: White and Black African
○ Mixed / Multiple Ethnic groups: White and Asian
○ Mixed / Multiple Ethnic groups: Any other Mixed / Multiple ethnic background,
○ Asian / Asian British: Indian
○ Asian / Asian British: Pakistani
○ Asian / Asian British: Bangladeshi
○ Asian / Asian British: Chinese
○ Asian / Asian British: Any other Asian background
○ Black / African / Caribbean / Black British: African
○ Black / African / Caribbean / Black British: Caribbean
○ Black / African / Caribbean / Black British: Any other Black / African / Caribbean background
○ Other Ethnic group: Arab
○ Other Ethnic group: Any other ethnic group
○ Do not wish to declare
○ Other [   ]

Do you identify as disabled?

❶ Choose one of the following answers
Please choose **only one** of the following:

○ Yes
○ No
○ Unknown
○ Prefer not to answer

Prize draw and further study

The data from this survey will be anonymised before it is shared.

Is there anything else you would like to add?

Please write your answer here:

[   ]

Leave your email address below if you would like to be included in our prize raw for £100 Love2Shop vouchers.

❶ Please check the format of your answer.
Please write your answer here:

[   ]

80

Leave your email address below if you're happy for us to get in touch to further discuss your thoughts on data, software and research skills.

❶ Please check the format of your answer.

Please write your answer here:

[   ]

08.03.2022 – 23:59

Submit your survey.
Thank you for completing this survey.

APPENDIX 4: INTERVIEW TOPIC GUIDE

Topic guide for Interviews

1. *Background*
- Explain purpose of research and interviews specifically
- Summary of types of question to be asked,
- Reiterate ethics processes right to withdraw
- Explain how they will get their voucher
- Any questions
- Read through consent questions and ask for individuals to respond

I'm going to record this session, and as we need to record your verbal consent, I'll turn the recording on now. The recording will be used to create a pseudonymised transcript.

So today, we're going to have a 45 minute conversation, which will take us through to [time + 45mins]. Hopefully you have some idea of what we will cover from our previous communication, but just to go over it: first of all we will ensure we have your consent to collect data; then we will dive straight into the questions beginning with your background, then some questions about skills, and finishing up with some questions about possible policy interventions. Are you happy with that?

Obviously, we would like to thank you for your participation today, and we'll email you a Love2Shop voucher for £30 in the next couple of weeks.

CONSENT

We're going to collect your consent verbally today. I'm going to read some statements and if you are happy to give your consent, please just say yes after each statement. If you're not happy at any point, you can say no and withdraw from the interview. Please do ask if you're not clear on any of the consent statements. We have six statements, because we need you to consent to both current and future use.

1. The first statement is that you have read and understood the participant information sheet and have had the opportunity to ask questions about the study. You should have received a copy of the participant information sheet and a consent form with your original email, and I'm showing it on my screen now.
   - Can you indicate verbally that you agree or identify if you don't
2. Next, state that you agree to take part in this research project and that you agree for your data to be used for the purpose of this study.
   - y/N
3. Now we ask you to state that you give permission for the pseudonymised transcript of this interview to be deposited in the researcher's data space (UoS SharePoint) and repository (Pure) as described in the Participant Information Sheet so it can be used for future ethically approved research and learning on the Arts and Humanities research community. This is so there is explicit consent for the data to be reused.

- Can you indicate verbally that you agree or identify if you don't

4.   Next statement is,you understand your participation is voluntary and you can withdraw (at any time) for any reason without my participation rights being affected.

- Can you indicate verbally that they agree or identify if they don't

5.   The next statement is related, which is that you understand that you can withdraw from this study up to a week after the date of the interview. You understand that should you withdraw from the study after this date, then the information collected about you up to this point may still be used for the purposes of achieving the objectives of the study only.

- Can you indicate verbally that you agree or identify if you don't

6.   Next, you understand that you may be quoted directly in reports of the research but that you will not be directly identified (e.g. that your name will not be used and identifiable details will be changed).

- Can you indicate verbally that you agree or identify if you don't

1. *Background*
- **Tell me a bit about your work**
    - Institution, discipline, role, area of research interest, recent projects
    - what does Data mean to you in your work? And do you use software?
    - Probe on research funding to support this work

- We're interested in whether there are barriers that particularly affect people with certain experiences and who identify as having certain characteristics. We know, for instance, that disability, ethnicity and gender can create barriers. Do you feel any of these are relevant to you?

2. *Skills questions*
1. **Tell me about your doctoral/professional training**?
    - New skills (data and/or software), embedded?, gained in formal training /other forms of learning? (community of practice - peers, supervisors, conferences etc)
2. **What new skills have you gained since** you completed your phd/ professional training?
    - Which tools? - data collection, data analysis data management software? (aim for a list and then with each one explore the process below)
    - Why? what motivated this? How do you decide you needed this? what was the process to get that training/gain that skill?
    - How did you find training – external/internal? , or other routes to learning (Community of practice: peers, supervisors, conferences etc..)
3. **Are there skills you wanted to gain but didn't**
    - or skills which would have been useful with hindsight ?
    - Why didn't you? What were the problems? (time /culture /cost /availability /awareness) [*AHRCs barriers list*]
    - *Probe academic culture, Work intensification, value of training*
4. **Thinking about now?**
    - Are there any skills you think would be useful in your current research?

- How do you know about them? How would you go about gaining that skill/those skills? Where would you look for training courses?
- Would you use other mechanisms? Community of practice: peers, supervisors, conferences etc..)
- Would you prefer to collaborate/cost in colleagues/experts within your institution (e.g. RSE groups, data experts) or gain the skills yourself?
- What things are likely to prevent you gaining the skills you need ?
- Thinking about recent research projects, what skills do you rely on others for? Why?

5. **Thinking more broadly about your peers and colleagues and research teams**
   - Do you have a sense of what data and software expertise they have?
   - Where do you see yourself in comparison to those you work with in terms of your skills in data collection data analysis data management, software etc?
   - Do you see yourself as part of a community of practice in relation to data and software?
   - How is your research community informing your knowledge of what (digital) skills you might need?

6. **Thinking about the institutional infrastructure that supports skills development**
   - How useful is your institution for providing the access to training skills and knowledge you need for your research – what is available? what works? what are the problems? (time/culture/cost/availability/awareness)
   - Probe formal training infrastructure, communities of practice methodological software, disciplinary communities.
   - What is your understanding of the AHRC's role in skills development?
   - In what ways do AHRC's research funding structures (response mode, calls research centres fellowships) encourage skills development?
   - Would you go to the AHRC for guidance about skills? Signposting, resources, information?
   - How do you think the AHRC should promote and engage the community in new approaches and tools? Should this come from UKRI? Cross council initiatives ? Individual councils?

7. *Policy /practice questions*
   We would like your thoughts on various scenarios for supporting skills development and knowledge around data and software
   1. *AHRC as a 'skills hub'* -
      - **AHRC Online Skills hub** -
        - **Website/Hu**b managed site detailing data and software skills the council is keen to promote and linking to preferred training suppliers and groups (i.e. RSE groups, NCRM, Turing, SSI, SDSA? Open Data Institute? Software carpentry)
        - **AHRC inhouse fellowship** – person will run the hub and other skills focused events, seminars, conferences, summer schools…(see for example the ESRCs Research methods festival) promoting methods, data and software skills to the community.
   2. *AHRC funding initiatives*

- **Update AHRC calls and bid process** – Improve prominence of Software and Data in **doctoral training requirements** - Data literacy and data management, ethics, workflow training for all doctoral students? Include software and data experts on writing calls panels and review teams strengthening links to the key orgs, embedding RSEs in bid process?
- **Data and skills knowledge exchange funding** – AHRC run a light touch competition to award conference bursaries aimed at recognised AHRC research community conferences for papers or training sessions that focus specifically on data and software practice, methodological/research innovation and development examples from recent and current research projects. What conference woud be appropriate in your discinple?
- **Academic Fellowship schemes** a) new scheme with a skills focus (RSE/AHRC fellowship; Data Skills Fellowships? Arts & Technology Fellowships for arts researchers (arts plus, rather than dropping technologist in) **b)** Embed or ringfence skills focus within an existing scheme (Post doc/future leaders etc). Fellows would promote skills development in a particular area - data collection, data manipulation, data management (repositories etc), ethics, software, research workflow etc building their own skills, talking to various institutional stakeholders (librarians etc) helping to streamline practice and disseminating at the institutional levels through seminars, advising colleagues, and institutional/disciplinary communities of practice

8. **Other policy and practice**
   What can others do? What is not in the UKRI remit?
   - **Changing organisational culture work intensification...** rewarding training and skills development
   - **HEIs and ROIs Research infrastructure and support** - offering more training and support on navigating internal research infrastructure, (e.g. ethics, finance, bid submission, workflows, data management etc) And entry level technological training in software and technical tools. Support could be an online course (tick box) or more personal - someone sits next to you and helps you navigate a particular system (Agresso) or thing (DPAs) or even models how to do it (Training as a Service) – see Nottingham Trent example
   - **Accountability** and audit trail – At the beginning of the research process bidding, what's the best way to ensure data management plans are fit for purpose? during grant awarding process? What question to ask in applications and interviews? At the end, how could we audit the outcomes of projects and DPAs? what if DMPs actually also recorded the methodology? And were filed with the data? So they became an externally facing record?
   - **Workflow (research) -** How do we enable understanding of a workflow process, rather than just individual tools? How do we teach the research infrastructures, rather than the tools?
   - How do we identify the crucial data to sustain?
   - How do we ensure data accessibility from a commercial and sociological rather than a technical point of view?
   - How can we enable early adopters with strategic skills to share their knowledge and experience with others? What resource does this require?

- Is there anything you would take away from this interview and reviewing your own practice on data and software?

  - Is there anything you would take away from this interview and reviewing your own practice on data and software?

APPENDIX 5: DATA MANAGEMENT PLAN

AHRC Data and Software Policy and Use

Data Management Plan v3

Document Nature	Internal project management
Dissemination Level	Public
Delivery Date	May 2021
Description	Metadata on all data sets collected in the project, along with regulatory and ethical compliance information and use guidance.
Version	3
Review date	August 2021 June 2022
Keywords	Data management, data plan, data curation, data protection, compliance
Coordinator	Johanna.walker@soton.ac.uk
Contributions	Rebecca Taylor, Simon Hettrick

Table of Contents

Introduction

The Data Management Plan (DMP) is concerned with the data generated or collected during the commission of the research project Data and Software Loss in the Arts and Humanities.

This document defines how data collected and generated within AHRC Data and Software Policy and Use will be stored, managed and reported on. It is anticipated all the research data will be digital in format, although information may be collected about artefacts. As and when more datasets are added, the DMP will be updated to reflect these.

D&S Policy and Use project overview

The project addresses the evidence gap in order to understand how AHRC can best support and build the capacity of its research communities to establish and adopt best practice in digital data manipulation and digital asset management. It seeks to understand the current policy landscape across UKRI with regard to data management; to map the constituent groups of AH researchers and other stakeholders supporting the research and to understand the broad extent of data and software loss, wider research practices and specific areas where there are skills gaps.

Objectives of the Data Management Plan

- Providing information on the data used in the project to the researchers and funders;
- Meeting data policy standards of UKRI;
- Ensuring legal obligations of data controlling and processing of personal data under the GDPR are met;
- Ensuring ethical compliance is achieved in line with what is stated in the ERGO application;
- Assisting with recovering data if data loss is suffered;
- Enabling research data set sharing where applicable;
- Enabling research data set review where necessary;
- Enabling data citation where appropriate.
-

Personal data

There are three types of personal data collected by D&SL:

1. **Contact information** – i.e. names, email addresses and phone numbers – in order to communicate with (potential) survey and interview respondents and. The legal basis for which are (unless otherwise specified*): (i) the performance of a contract; and (ii) legitimate interests – to identify potential survey respondents. (Dataset DS-M2)

2. **Consent** – The legal basis for which are (unless otherwise specified*): (i) consent; (ii) legal obligation (to comply with ethics regulations). (Dataset DS-S2)

3. **Descriptions of data management activities** – survey and interview participants agree to submit/share information, that may necessarily be identifiable, about their data and software management activities. The legal basis for which are, (unless otherwise specified): (i) consent (ii) legitimate interests – to gather research data. (Dataset DS-S5)

Full List of Datasets

The naming convention is as follows: DS-Mn pertains to the mapping section of the research: DS-Sn pertains to the survey/interview data collection section of the research and DS-An pertains to the analysis. N is the sequential number of the dataset as it is recorded in the DMP.

	Dataset reference	Dataset Name	First recorded in DMP
1	DS-M1	Contacts Who Know About the AHRC Community	M1
2	~~DS-M2~~	~~Background Literature~~	~~M1~~
3	DS-M3	GtR_Analysed	M1
4	~~DS-M4~~	~~Arts and Humanities Institutions~~	~~M2~~
5	~~DS-S1~~	~~Contacts for Survey~~	~~M2~~
6	DS-S2	Contacts for Interview/Prize draw	M9
7	~~DS-S3~~	~~Arts and Humanities Repositories~~	~~M6~~
8	DS-A1	Survey Results Complete	M9
9	DS-A2	Survey Results Without Demographics	M9
10	DS - S3	Interview Recordings	M10
11	DS- S4	Interview Transcripts	M10

Dataset Template

Dataset Reference	Eg DS-M4
Dataset Name	Name as shown in Project File on Sharepoint
Type/Methodology	It is anticipated most data will be collected but some might be generated. This includes notes made by the researchers on various processes, or consent forms. For example, survey response data is collected, and the methodology may be 'online, Qualtrics'. The survey respondents data would be a separate dataset.
Origin	The source of the data, eg, interviewees, websites, document repositories, etc.
Scale (Approx End Volume)	It is not anticipated that the project will create large datasets
Description	A brief explanation of the data set, eg 'Contact details (name, title and academic institution) of selection of arts and humanities researchers in UK.'
Purpose/Methodology	Why we have collected/generated this data, and how we plan to process it (eg, using NVIVO, Leximancer, or simply to contact people)
Associated Documentation	Formal/informal documentation describing any part of the collection process or planned use of data; any relevant outputs such as papers
Useful to	This will generally be the researchers and the funder, but could also potentially be of use to other researchers.
Standards	With non-technical data, generally refers to .xls etc

Data Sharing Policy & Lawful Basis for Sharing	How the data might be used in future internally/externally. Whether or not the data will be released in line with UKRI guidelines (default availability) or it is protected by confidentiality, GDPR or IP. AHRC expects that grant holders 'must make any significant electronic resources or datasets created as a result of research funded by the Council available in an accessible and appropriate depository for at least three years after the end of their grant' Data status: will it be open, selectively shared or charged
Dissemination	How the data will be disseminated/shared (if appropriate) How interested parties should be informed How the data should be cited (if appropriate)
Location, Archiving and Storage	Where the data will be stored How the data will be destroyed at the end of the preservation period Who is responsible for this
Preservation Time and Cost	This could be in line with Ethics requirements, in line with AHRC project requirements or in perpetuity if it is a dataset for the purpose of future research. In general, the Research Councils expect data that underpins findings in publications should be accessible for at least ten years after publication. Cost of preserving data (to the project) beyond end of project

Dataset Information

Dataset Reference	DS-M1
Dataset Name	CONTACTS WHO KNOW ABOUT THE AHRC COMMUNITY

Type/Methodology	Excel spreadsheet, collected from publicly available information
Origin	Identified by members of the project team,
Scale (Approx End Volume)	Small
Description	'Contact details (name, title and academic institution) and socio-demographics (gender, ethnicity, disability) of selection of arts and humanities researchers in UK.'
Purpose/Methodology	This data was collected in order to contact people for the Stage 1 exploratory interviews as described in the report methodology; a scoping stage which involved a mapping of the arts and humanities community; a review of literature and UKRI data policies and 12 exploratory interviews with researchers and stakeholders. Approved under ERGO 65039.
Associated Documentation	Report: Shaping Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC Rebecca Taylor, Johanna Walker, Simon Hettrick (University of Southampton) and David de Roure (University of Oxford)
Useful to	The researchers during the project
Standards	xls
Data Sharing Policy & Lawful Basis for Sharing	Personally identifying information – therefore None.

Dissemination	N/A
Location, Archiving and Storage	In line with Ethics requirements this data will be destroyed at the end of the project (August 2022) by removal from Sharepoint by the PI in line with whatever method is currently employed for secure destruction from Sharepoint.
Preservation Time and Cost	N/A

Dataset Reference	DS-M3
Dataset Name	GtR_Analysed
Type/Methodology	Excel, collected, generated
Origin	https://gtr.ukri.org plus email addresses (web search and algorithmically created)
Scale (Approx End Volume)	Small
Description	Enhanced version of publicly available Gateway to Research Data.

Purpose/Methodology	Developing contact database for dissemination of a large-scale survey of researchers from across the community including those in a wide range of higher education institutions and IROs, in various roles both academic and professional/technical, at various career stages (junior, early, mid and senior), and from a wide range of arts and humanities disciplines. Approved under ERGO 65039.
Associated Documentation	Report Shaping Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC Rebecca Taylor, Johanna Walker, Simon Hettrick (University of Southampton) and David de Roure (University of Oxford)
Useful to	Research Team
Standards	xls
Data Sharing Policy & Lawful Basis for Sharing	None
Dissemination	None
Location, Archiving and Storage	In line with Ethics requirements this data will be destroyed at the end of the project (August 2022) by removal from Sharepoint by the PI in line with whatever method is currently employed for secure destruction from Sharepoint.

Preservation Time and Cost	N/A

Dataset Reference	DS-M9
Dataset Name	Contacts for Interview/Prize Draw
Type/Methodology	Excel, collected, generated
Origin	Subset of those survey respondents who wished to volunteer for interview.
Scale (Approx End Volume)	Small (<100 lines)
Description	Subset of socio-demographic, name and contact information from Survey responses, enhanced with administrative data (eg date of interview, other notes) and pseudonymisation keys.
Purpose/Methodology	Organisation of In-depth semi structured interviews with 30 survey participants who had opted in to further research and were purposively selected to capture disciplinary and career stage diversity. The interviews explored research practices, experiences, cultures and skills including barriers and facilitators to training and development, forms of institutional support and future policy scenarios.
Associated Documentation	Report Shaping Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC Rebecca Taylor, Johanna Walker, Simon Hettrick (University of Southampton) and David de Roure (University of Oxford)

Useful to	Research Team
Standards	xls
Data Sharing Policy & Lawful Basis for Sharing	None
Dissemination	None
Location, Archiving and Storage	In line with Ethics requirements this data will be destroyed at the end of the project (August 2022) by removal from Sharepoint by the PI in line with whatever method is currently employed for secure destruction from Sharepoint.
Preservation Time and Cost	N/A

Dataset Reference	DS-A1
Dataset Name	Survey Complete Results
Type/Methodology	Excel, collected, analysis

Origin	Complete results of survey carried out of Arts and Humanities researchers in Feb/March 2022
Scale (Approx End Volume)	Small
Description	Raw results and analysis of responses to all survey questions of inc univariate and multivariate This version contains some personally identifiable information and sensitive personal information
Purpose/Methodology	A large-scale survey of researchers from across the community including those in a wide range of higher education institutions and IROs, in various roles both academic and professional/technical, at various career stages (junior, early, mid and senior), and from a wide range of arts and humanities disciplines. Responses were collected from 335 researchers (281 complete). The survey measured current research practices including data collections and analysis, data management and sustainability, use of software, institutional support and research skills.
Associated Documentation	Survey questions - Appendix 3 of Report Shaping Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC

The report contains detailed analysis of the socio-demographics of the survey sample compared to the overall community.

Rebecca Taylor, Johanna Walker, Simon Hettrick (University of Southampton) and David de Roure (University of Oxford) |
| Useful to | Research team |
| Standards | XLS |

Data Sharing Policy & Lawful Basis for Sharing	None
Dissemination	None
Location, Archiving and Storage	In line with Ethics requirements this data will be destroyed at the end of the project (August 2022) by removal from Sharepoint by the PI in line with whatever method is currently employed for secure destruction from Sharepoint.
Preservation Time and Cost	N/A

Dataset Reference	DS-A2
Dataset Name	Survey Results (with all personally identifiable and sensitive information removed.)
Type/Methodology	Excel, collected, analysis
Origin	Abridged results of survey carried out of Arts and Humanities researchers in Feb/March 2022
Scale (Approx End Volume)	Small

Description	Raw results and analysis of responses to survey questions of inc univariate and multivariate.
Purpose/Methodology	A large-scale survey of researchers from across the community including those in a wide range of higher education institutions and IROs, in various roles both academic and professional/technical, at various career stages (junior, early, mid and senior), and from a wide range of arts and humanities disciplines. Responses were collected from 335 researchers (281 complete). The survey measured current research practices including data collections and analysis, data management and sustainability, use of software, institutional support and research skills.
Associated Documentation	Survey questions - Appendix 3 of Report Shaping Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC

The report contains detailed analysis of the socio-demographics of the survey sample compared to the overall community.

Rebecca Taylor, Johanna Walker, Simon Hettrick (University of Southampton) and David de Roure (University of Oxford) |
| Useful to | Research team, other researchers exploring the arts and humanities research community |
| Standards | XLS |
| Data Sharing Policy & Lawful Basis for Sharing | Open Data under CC-BY

Fully consented under ERGO 65039 for future use |

Dissemination	DOI will be created for this data and added to final report Contact report authors
Location, Archiving and Storage	UoS repository (PURE)
Preservation Time and Cost	UoS repository requirements (10 years)

Dataset Reference	DS-S4
Dataset Name	Interview Recordings
Type/Methodology	Collected video (Teams/Streams)
Origin	Raw interview footage for qualitative data for report 1. 12 exploratory interviews with researchers and stakeholders 2. In-depth semi structured interviews with 30 survey participants who had opted in to further research and were purposively selected to capture disciplinary and career stage diversity.
Scale (Approx End Volume)	Medium
Description	45 – 60 minutes interviews and group interviews with members of the AHRC research community

Purpose/Methodology	The interviews explored research practices, experiences, cultures and skills including barriers and facilitators to training and development, forms of institutional support and future policy scenarios. Transcribed for thematic analysis in NVivo.
Associated Documentation	Appendix 2 - ERGO 65039 Consent and PIS statements Survey questions - Appendix 4 – ERGO 65039 Interview Topic Guide

Report Shaping Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC

The report contains detailed analysis of the socio-demographics of the interview sample compared to the overall community as well as explanation of how the exploratory interviews were used

Rebecca Taylor, Johanna Walker, Simon Hettrick (University of Southampton) and David de Roure (University of Oxford) |
Useful to	Research team, future research on the arts and humanities community
Standards	mp4
Data Sharing Policy & Lawful Basis for Sharing	None
Dissemination	Project only

Location, Archiving and Storage	In line with Ethics requirements this data will be destroyed at the end of the project (August 2022) by removal from Sharepoint by the PI in line with whatever method is currently employed for secure destruction from Sharepoint.
Preservation Time and Cost	None

Dataset Reference	DS-S3
Dataset Name	Interview Recording transcripts
Type/Methodology	Collected video (Teams/Streams) Transcribed, pseudonymised then anonymised
Origin	Raw interview footage for qualitative data for report 3. 12 exploratory interviews with researchers and stakeholders 4. In-depth semi structured interviews with 30 survey participants who had opted in to further research and were purposively selected to capture disciplinary and career stage diversity.
Scale (Approx End Volume)	Medium
Description	45 – 60 minutes anonymised interviews and group interviews with members of the AHRC research community
Purpose/Methodology	The interviews explored research practices, experiences, cultures and skills including barriers and facilitators to training and development, forms of institutional support and future policy scenarios.

Associated Documentation	Appendix 2 - ERGO 65039 Consent and PIS statements Survey questions - Appendix 4 – ERGO 65039 Interview Topic Guide of Report Shaping Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC The report contains detailed analysis of the socio-demographics of the interview sample compared to the overall community as well as explanation of how the exploratory interviews were used Rebecca Taylor, Johanna Walker, Simon Hettrick (University of Southampton) and David de Roure (University of Oxford)
Useful to	Research Team, Researchers looking at the Arts and Humanities research communities
Standards	PDF/Word
Data Sharing Policy & Lawful Basis for Sharing	Fully consented (consent is in transcripts) Open Data (no personally identifiable information) CC-BY
Dissemination	DOI in project report Contact authors
Location, Archiving and Storage	PURE

Preservation Time and Cost	10 years in line with UoS requirements.

Sources

This data management plan was informed by the following sources:
- AHRC Data Management Plan Guidance
- UKRI Open Research Policy
- Data Pitch H2020 Data Management Plan Deliverable
- School of Advanced Study University of London Data Management Plans Content Guide

<div align="center">

AHRC Survey Results

April 2022

</div>

1 Highlights

1.1 Univariate Analysis

- The respondents of this survey were comprised of lower numbers of mid-career researchers (Phase 2: 12.96 %, Phase 3: 16.67 %), and higher numbers of Phase 1 (37.65 %) and Phase 4 (32.72 %).

- Only 36.72 % of respondents work with numerical data, which is arguably the data type that current storing and sharing methods are best suited to.

- The most common typical data storage location is a shared area provided by the institution (typically used by 38.03 % respondents).

- Only 27.87 % regularly share research data, although it was noted in the free text questions that sensitivity of data often prevents sharing.

- 32.13 % of respondents had not been involved in research that included a data management plan. However, when a data management plan was used, it was most likely to be 'mostly followed'.

- The most common method of sharing data was by citation in publications (56.72 %)

- Where research software was used, it was most likely to have been provided by the University (56.00 %).

- Relatively low numbers of respondents (N = 39, 11.68 %) were involved in research software development.

- 84.40% of respondents had learnt at least one new data skill over the last five years, whereas only 41.49 % of respondents had learnt at least one software skills over the last five years.

- The most common method for learning new skills was learning on the job (self-taught) which was reported by 71.28 % of respondents.

- The most common barrier to learning skills was lack of time to engage (reported by 44.13 %).

- Confidence in policies and responsibilities was on average higher for data than research software.

1.2 Multivariate Analysis

- Phase 1 (PhD student) respondents were much more likely to report that their typical approach to storing data was on a laptop or desktop, to have not shared their research data, and to have not used a data management plan.

- Phase 1 respondents were more likely to have completed data or software skills training during PGR and PGT courses.

- Phase 2 (Early Career Researcher) respondents were most likely to lack time and funding to attend training.

- On average, confidence in data and software skills increased with increasing career stage.

- Perhaps unsurprisingly, respondents who work in Digital Humanities shared data more regularly, were more likely to licence and create a DOI for their data, had learnt more data and software skills, and were more confident in their data and software responsibilities.

- Respondents who work in Languages and Literature were more likely to typically store data on a laptop or desktop computer.

<div align="center">1</div>

- When the data was segmented by institution, lower levels of data skills and use of reliable data practices were reported by respondents from the University of Oxford but this is likely due to the fact that there was a relatively higher proportion of Phase 1 (PhD student) respondents from this institution as compared to other institutions.

- Data skills and reliable data practices were higher among respondents who work with numerical data, as compared to overall figures. This may be because the systems for storing and sharing data are more suited to numerical data than other data types.

Note that 'Other' responses and free text responses can be found at the end of the document.

2 Which institution/organisation do you work for? (N = 322)

Institution	N	Percent
University of Glasgow	15	4.66
University of Oxford	13	4.04
University of York	13	4.04
University of Liverpool	12	3.73
University of Birmingham	12	3.73
Royal Holloway University of London	10	3.11
The Open University	9	2.8
University of Leeds	9	2.8
University College London	9	2.8
Newcastle University	8	2.48
University of Exeter	7	2.17
Durham University	7	2.17
University of Edinburgh	7	2.17
University of Manchester	7	2.17
Birkbeck College University of London	6	1.86
University of Cambridge	5	1.55
University of Sheffield	5	1.55
University of Leicester	5	1.55
Birmingham City University	5	1.55
Manchester Metropolitan University	5	1.55
University of Nottingham	5	1.55
Queen's University Belfast	5	1.55
Loughborough University	5	1.55
University of Sussex	5	1.55
University of Surrey	4	1.24
University of Kent	4	1.24
King's College London	4	1.24
University of the Arts London	4	1.24
University of Bristol	3	0.93
University of St Andrews	3	0.93
University of Strathclyde	3	0.93
Bournemouth University	3	0.93
Queen Mary University of London	3	0.93
Nottingham Trent University	3	0.93
Bath Spa University	3	0.93
Lancaster University	3	0.93
Edge Hill University	3	0.93
University of Aberdeen	3	0.93
University of Lincoln	3	0.93
Cardiff University	3	0.93
Northumbria University	3	0.93
University of Warwick	2	0.62
The National Archives	2	0.62
University of Stirling	2	0.62
University of the West of England	2	0.62

2

Institution	N	Percent
De Montfort University	2	0.62
Open University	2	0.62
University of Hertfordshire	2	0.62
Victoria and Albert Museum	2	0.62
Sheffield Hallam University	2	0.62
University of Hull	2	0.62
British Library	2	0.62
The Alan Turing Institute	2	0.62
Historic Environment Scotland	2	0.62
Aberystwyth University	2	0.62
Bangor University	1	0.31
Coventry University	1	0.31
F1000	1	0.31
HEC Montreal	1	0.31
University of Bradford	1	0.31
City University of London	1	0.31
Loughborough University London	1	0.31
University of Wales Trinity St David	1	0.31
Trinity College Dublin	1	0.31
University of Catania	1	0.31
University of Westminster	1	0.31
Glasgow School of Art	1	0.31
University of Antwerp	1	0.31
Cranfield University	1	0.31
Abertay University	1	0.31
University of Roehampton	1	0.31
Liverpool John Moores University	1	0.31
University of East Anglia	1	0.31
Keele University	1	0.31
University of Worcester	1	0.31
University of Southampton	1	0.31
Teesside University	1	0.31
Goldsmiths University of London	1	0.31
The British Library	1	0.31
National Gallery	1	0.31
Ryerson University	1	0.31
Liverpool University	1	0.31
National Museums Scotland	1	0.31
University of Reading	1	0.31
Stellenbosch University	1	0.31
King's Digital Lab	1	0.31
Ulster University	1	0.31
Edinburgh Napier University	1	0.31
Royal Museums Greenwich	1	0.31
Oxford Brookes University	1	0.31
University of Huddersfield	1	0.31
Imperial College London	1	0.31
Univesity of Aberdeen	1	0.31
Univeristy of Leeds	1	0.31
Glasgow Caledonian University	1	0.31
Courtauld Institute of Art	1	0.31
National Galleries of Scotland	1	0.31
Middlesex University	1	0.31
UAL Creative Computing Institute	1	0.31
University of Brighton	1	0.31
Falmouth University	1	0.31
Staffordshire University	1	0.31
British Museum	1	0.31
Royal Central School of Speech and Drama	1	0.31

3

3 In which disciplines do you work? (N = 334)

Discipline	N	Percent
History	94	28.14
Other	64	19.16
Languages and Literature	59	17.66
Digital Humanities	54	16.17
Archaeology	38	11.38
Cultural and Museum Studies	35	10.48
Music and Visual Arts	35	10.48
Media	31	9.28
Design	24	7.19
Linguistics	23	6.89
Information and Communication Technologies	23	6.89
Library and Information Studies	20	5.99
Philosophy	19	5.69
Classics	14	4.19
Drama and Theatre Studies	12	3.59
Development Studies	9	2.69
Divinity and Religion	9	2.69
Theology	9	2.69
Dance	5	1.5
Law and Legal Studies	5	1.5
Political Science and International Studies	5	1.5

4　To which of the following organisations have you applied for funding? (N = 334)

Funding Organisations Applied To	N	Percent
AHRC	290	86.83
My institution	126	37.72
British Academy	109	32.63
Leverhulme Foundation	88	26.35
European Research Council	56	16.77
ESRC	53	15.87
EPSRC	45	13.47
Other	40	11.98
Scholarly Societies	34	10.18
Wellcome Trust	34	10.18
Heritage Lottery Fund	28	8.38
Arts Council England	24	7.19
Innovate UK	19	5.69
Charitable/Philanthropic society	15	4.49
Andrew Mellon Foundation	14	4.19
NERC	13	3.89
Humanities in the European Research Area	13	3.89
MRC	11	3.29
Wolfson Foundation	7	2.1
Creative Scotland	4	1.2
BBSRC	3	0.9
STFC	1	0.3
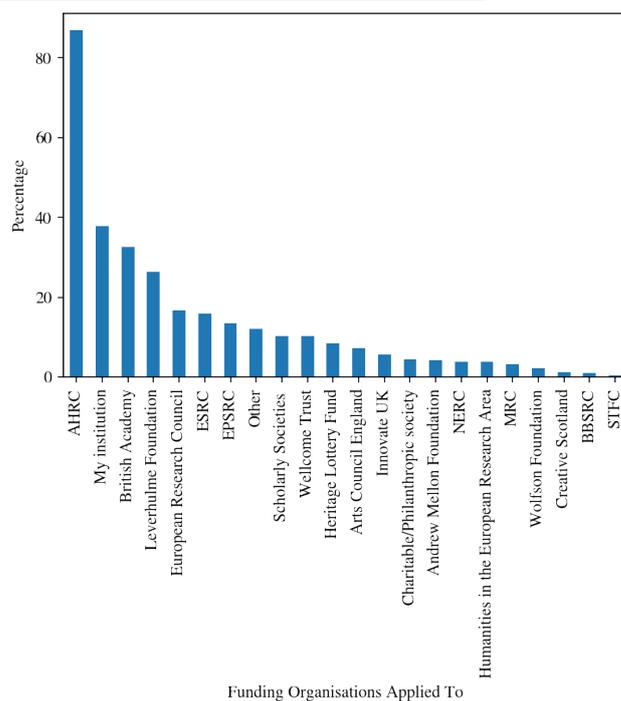
5 What is your career stage? (N = 324)

Career Stage	N	Percent
Phase 1 - Junior	122	37.65
Phase 2 - Early	42	12.96
Phase 3 - Mid / Recognised	54	16.67
Phase 4 - Established / Experienced / Senior	106	32.72
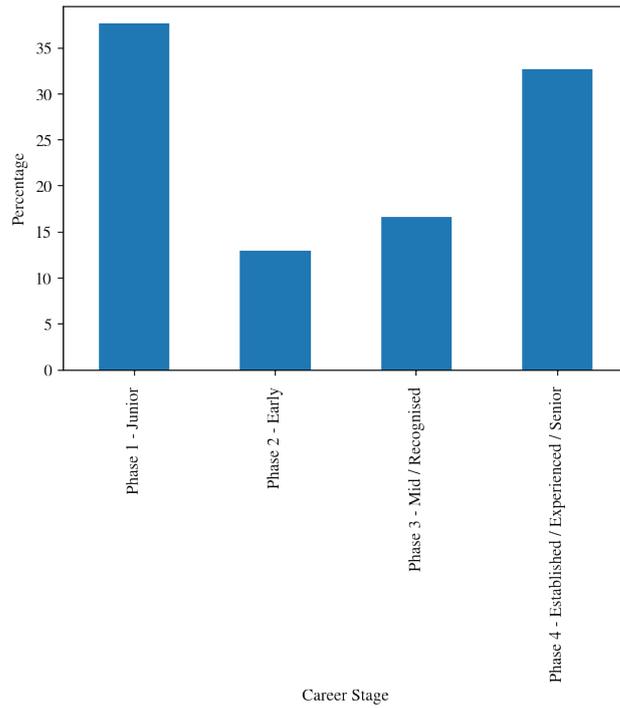
6 Over the last five years, what types of research data have you or your team worked with? (N = 305)

Types of Research Data	N	Percent
Textual	267	87.54
Image	213	69.84
Audio	137	44.92
Physical artefacts	129	42.3
Video	113	37.05
Numerical	112	36.72
Other	24	7.87
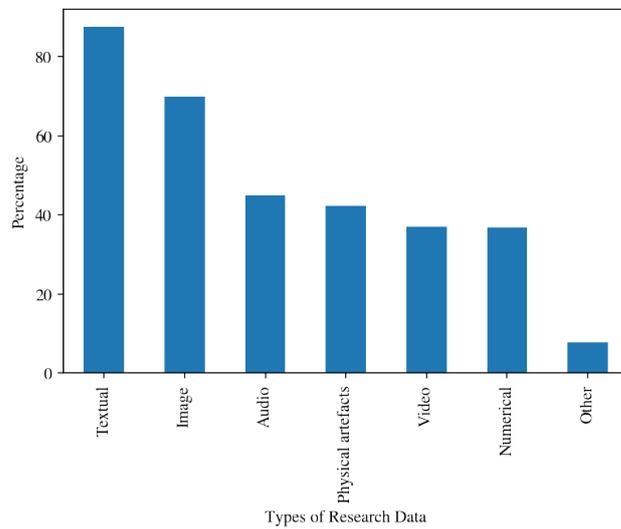
6

7 **Over the last five years - once you have completed a research project/study - which of the following options have you used to store the data that were generated? (N = 305)**

Data Storage Locations	N	Percent
On a laptop or desktop computer	234	76.72
In a shared area provided by your institution	226	74.1
In a shared area provided by yourself	164	53.77
In a data repository provided by your instution	107	35.08
In a data repository outside of your institution	62	20.33
Other	27	8.85
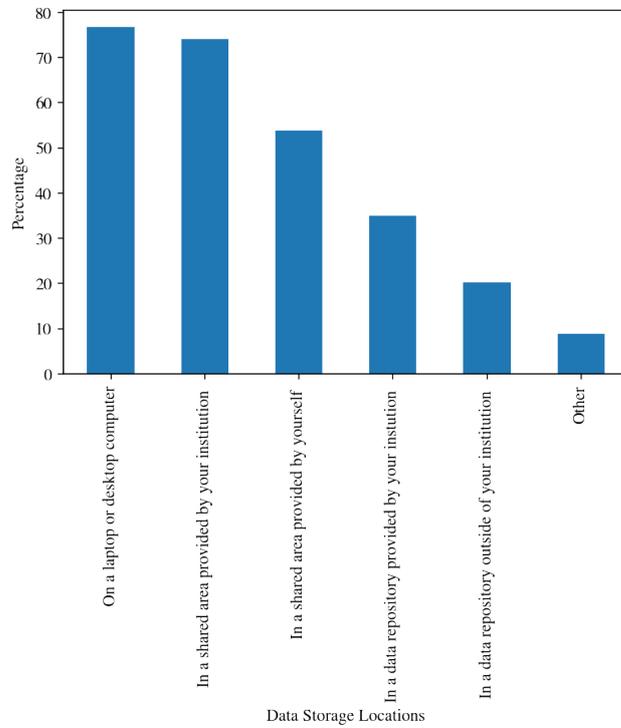
8 **If you selected more than one choice above, which of the following options describes your typical approach to storing the data that were generated? (N = 305)**

Typical Data Storage Location	N	Percent
In a shared area provided by your institution	116	38.03
On a laptop or desktop computer	80	26.23
In a shared area provided by yourself	53	17.38
In a data repository provided by your institution	24	7.87
In a data repository outside of your institution	18	5.9
Not applicable	14	4.59
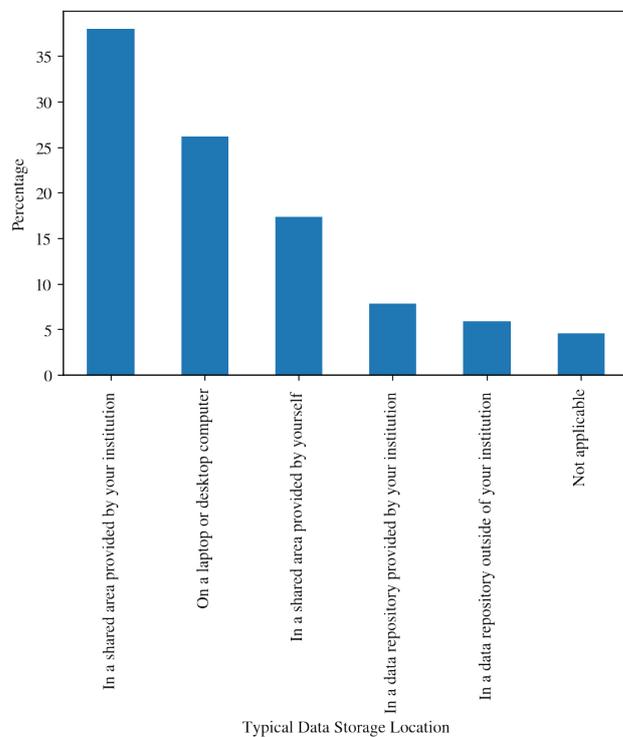
9 Over the last five years, which of the following statements best applies to your research data? (N = 305)

Data Sharing Frequency	Percent	N
I do not have research data to share	7.54	29.0
I haven't shared my research data	22.3	85.0
I have shared my research data but I rarely do so	18.03	69.0
I occasionally share my research data	24.26	93.0
I regularly share my research data	27.87	107.0
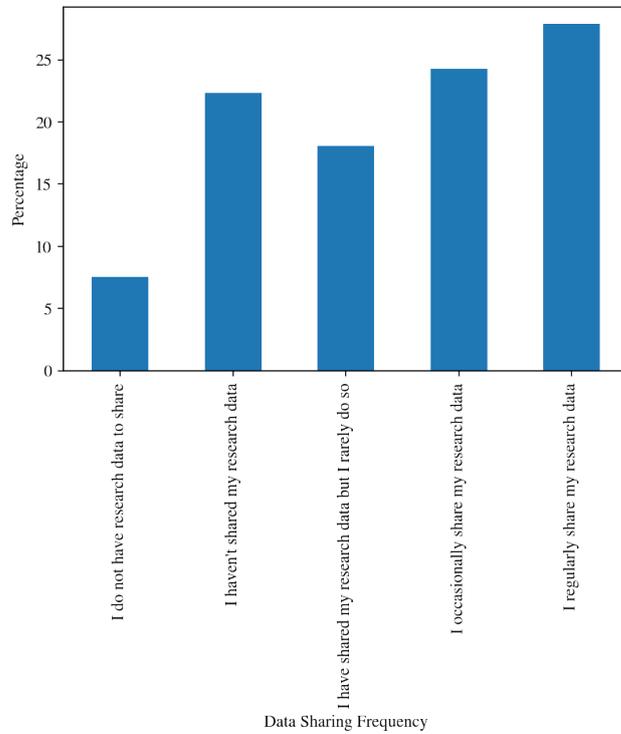
Chart showing Percentage by Data Sharing Frequency with categories: I do not have research data to share, I haven't shared my research data, I have shared my research data but I rarely do so, I occasionally share my research data, I regularly share my research data.

10 Over the last five years, what best describes your experience of data management plans on your projects? (N = 305)

Data Management Plan Experiences	Percent	N
I haven't been involved in research that included a data management plan	32.13	123.0
In my experience data management plans are not followed	3.61	14.0
In my experience data management plans are followed in part	19.67	75.0
In my experience data management plans are mostly followed	38.03	146.0
In my experience data management plans are followed exactly	6.56	25.0

10

Data Management Plan Experiences
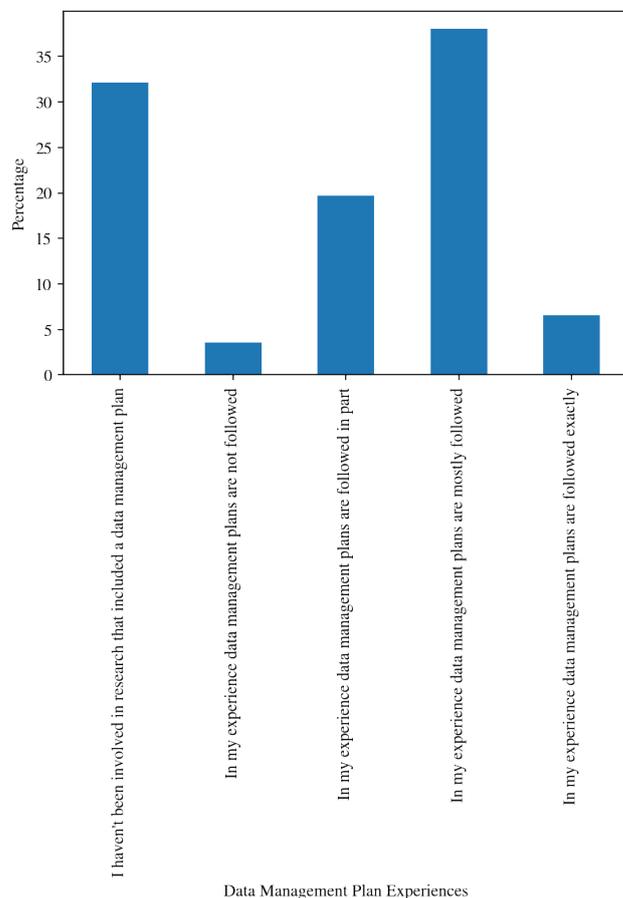
11 Over the last five years, which of the following practices have been part of your standard research process? (N = 305)

Methods of Data Sharing	N	Percent
I cite my research data in publications	173	56.72
I share my data with individuals/groups that have requested access	159	52.13
I create a unique identifier to make my data findable	72	23.61
I license my data to allow it to be shared	60	19.67

11

Percentage vs Methods of Data Sharing bar chart with categories: I cite my research data in publications; I share my data with individuals/groups that have requested access; I create a unique identifier to make my data findable; I license my data to allow it to be shared.
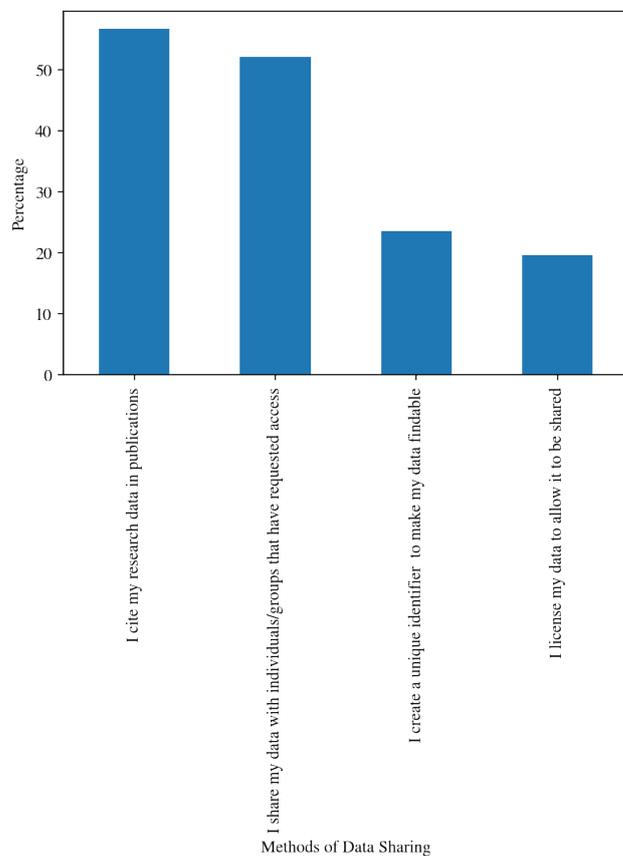
12 Over the last five years, which of the following have applied to your research? (N = 300)

Research Software Development	N	Percent
I have used research software provided by university	168	56.0
I have used specialist research software which was provided for free	103	34.33
I have not used research software	76	25.33
I have purchased specialist research software with research funding	46	15.33
I have developed my own research software	26	8.67
I have used research software that was developed for me	25	8.33
Other	9	3.0

12

Percentage

Research Software Development
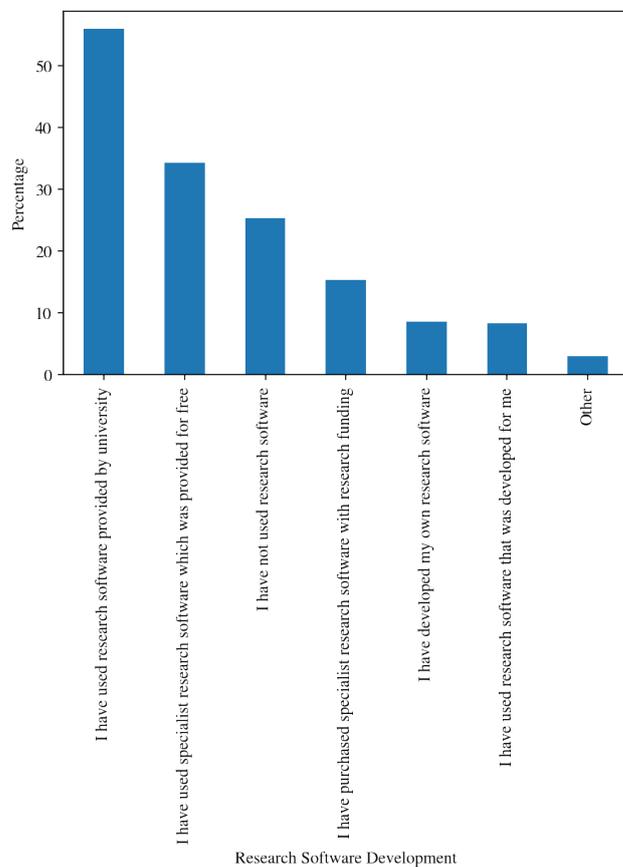
13 **Over the last five years, at the end of a research project/study, where have you typically stored the research software that was created? (N=39)**

Research Software Storage Locations	N	Percent
In a repository outside my institution	22	56.41
In an institutional software repository	6	15.38
On a laptop/desktop computer	5	12.82
Other	3	7.69
In a shared area provided by your institution	3	7.69

13

Research Software Storage Locations
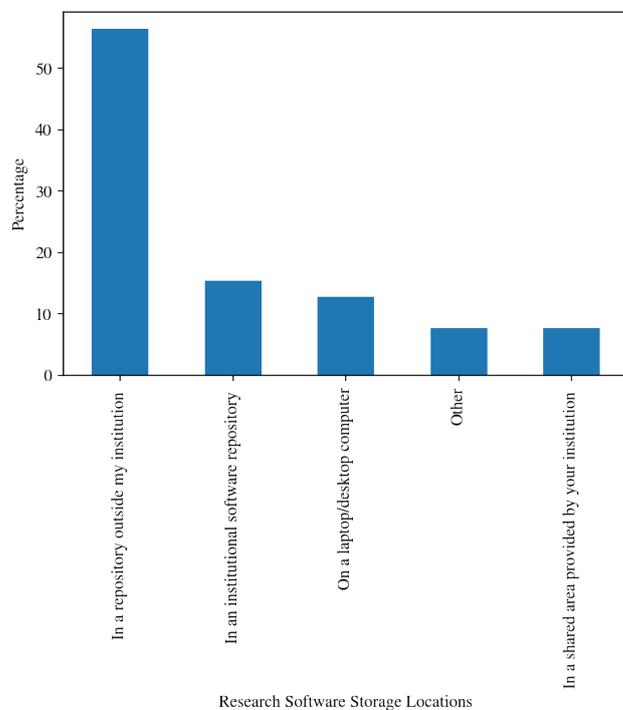
14 Over the last five years, which of the following practices have been part of your standard research process? (N = 39)

Methods of Sharing Research Software	N	Percent
I license my research software to allow to be shared	26	66.67
I share my research software with individuals/groups that have requested access	21	53.85
I cite my software in publications and other outputs from my research	21	53.85
I create a unique identifier to make my research software findable	10	25.64
I do not share my research software	1	2.56

Methods of Sharing Research Software
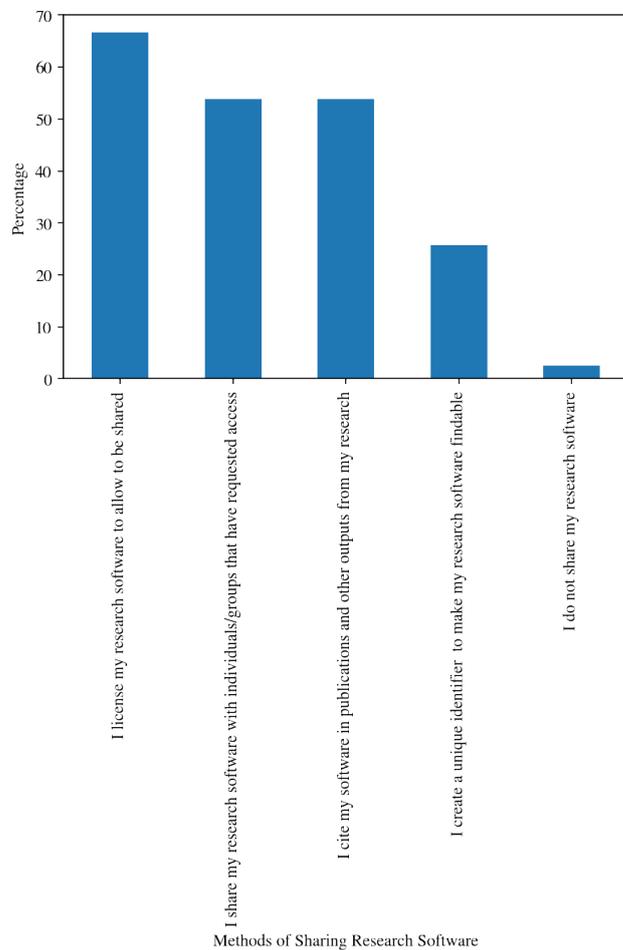
15   Who develops your research software? (N = 39)

Who Develops Research Software	N	Percent
Research Software Engineers from within academia	26	66.67
I develop my own research software	20	51.28
PhD students and postdoctoral researchers	18	46.15
Software developers from industry	11	28.21

15

Who Develops Research Software
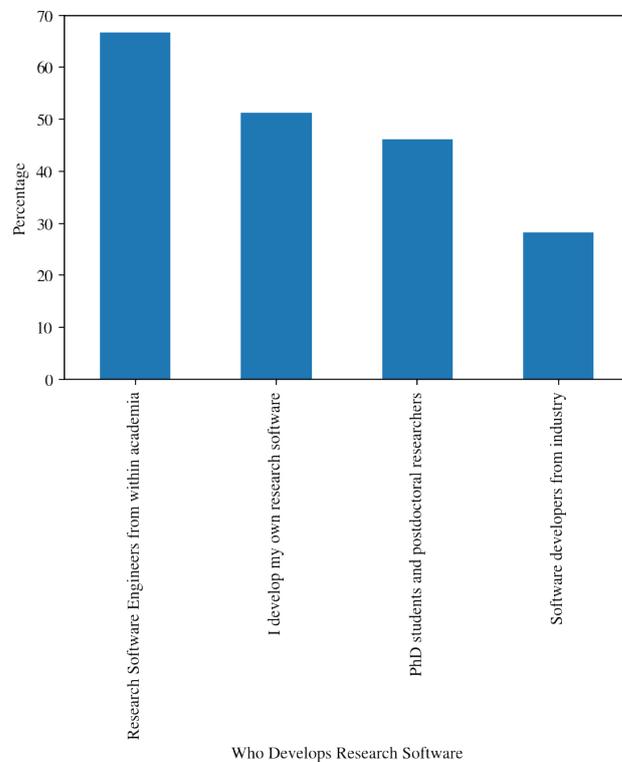
16 **Over the last five years, have you learned new skills or knowledge in relation to any of the following? (N = 282)**

Data Skills Learnt	N	Percent
Data management/storage	196	69.5
Data analysis	160	56.74
Data sharing	124	43.97
Data collection	122	43.26
None of the above	44	15.6

16
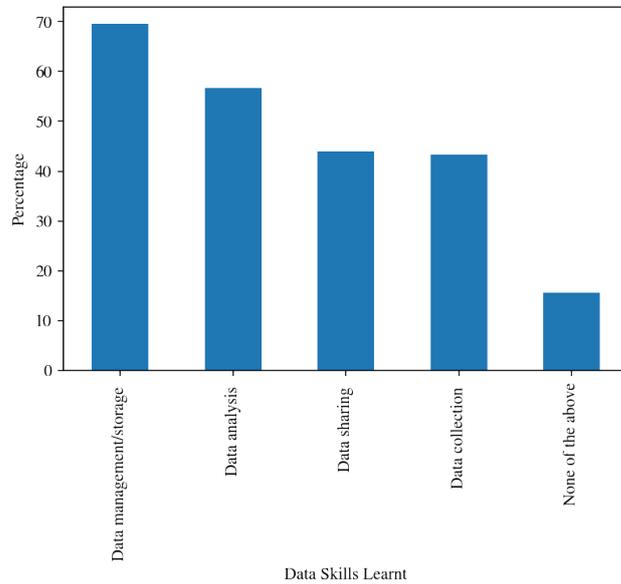
Data Skills Learnt

17 Over the last five years, have you learned new skills or knowledge in relation to any of the following? (N = 282)

Software Skills Learnt	N	Percent
None of the above	165	58.51
Software sharing	83	29.43
Software development	59	20.92
Software requirements collection	37	13.12

17

Software Skills Learnt
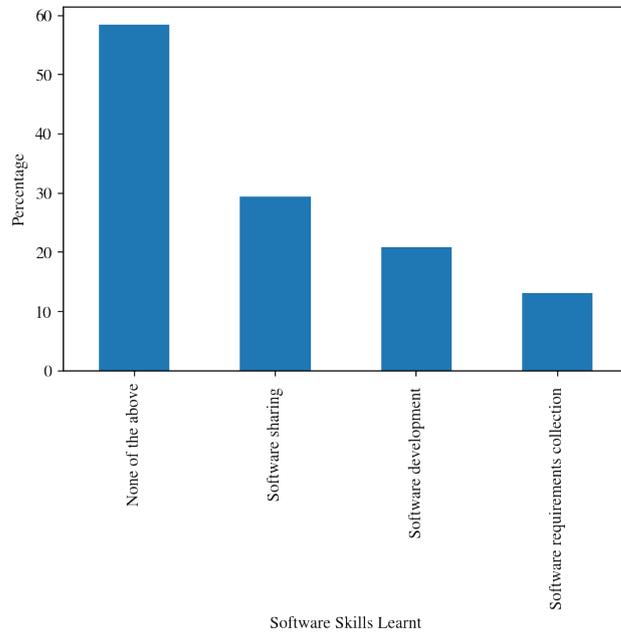
18 Where did you acquire your skills and knowledge about data and/or software? (N = 282)

Where Skills were Acquired	N	Percent
Learning on the job - self-taught	201	71.28
Learning on the job - peer taught	148	52.48
Training/advice provided in-house by my institution	100	35.46
Masters/Doctoral training	89	31.56
Training/advice provided externally	38	13.48
Other	11	3.9
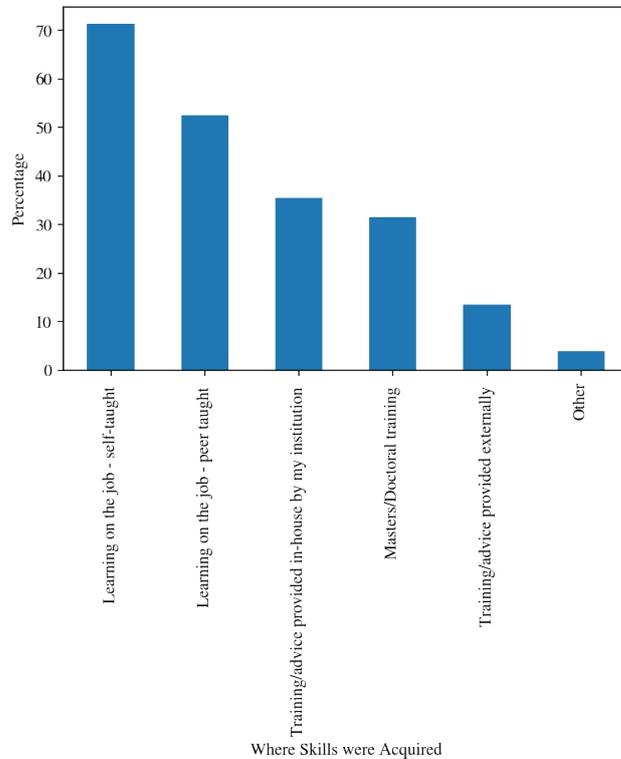
19 Thinking about your current skill set, what is the main data or software skill that you would like to obtain to improve your research? (N = 272)

Category	N
Data analysis	60
Not interested	25
Unsure but willing/Want to know what is available/Want to get started	23
Data sharing	23
Software development	22
Data collection	22
Other	18
Data storage	11
Data visualisation	10
Databases	10
General coding/programming	9
Bibliographic software	7
Web data skills	6
Data organisation	5

Software packages	N
R	17
Python	10
NVivo	10
GIS	9
Excel/spreadsheets	5
TEI XML	3
Matlab	2
SQL	2
I design	1
lidar	1
OBS Studio	1
Gephi	1
SPSS	1
HTML	1
javascript	1
CIDOC-CRM	1
Docker	1
Tableau	1
Notion	1
Korsakow programming	1
Sketch engine	1

20 Do any of the following issues prevent you from gaining the the main data or software skill that you would like to obtain to improve your research? (N = 281)

Barriers to Learning Skill	N	Percent
I'm aware of training/learning opportunities but I have no time to engage with them	124	44.13
I'm not aware of any training or opportunities to learn this skill	75	26.69
None of the above	56	19.93
There are only external training/learning opportunities but I don't have funding to attend them	48	17.08
My institution doesn't support the infrastructure necessary to make use of the skill	29	10.32
Other	23	8.19

Barriers to Learning Skill
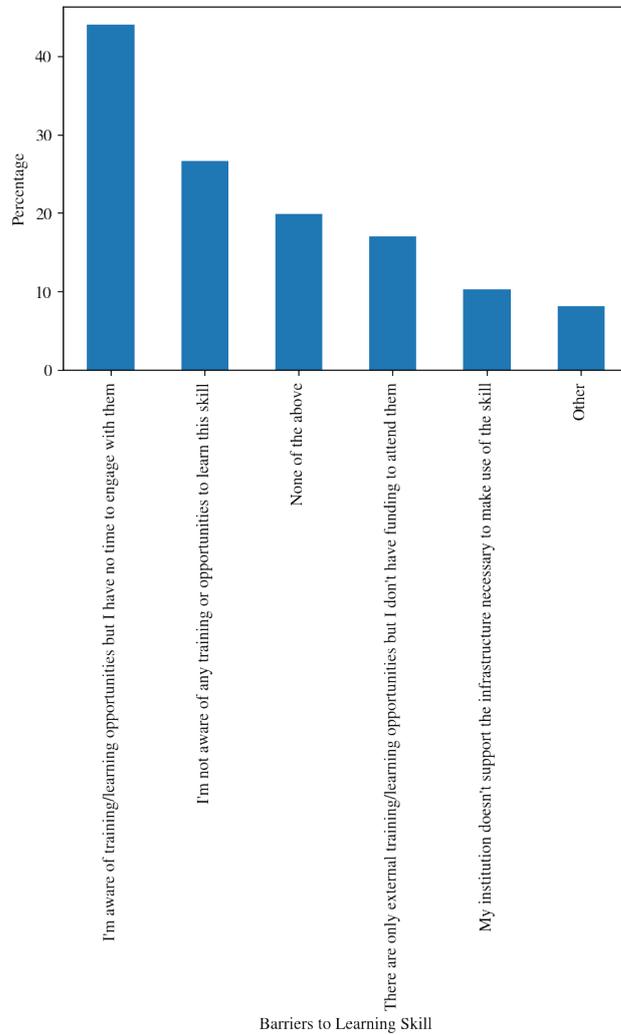
21 Are you confident that you understand the policies and responsibilities related to research data that are placed on you during a project/study? (N = 272)

Data Responsibilities During Study	N	Percent
1 (Not at all confident)	5	1.84
2	19	6.99
3	71	26.1
4	122	44.85
5 (Very confident)	55	20.22

21

Data Responsibilities During Study
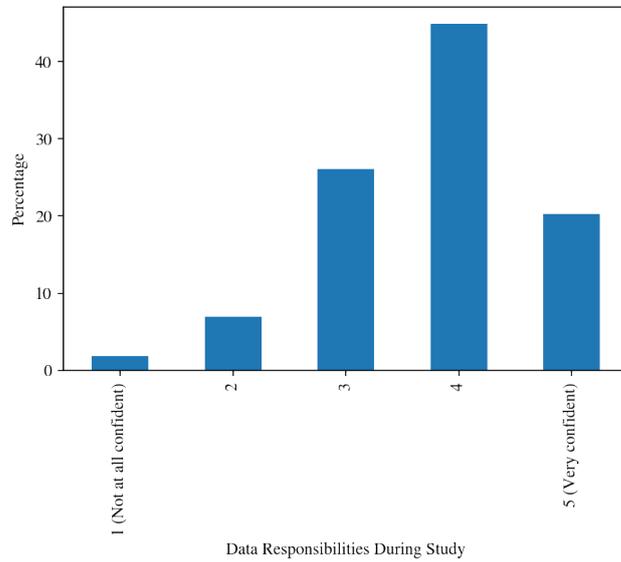
22 **Are you confident that you understand the policies and responsibilities related to research data that are placed on you data after a project/study has ended? (N = 272)**

Data Responsibilities After Study	N	Percent
1 (Not at all confident)	13	4.78
2	38	13.97
3	77	28.31
4	97	35.66
5 (Very confident)	47	17.28

22

Data Responsibilities After Study
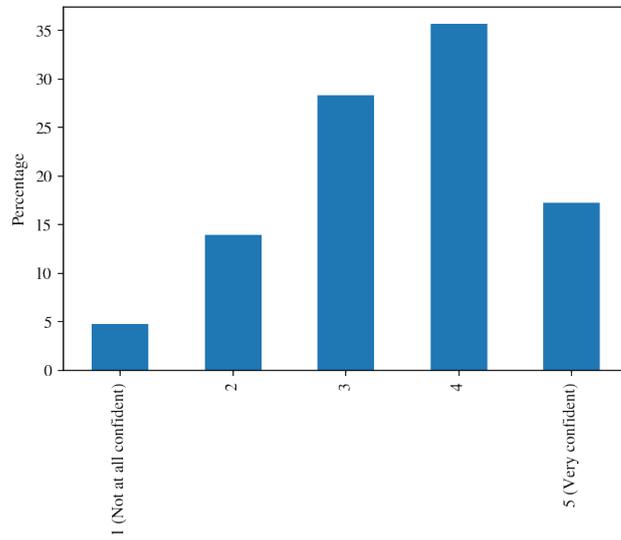
23   Are you confident that you understand the policies and responsibilities related to research software that are placed on you during a project/study? (N = 264)

Software Responsibilities During Study	N	Percent
1 (Not at all confident)	28	10.61
2	49	18.56
3	89	33.71
4	71	26.89
5 (Very confident)	27	10.23

23

Software Responsibilities During Study
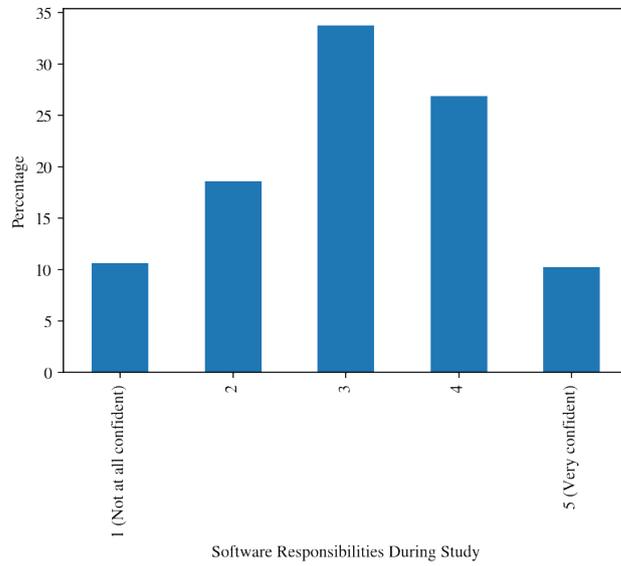
24 Are you confident that you understand the policies and responsibilities related to research software that are placed on you data after a project/study has ended? (N = 264)

Software Responsibilities After Study	N	Percent
1 (Not at all confident)	35	13.46
2	63	24.23
3	77	29.62
4	63	24.23
5 (Very confident)	22	8.46

24

Software Responsibilities After Study
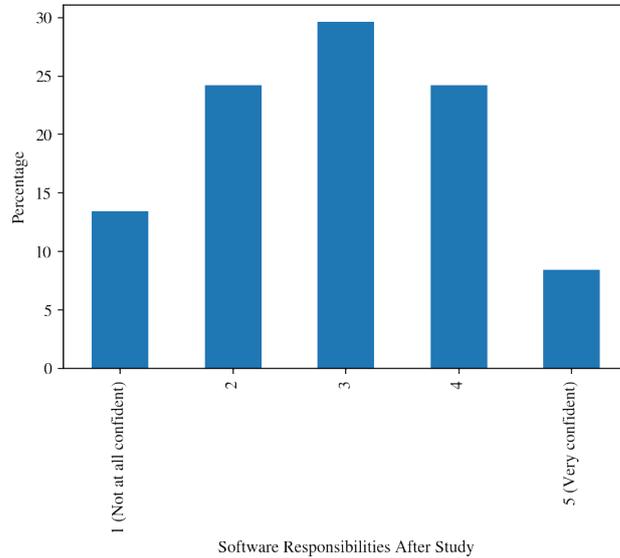
25   Analyses Grouped by Career Stage

Career Stage	N	Percent
Phase 1 - Junior	122	37.65
Phase 2 - Early	42	12.96
Phase 3 - Mid / Recognised	54	16.67
Phase 4 - Established / Experienced / Senior	106	32.72

25.1   Over the last five years, what types of research data have you or your team worked with? (As a percentage of all respondents from the career stage)

What is your career stage?	Textual	Image	Audio	Physical artefacts	Video	Numerical
Phase 1 - Junior	81.97	56.56	29.51	36.07	20.49	21.31
Phase 2 - Early	88.1	73.81	50.0	47.62	30.95	45.24
Phase 3 - Mid / Recognised	81.48	70.37	46.3	42.59	37.04	35.19
Phase 4 - Established / Experienced / Senior	81.13	70.75	51.89	39.62	51.89	45.28

25.2 Over the last five years - once you have completed a research project/study - which of the following options have you used to store the data that were generated? (As a percentage of all respondents from the career stage)

What is your career stage?	In a shared area provided by your institution	On a laptop or desktop computer	In a shared area provided by yourself	In a data repository provided by your instution	In a data repository outside of your institution
Phase 1 - Junior	58.2	77.87	46.72	2.46	2.46
Phase 2 - Early	83.33	69.05	57.14	54.76	23.81
Phase 3 - Mid / Recognised	72.22	72.22	48.15	46.3	22.22
Phase 4 - Established / Experienced / Senior	76.42	66.98	53.77	52.83	34.91

26

25.3 Which of the following options describes your typical approach to storing the data that were generated? (As a percentage of all respondents from the career stage)

What is your career stage?	In a shared area provided by your institution	On a laptop or desktop computer	In a shared area provided by yourself	In a data repository provided by your institution	In a data repository outside of your institution	Not applicable
Phase 1 - Junior	30.7	40.35	19.3	0.88	1.75	7.02
Phase 2 - Early	35.71	14.29	21.43	16.67	9.52	2.38
Phase 3 - Mid / Recognised	40.82	18.37	16.33	8.16	6.12	10.2
Phase 4 - Established / Experienced / Senior	46.0	19.0	14.0	12.0	9.0	0.0

27

Phase 1 - Junior
Phase 2 - Early
Phase 3 - Mid / Recognised
Phase 4 - Established / Experienced / Senior

Percentage

40

30

20

10

0

In a shared area provided by your institution

On a laptop or desktop computer

In a shared area provided by yourself

In a data repository provided by your institution

In a data repository outside of your institution

Not applicable

Typical Data Storage Location

25.4 Data Sharing Frequency - Over the last five years, which of the following statements best applies to your research data? (As a percentage of all respondents from the career stage)

What is your career stage?	I do not have research data to share	I haven't shared my research data	I have shared my research data but I rarely do so	I occasionally share my research data	I regularly share my research data
Phase 1 - Junior	13.16	41.23	22.81	15.79	7.02
Phase 2 - Early	7.14	16.67	19.05	26.19	30.95
Phase 3 - Mid / Recognised	6.12	8.16	6.12	26.53	53.06
Phase 4 - Established / Experienced / Senior	2.0	10.0	18.0	32.0	38.0

28

Phase 1 - Junior
Phase 2 - Early
Phase 3 - Mid / Recognised
Phase 4 - Established / Experienced / Senior

Data Sharing Over Last Five Years
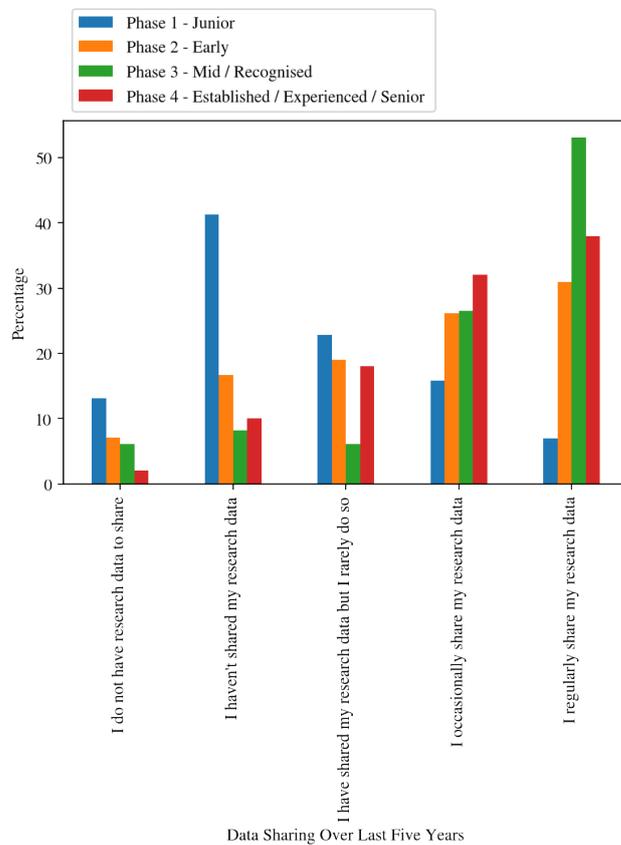
25.5 Over the last five years, what best describes your experience of data management plans on your projects? (As a percentage of all respondents from the career stage)

What is your career stage?	I haven't been involved in research that included a data management plan	In my experience data management plans are not followed	In my experience data management plans are followed in part	In my experience data management plans are mostly followed	In my experience data management plans are followed exactly
Phase 1 - Junior	60.53	2.63	11.4	22.81	2.63
Phase 2 - Early	14.29	11.9	26.19	40.48	7.14
Phase 3 - Mid / Recognised	14.29	2.04	16.33	57.14	10.2
Phase 4 - Established / Experienced / Senior	16.0	2.0	28.0	45.0	9.0

29

Data Management Plan Experience
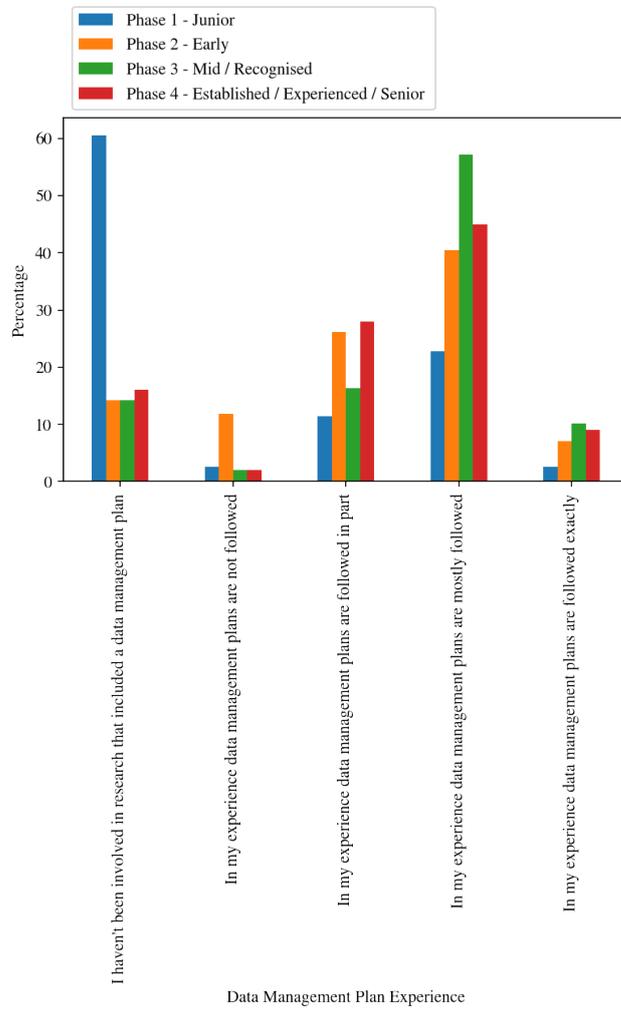
25.6 Data Sharing Methods - Over the last five years, which of the following practices have been part of your standard research process? (As a percentage of all respondents from the career stage)

What is your career stage?	I license my data to allow it to be shared	I share my data with individuals/groups that have requested access	I create a unique identifier (e.g. a DOI) to make my data findable	I cite my research data in publications
Phase 1 - Junior	5.74	34.43	8.2	33.61
Phase 2 - Early	28.57	50.0	28.57	61.9
Phase 3 - Mid / Recognised	31.48	57.41	29.63	66.67
Phase 4 - Established / Experienced / Senior	22.64	61.32	32.08	66.04

31
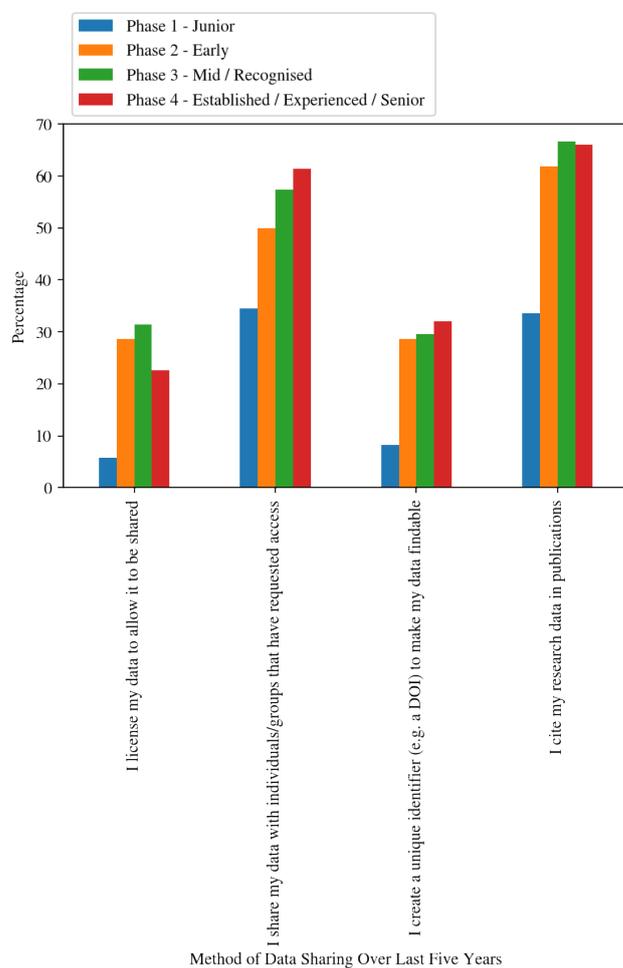
25.7 Research Software Use - Over the last five years, which of the following have applied to your research? (As a percentage of all respondents from the career stage)

What is your career stage?	I have not used research software	I have used research software provided by university	I have purchased specialist research software with research funding	I have used specialist research software which was provided for free	I have used research software that was developed for me	I have developed my own research software
Phase 1 - Junior	32.79	47.54	4.1	24.59	0.0	2.46
Phase 2 - Early	30.95	54.76	11.9	30.95	11.9	11.9
Phase 3 - Mid / Recognised	14.81	50.0	25.93	38.89	16.67	9.26
Phase 4 - Established / Experienced / Senior	14.15	56.6	20.75	36.79	10.38	12.26

32
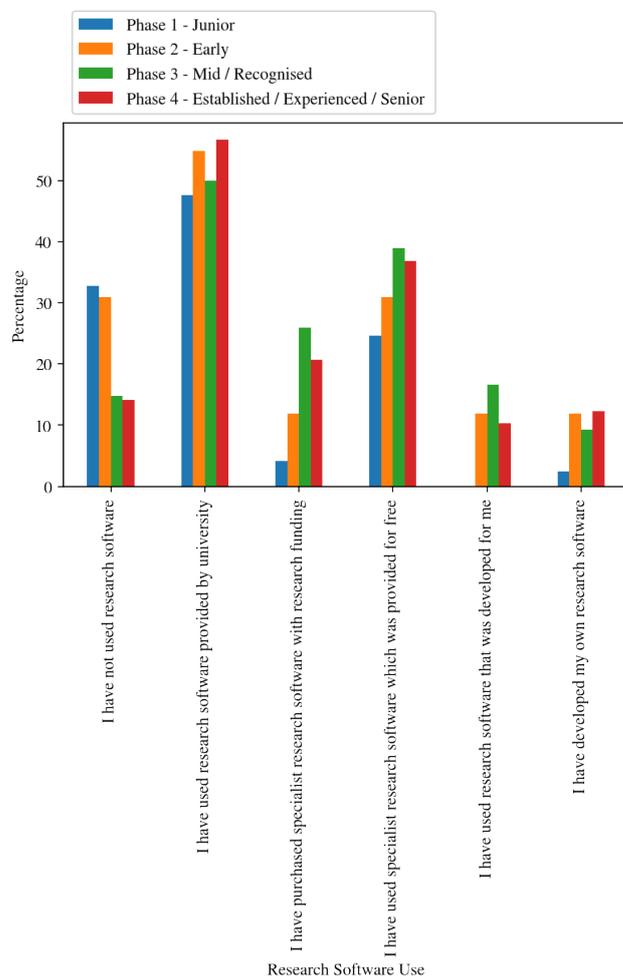
25.8 Over the last five years, at the end of a research project/study, where have you typically stored the research software that was created? (N=39) (As a percentage of all respondents from the career stage)

What is your career stage?	In a repository outside my insti-tution	In an institu-tional software reposi-tory	On a lap-top/desktop computer	In a shared area pro-vided by your in-stitution
Phase 1 - Junior	66.67	0.0	33.33	0.0
Phase 2 - Early	66.67	16.67	16.67	0.0
Phase 3 - Mid / Recognised	72.73	9.09	9.09	0.0
Phase 4 - Established / Experienced / Senior	42.11	21.05	10.53	15.79
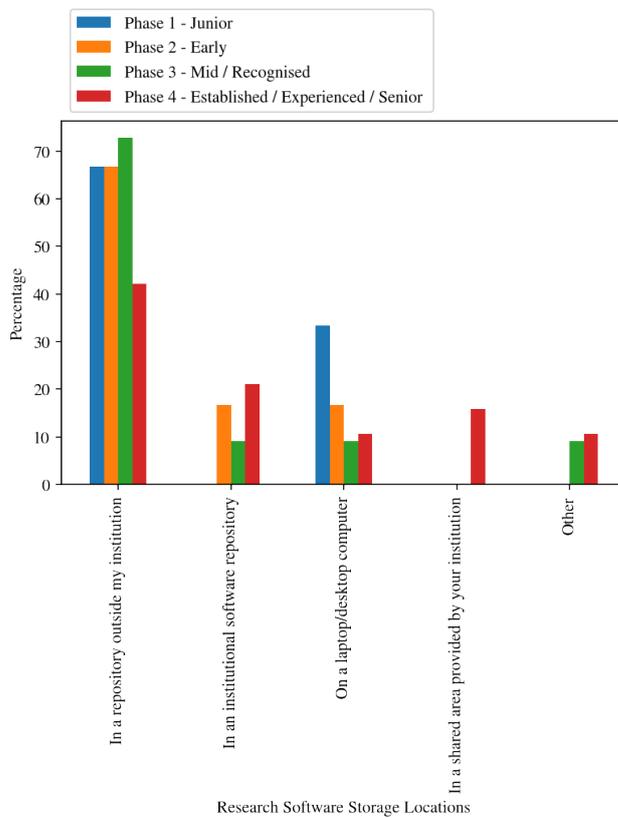
25.9 Over the last five years, which of the following practices have been part of your standard research process? (Research software sharing) (N = 39) (As a percentage of all respondents from the career stage)

What is your career stage?	I do not share my research software	I license my research software to allow to be shared	I share my research software with individuals/groups that have requested access	I cite my software in publications and other outputs from my research	I create a unique identifier to make my research software findable
Phase 1 - Junior	0.0	0.0	100.0	0.0	0.0
Phase 2 - Early	0.0	83.33	83.33	66.67	33.33
Phase 3 - Mid / Recognised	0.0	81.82	27.27	45.45	18.18
Phase 4 - Established / Experienced / Senior	5.26	63.16	52.63	63.16	31.58

34

Methods of Sharing Research Software
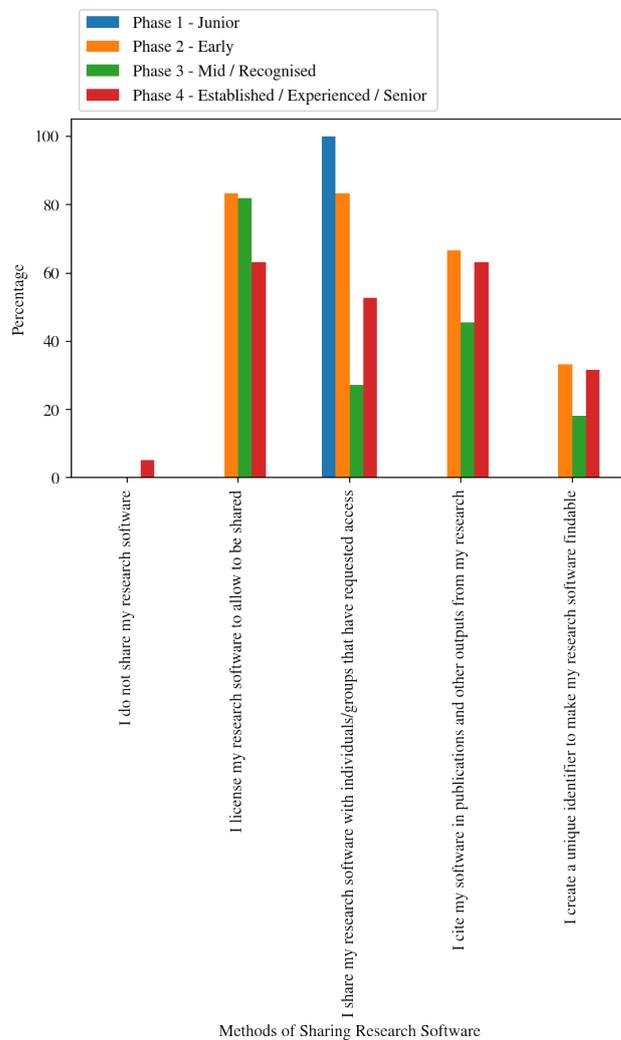
25.10 Who develops your research software? (N = 39) (As a percentage of all respondents from the career stage)

What is your career stage?	I develop my own research software	PhD students and post-doctoral researchers	Software developers from industry	Research Software Engineers from within academia
Phase 1 - Junior	100.0	0.0	33.33	33.33
Phase 2 - Early	66.67	50.0	16.67	66.67
Phase 3 - Mid / Recognised	36.36	72.73	9.09	63.64
Phase 4 - Established / Experienced / Senior	47.37	36.84	42.11	73.68

35

Who Develops Research Software
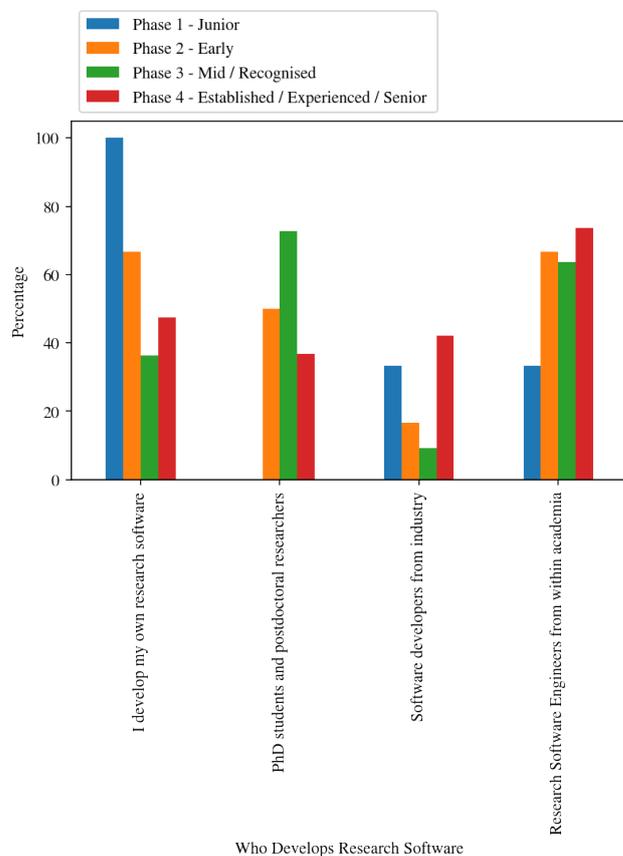
25.11 Data Skills - Over the last five years, have you learned new skills or knowledge in relation to any of the following? (As a percentage of all respondents from the career stage)

What is your career stage?	Data management/storage	Data analysis	Data sharing	Data collection
Phase 1 - Junior	45.9	45.9	21.31	42.62
Phase 2 - Early	80.95	69.05	64.29	45.24
Phase 3 - Mid / Recognised	66.67	57.41	51.85	42.59
Phase 4 - Established / Experienced / Senior	66.04	41.51	40.57	26.42

36

Data Skills Learnt
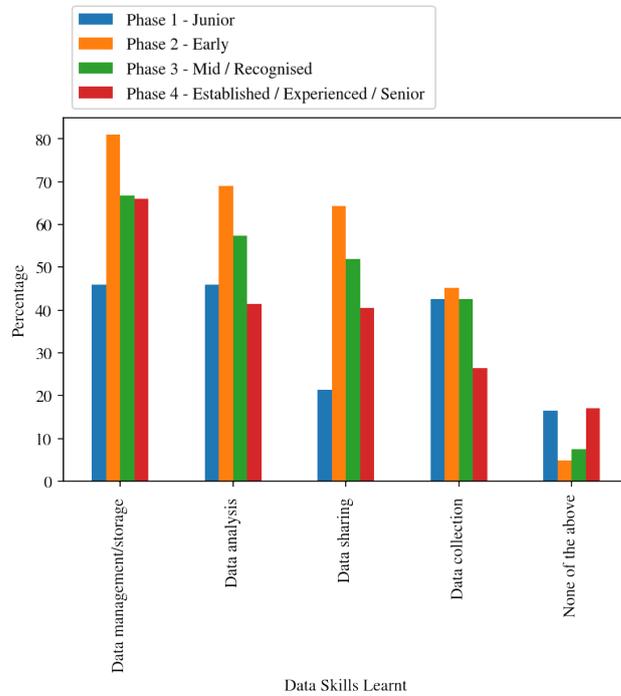
25.12   Research Software Skills - Over the last five years, have you learned new skills or knowledge in relation to any of the following? (As a percentage of all respondents from the career stage)

What is your career stage?	None of the above	Software sharing	Software development	Software requirements collection
Phase 1 - Junior	59.02	13.93	13.11	7.38
Phase 2 - Early	42.86	42.86	26.19	14.29
Phase 3 - Mid / Recognised	48.15	22.22	25.93	16.67
Phase 4 - Established / Experienced / Senior	46.23	33.96	16.98	12.26

37
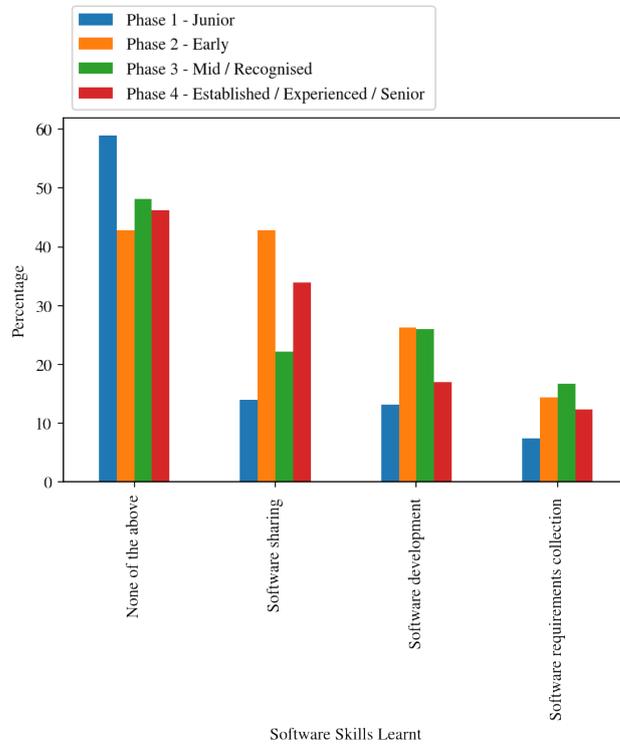
25.13 Where did you acquire your skills and knowledge about data and/or software?

What is your career stage?	Learning on the job - self-taught	Learning on the job - peer taught	Training provided in-house by my institution	Completed during PGR and PGT courses	Training provided externally
Phase 1 - Junior	53.28	23.77	18.03	53.28	9.84
Phase 2 - Early	80.95	52.38	59.52	23.81	19.05
Phase 3 - Mid / Recognised	72.22	61.11	37.04	12.96	18.52
Phase 4 - Established / Experienced / Senior	59.43	60.38	31.13	6.6	7.55
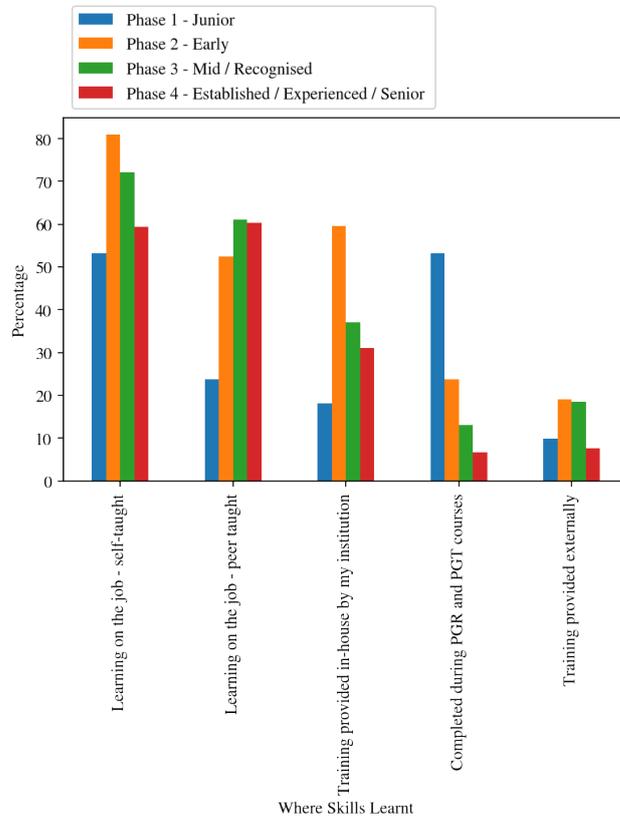
25.14 Do any of the following issues prevent you from gaining the skill you entered above (the main skill you want to improve)? (As a percentage of all respondents from the career stage)

What is your career stage?	I'm not aware of any training to learn this skill	I'm aware of training but I have no time to engage with it	There is only external training but I don't have funding to attend it	My institution doesn't support the infrastructure necessary to make use of the skill	None of the above
Phase 1 - Junior	28.69	28.69	10.66	7.38	21.31
Phase 2 - Early	21.43	50.0	28.57	11.9	9.52
Phase 3 - Mid / Recognised	27.78	38.89	14.81	3.7	11.11
Phase 4 - Established / Experienced / Senior	15.09	44.34	14.15	12.26	18.87

Barriers to Training
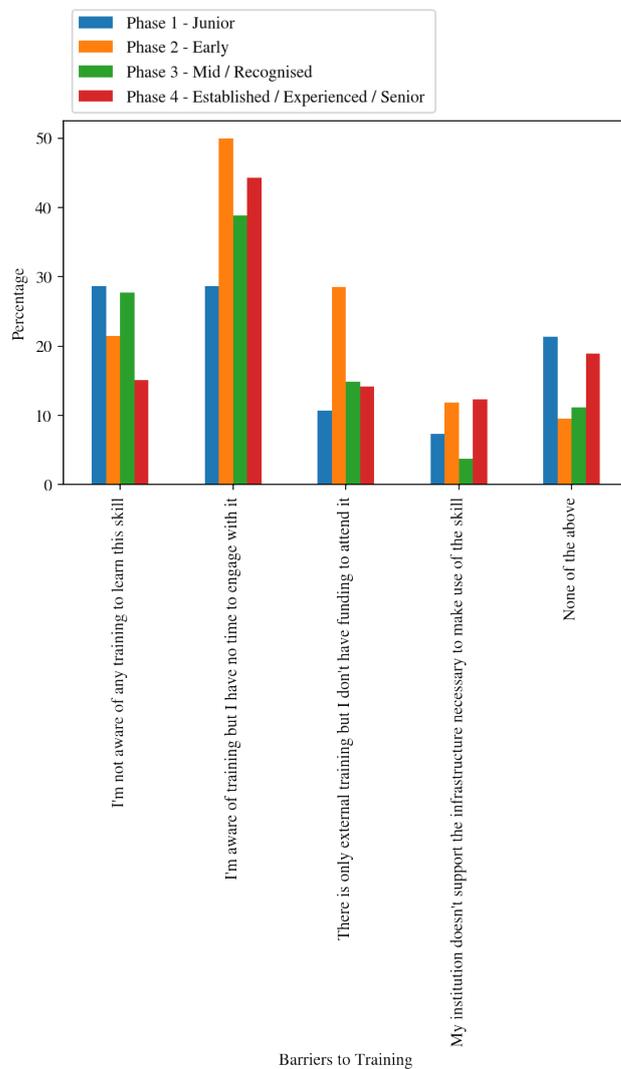
25.15 Are you confident that you understand the policies and responsibilities related to research data that are placed on you during a project/study? (As a percentage of all respondents from the career stage)

What is your career stage?	1 (Not at all confident)	2	3	4	5 (Very confident)
Phase 1 - Junior	4.4	9.89	41.76	36.26	7.69
Phase 2 - Early	0.0	2.5	17.5	50.0	30.0
Phase 3 - Mid / Recognised	0.0	6.52	17.39	45.65	30.43
Phase 4 - Established / Experienced / Senior	1.05	6.32	18.95	50.53	23.16

Confidence in Data Responsibilities During Study
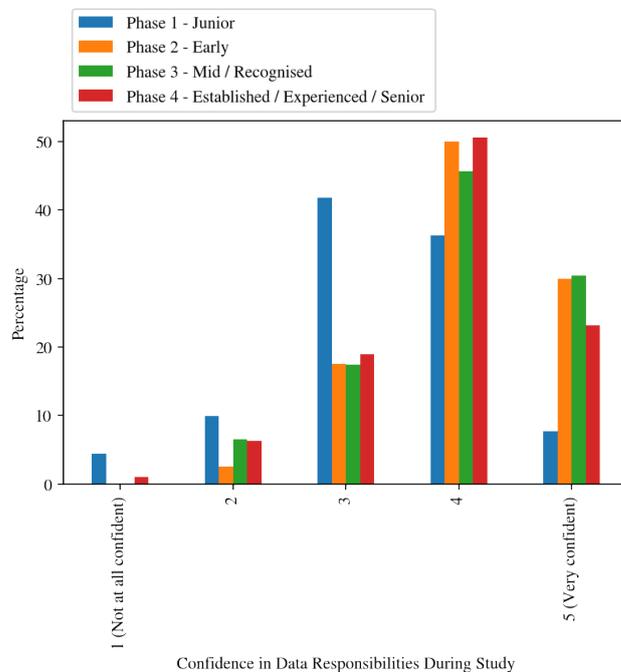
25.16 **Are you confident that you understand the policies and responsibilities related to research data that are placed on you after a project/study has ended? (As a percentage of all respondents from the career stage)**

What is your career stage?	1 (Not at all confident)	2	3	4	5 (Very confident)
Phase 1 - Junior	9.89	21.98	40.66	24.18	3.3
Phase 2 - Early	0.0	5.0	25.0	57.5	12.5
Phase 3 - Mid / Recognised	0.0	17.39	21.74	32.61	28.26
Phase 4 - Established / Experienced / Senior	4.21	8.42	21.05	38.95	27.37

41

Confidence in Data Responsibilities After Study

25.17 **Are you confident that you understand the policies and responsibilities related to research software that are placed on you during a project/study? (As a percentage of all respondents from the career stage)**

What is your career stage?	1 (Not at all confident)	2	3	4	5 (Very confident)
Phase 1 - Junior	23.53	31.76	35.29	9.41	0.0
Phase 2 - Early	5.26	15.79	42.11	31.58	5.26
Phase 3 - Mid / Recognised	6.82	22.73	25.0	25.0	20.45
Phase 4 - Established / Experienced / Senior	10.75	21.51	21.51	34.41	11.83

42

Confidence in Research Software Responsibilities During Study

25.18 Are you confident that you understand the policies and responsibilities related to research software that are placed on you after a project/study has ended? (As a percentage of all respondents from the career stage)

What is your career stage?	1 (Not at all confident)	2	3	4	5 (Very confident)
Phase 1 - Junior	23.53	31.76	35.29	9.41	0.0
Phase 2 - Early	5.26	15.79	42.11	31.58	5.26
Phase 3 - Mid / Recognised	6.82	22.73	25.0	25.0	20.45
Phase 4 - Established / Experienced / Senior	10.75	21.51	21.51	34.41	11.83

43

Confidence in Research Software Responsibilities After Study

26 Analyses Grouped by Discipline

Discipline	N	Percent
History	94	28.14
Other	64	19.16
Languages and Literature	59	17.66
Digital Humanities	54	16.17
Archaeology	38	11.38
Cultural and Museum Studies	35	10.48
Music and Visual Arts	35	10.48
Media	31	9.28
Design	24	7.19
Linguistics	23	6.89
Information and Communication Technologies	23	6.89
Library and Information Studies	20	5.99
Philosophy	19	5.69
Classics	14	4.19
Drama and Theatre Studies	12	3.59
Development Studies	9	2.69
Divinity and Religion	9	2.69
Theology	9	2.69
Dance	5	1.5
Law and Legal Studies	5	1.5
Political Science and International Studies	5	1.5

44

26.1 What is your career stage? (As a percentage of all respondents from the discipline)

	Phase 1 - Junior	Phase 2 - Early	Phase 3 - Mid / Recognised	Phase 4 - Established / Experienced / Senior
History	42.55	15.96	15.96	25.53
Languages and Literature	40.68	6.78	18.64	33.9
Digital Humanities	14.81	22.22	35.19	27.78
Archaeology	36.84	15.79	21.05	26.32
Cultural and Museum Studies	45.71	11.43	22.86	20.0
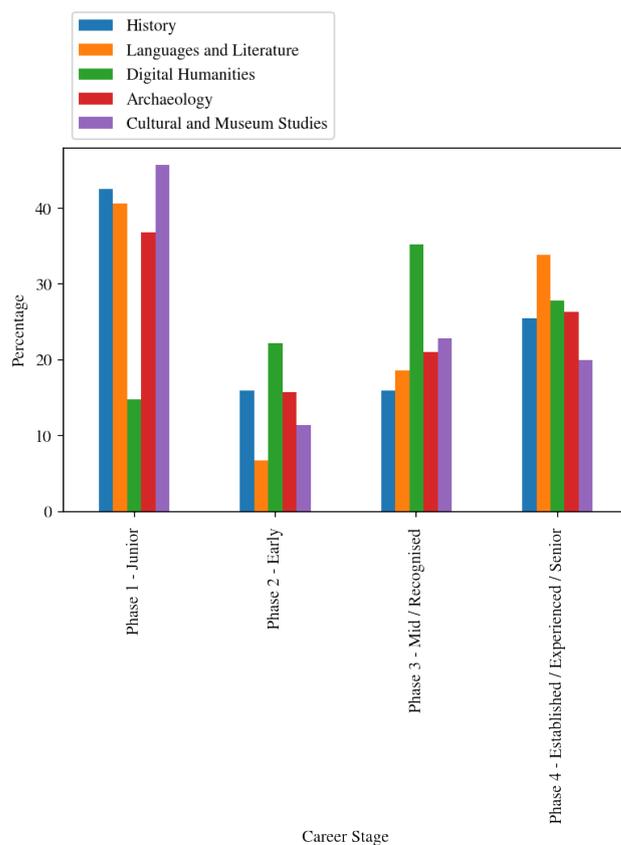
26.2 Over the last five years, what types of research data have you or your team worked with? (As a percentage of all respondents from the discipline)

	Textual	Image	Audio	Physical artefacts	Video	Numerical
History	93.18	78.41	37.5	51.14	20.45	39.77
Languages and Literature	94.83	63.79	41.38	29.31	29.31	18.97
Digital Humanities	94.23	76.92	44.23	48.08	34.62	57.69
Archaeology	73.68	86.84	28.95	76.32	28.95	63.16
Cultural and Museum Studies	90.91	81.82	51.52	45.45	39.39	30.3

45
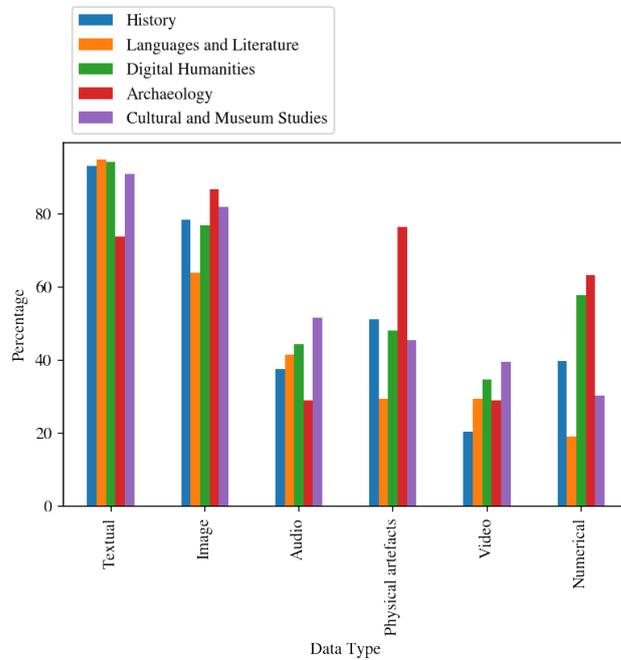
26.3 Over the last five years - once you have completed a research project/study - which of the following options have you used to store the data that were generated? (As a percentage of all respondents from the discipline)

	In a shared area provided by your institution	On a laptop or desktop computer	In a shared area provided by yourself	In a data repository provided by your instution	In a data repository outside of your institution
History	68.18	81.82	55.68	30.68	18.18
Languages and Literature	70.69	91.38	63.79	32.76	13.79
Digital Humanities	88.46	73.08	55.77	53.85	34.62
Archaeology	65.79	71.05	55.26	36.84	31.58
Cultural and Museum Studies	81.82	63.64	45.45	30.3	21.21

46
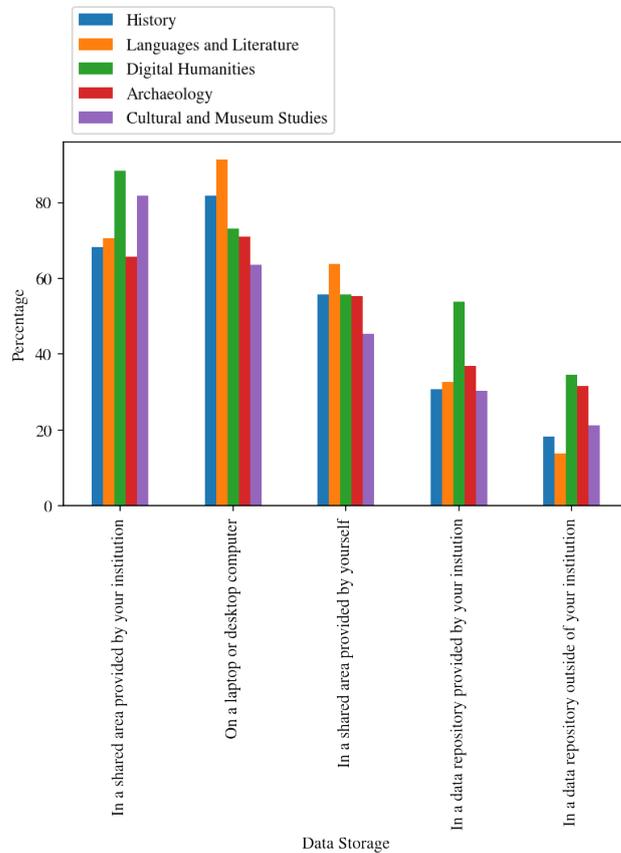
26.4 Which of the following options describes your typical approach to storing the data that were generated? (As a percentage of all respondents from the discipline)

	On a laptop or desktop computer	In a shared area provided by yourself	In a shared area provided by your institution	In a data repository provided by your institution	In a data repository outside of your institution	Not applicable
History	25.0	21.59	32.95	7.95	5.68	6.82
Languages and Literature	39.66	18.97	32.76	5.17	1.72	1.72
Digital Humanities	13.46	21.15	36.54	13.46	7.69	7.69
Archaeology	21.05	7.89	36.84	5.26	15.79	13.16
Cultural and Museum Studies	21.21	18.18	33.33	9.09	9.09	9.09

Typical Method of Data Storage
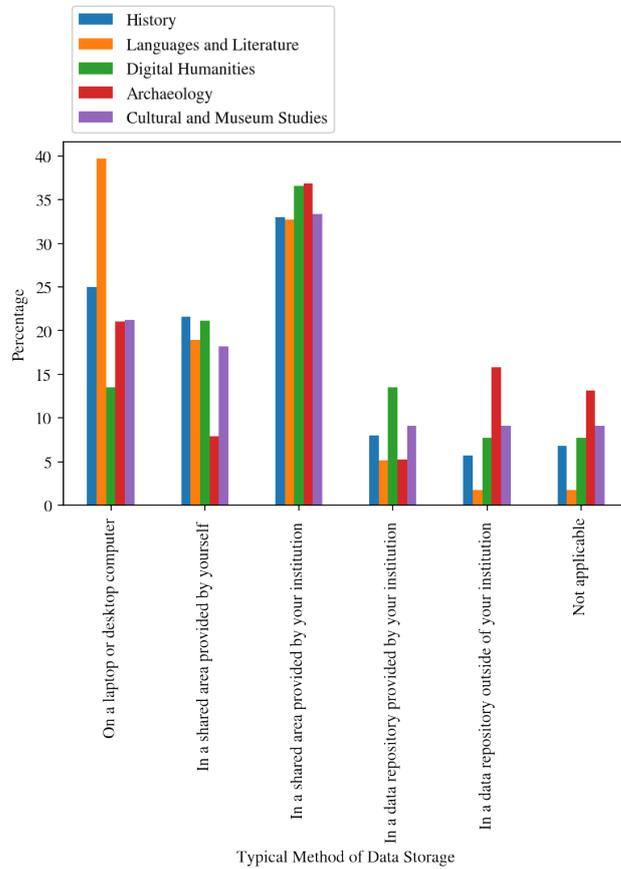
26.5 Data Sharing Frequency - Over the last five years, which of the following statements best applies to your research data? (As a percentage of all respondents from the discipline)

	I do not have research data to share	I haven't shared my research data	I have shared my research data but I rarely do so	I occasionally share my research data	I regularly share my research data
History	7.95	19.32	22.73	22.73	27.27
Languages and Literature	6.9	24.14	18.97	27.59	22.41
Digital Humanities	3.85	9.62	11.54	19.23	55.77
Archaeology	2.63	18.42	10.53	26.32	42.11
Cultural and Museum Studies	3.03	42.42	15.15	12.12	27.27

48

History
Languages and Literature
Digital Humanities
Archaeology
Cultural and Museum Studies

Percentage

I do not have research data to share

I haven't shared my research data

I have shared my research data but I rarely do so

I occasionally share my research data

I regularly share my research data

Data Sharing Frequency
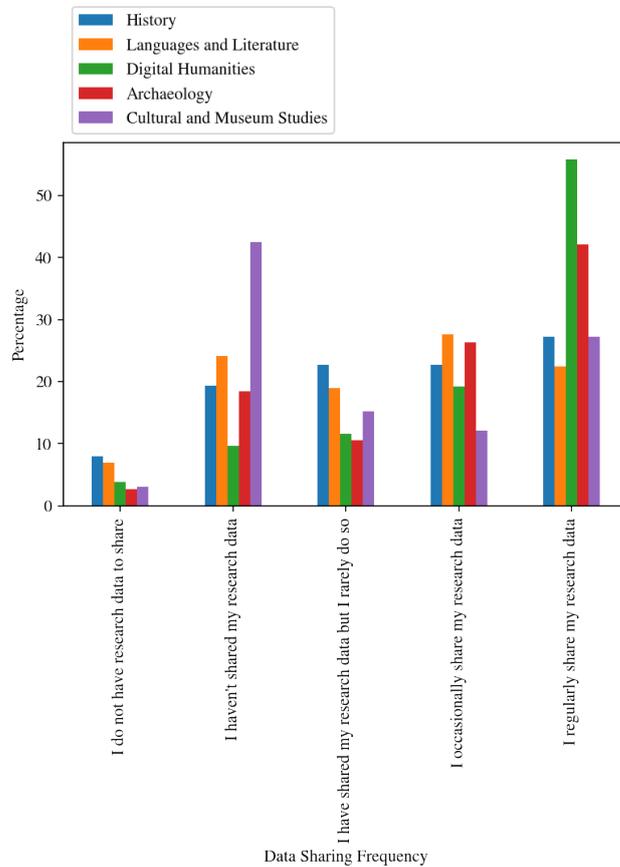
26.6 Over the last five years, what best describes your experience of data management plans on your projects? (As a percentage of all respondents from the discipline)

	I haven't been involved in research that included a data management plan	In my experience data management plans are not followed	In my experience data management plans are followed in part	In my experience data management plans are mostly followed	In my experience data management plans are followed exactly
History	39.77	5.68	18.18	32.95	3.41
Languages and Literature	43.1	1.72	17.24	36.21	1.72
Digital Humanities	9.62	3.85	23.08	59.62	3.85
Archaeology	21.05	7.89	15.79	50.0	5.26
Cultural and Museum Studies	21.21	6.06	24.24	33.33	15.15

Legend:
- History
- Languages and Literature
- Digital Humanities
- Archaeology
- Cultural and Museum Studies

Percentage

I haven't been involved in research that included a data management plan

In my experience data management plans are not followed

In my experience data management plans are followed in part

In my experience data management plans are mostly followed

In my experience data management plans are followed exactly

Data Management Plan Experience

50

26.7 Data Sharing Methods - Over the last five years, which of the following practices have been part of your standard research process? (As a percentage of all respondents from the discipline)

	I license my data to allow it to be shared	I share my data with individuals/groups that have requested access	I create a unique identifier to make my data findable	I cite my research data in publications
History	20.45	52.27	19.32	54.55
Languages and Literature	13.79	51.72	17.24	56.9
Digital Humanities	48.08	63.46	51.92	67.31
Archaeology	26.32	60.53	34.21	60.53
Cultural and Museum Studies	27.27	39.39	15.15	51.52

51
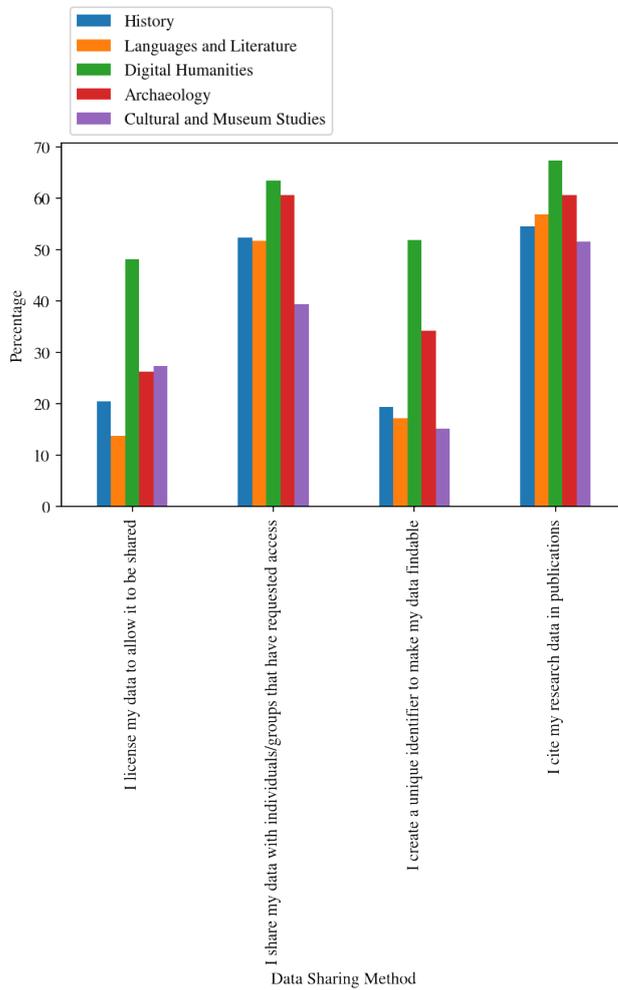
26.8 Research Software Use - Over the last five years, which of the following have applied to your research? (As a percentage of all respondents from the discipline)

	I have not used research software	I have used research software provided by university	I have purchased specialist research software with research funding	I have used specialist research software which was provided for free	I have used research software that was developed for me	I have developed my own research software
History	32.18	47.13	13.79	33.33	9.2	10.34
Languages and Literature	35.09	47.37	12.28	33.33	10.53	8.77
Digital Humanities	4.0	58.0	32.0	62.0	32.0	26.0
Archaeology	18.42	65.79	15.79	31.58	0.0	5.26
Cultural and Museum Studies	24.24	57.58	15.15	33.33	6.06	12.12
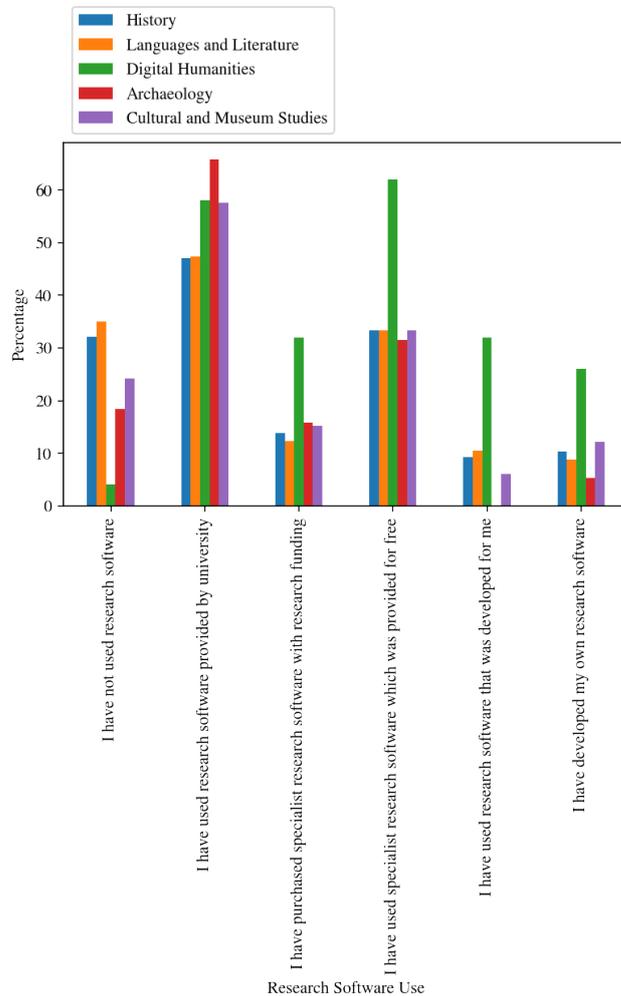
26.9 **Over the last five years, at the end of a research project/study, where have you typically stored the research software that was created? (N=39) (As a percentage of all respondents from the discipline)**

	In a repository outside my institution	In an institutional software repository	On a laptop/desktop computer	In a shared area provided by your institution
History	53.85	15.38	23.08	0.0
Languages and Literature	37.5	25.0	12.5	0.0
Digital Humanities	75.0	15.0	0.0	0.0
Archaeology	50.0	0.0	0.0	0.0
Cultural and Museum Studies	33.33	50.0	16.67	0.0

53

Research Software Storage Locations
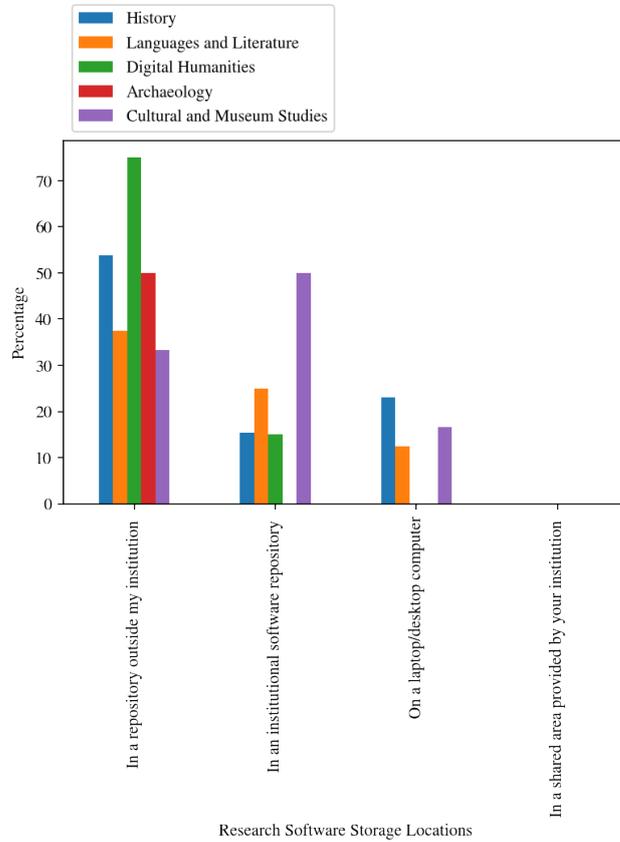
26.10 Over the last five years, which of the following practices have been part of your standard research process? (Research software sharing) (N = 39) (As a percentage of all respondents from the discipline)

	I do not share my research software	I license my research software to allow to be shared	I share my research software with individuals/groups that have requested access	I cite my software in publications and other outputs from my research	I create a unique identifier to make my research software findable
History	0.0	76.92	53.85	38.46	15.38
Languages and Literature	0.0	75.0	50.0	37.5	25.0
Digital Humanities	0.0	75.0	50.0	60.0	30.0
Archaeology	0.0	0.0	100.0	50.0	0.0
Cultural and Museum Studies	0.0	83.33	33.33	33.33	33.33

54

Methods of Sharing Research Software

Legend:
- History
- Languages and Literature
- Digital Humanities
- Archaeology
- Cultural and Museum Studies

X-axis categories:
- I do not share my research software
- I license my research software to allow to be shared
- I share my research software with individuals/groups that have requested access
- I cite my software in publications and other outputs from my research
- I create a unique identifier to make my research software findable

26.11 Who develops your research software? (N = 39) (As a percentage of all respondents from the discipline)

	I develop my own research software	PhD students and post-doctoral researchers	Software developers from industry	Research Software Engineers from within academia
History	53.85	46.15	15.38	76.92
Languages and Literature	50.0	25.0	25.0	87.5
Digital Humanities	40.0	50.0	25.0	85.0
Archaeology	100.0	0.0	50.0	50.0
Cultural and Museum Studies	66.67	50.0	33.33	83.33

Who Develops Research Software
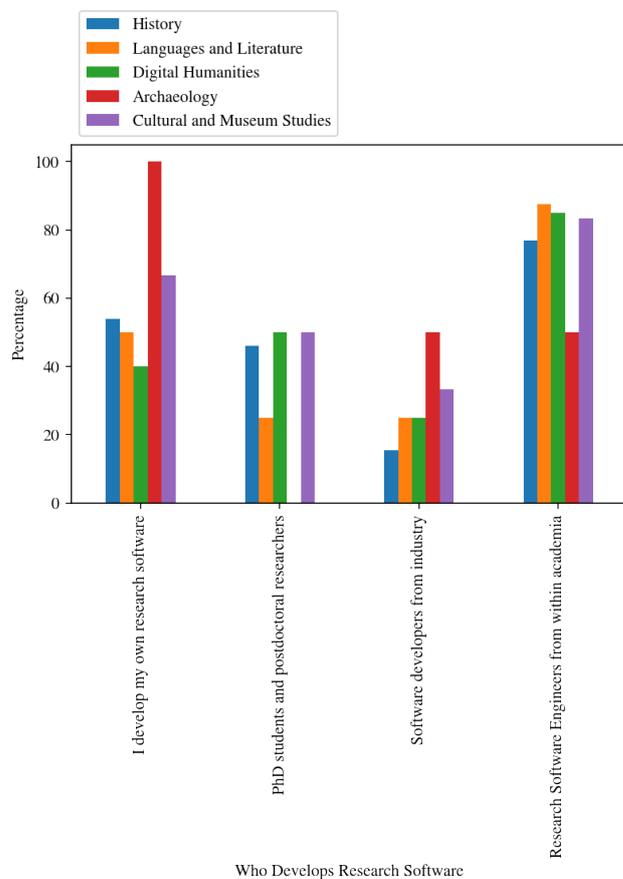
26.12 Data Skills - Over the last five years, have you learned new skills or knowledge in relation to any of the following? (As a percentage of all respondents from the discipline)

	Data management/storage	Data analysis	Data sharing	Data collection
History	64.2	55.56	44.44	40.74
Languages and Literature	66.04	50.94	45.28	43.4
Digital Humanities	86.96	82.61	76.09	65.22
Archaeology	75.68	70.27	56.76	51.35
Cultural and Museum Studies	75.0	68.75	37.5	62.5

Data Skills Learnt
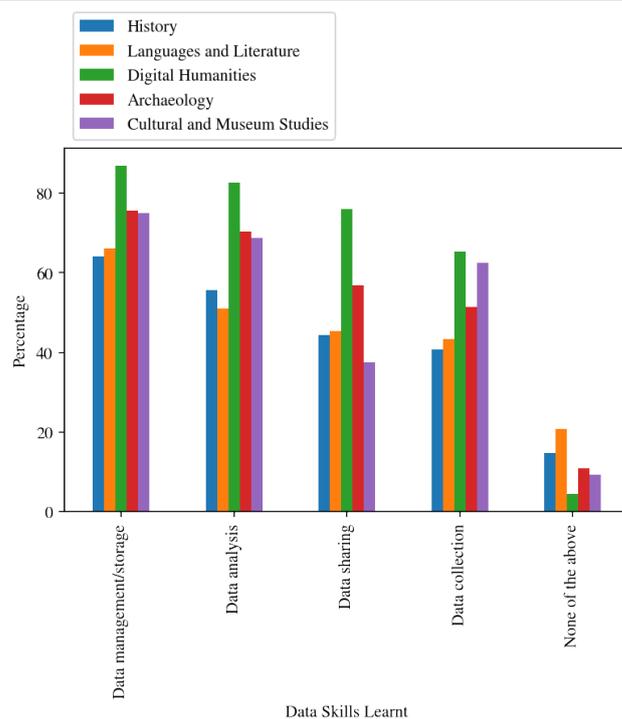
26.13 Research Software Skills - Over the last five years, have you learned new skills or knowledge in relation to any of the following? (As a percentage of all respondents from the discipline)

	None of the above	Software sharing	Software development	Software requirements collection
History	60.49	30.86	22.22	12.35
Languages and Literature	62.26	30.19	15.09	11.32
Digital Humanities	26.09	54.35	45.65	28.26
Archaeology	62.16	16.22	18.92	21.62
Cultural and Museum Studies	53.12	28.12	31.25	28.12

Software Skills Learnt
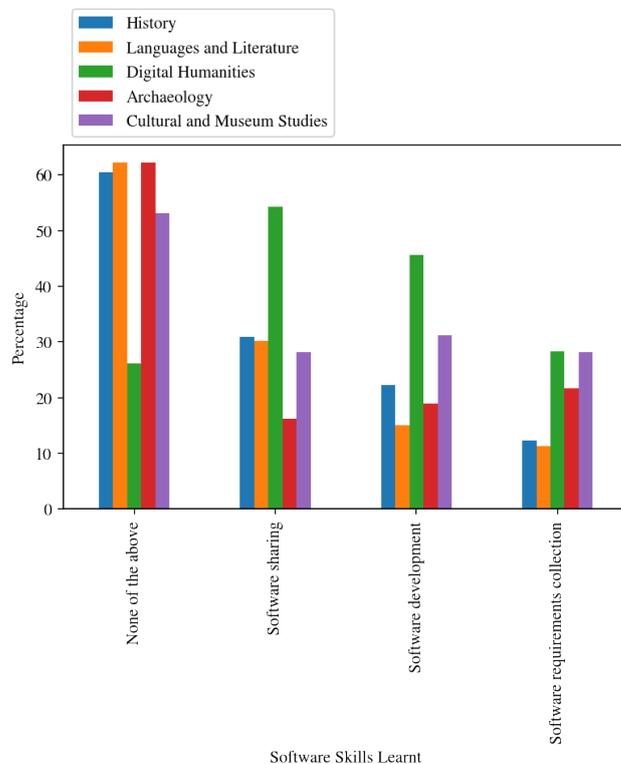
26.14 Where did you acquire your skills and knowledge about data and/or software? (As a percentage of all respondents from the discipline)

	Learning on the job – self-taught	Learning on the job – peer taught	Training/advice provided in-house by my institution	Masters/Doctoral training	Training/advice provided externally
History	69.14	56.79	33.33	33.33	13.58
Languages and Literature	64.15	45.28	26.42	30.19	9.43
Digital Humanities	78.26	76.09	47.83	21.74	19.57
Archaeology	86.49	51.35	32.43	43.24	24.32
Cultural and Museum Studies	62.5	53.12	43.75	40.62	25.0

58

Where Skills Learnt
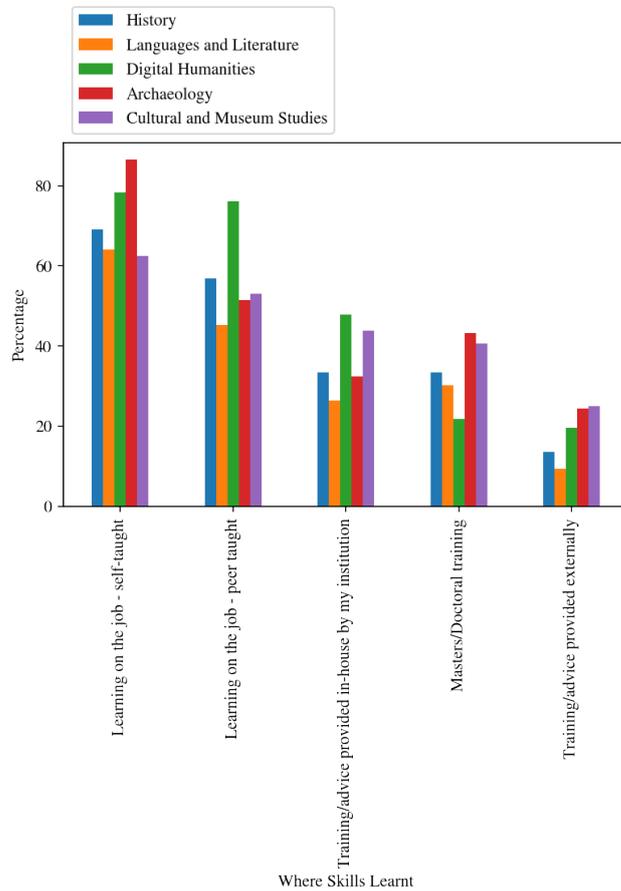
26.15 **Do any of the following issues prevent you from gaining the skill you entered above (the main skill you want to improve)? (As a percentage of all respondents from the discipline)**

	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 with them	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	None of the above
History	35.8	35.8	19.75	13.58	19.75
Languages and Literature	26.42	41.51	11.32	5.66	22.64
Digital Humanities	21.74	56.52	23.91	10.87	8.7
Archaeology	16.22	54.05	24.32	10.81	18.92
Cultural and Museum Studies	43.75	37.5	15.62	12.5	9.38

Legend:
- History
- Languages and Literature
- Digital Humanities
- Archaeology
- Cultural and Museum Studies

Percentage

- 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 with them
- 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
- None of the above

Barriers to Learning Skills

60

26.16 Are you confident that you understand the policies and responsibilities related to research data that are placed on you during a project/study? (As a percentage of all respondents from the discipline)

	1 (Not at all confident)	2	3	4	5 (Very confident)
History	3.8	10.13	21.52	46.84	17.72
Languages and Literature	1.96	7.84	37.25	45.1	7.84
Digital Humanities	0.0	2.17	15.22	54.35	28.26
Archaeology	0.0	2.7	24.32	54.05	18.92
Cultural and Museum Studies	6.25	3.12	18.75	50.0	21.88

Confidence in Data Responsibilities During Study
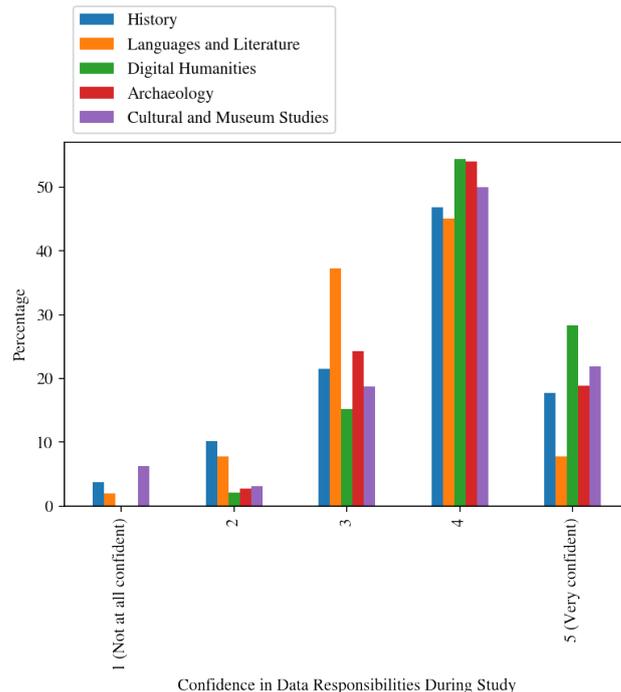
26.17 Are you confident that you understand the policies and responsibilities related to research data that are placed on you after a project/study has ended? (As a percentage of all respondents from the discipline)

	1 (Not at all confident)	2	3	4	5 (Very confident)
History	7.59	13.92	29.11	35.44	13.92
Languages and Literature	6.0	16.0	38.0	30.0	10.0
Digital Humanities	0.0	11.11	20.0	48.89	20.0
Archaeology	2.7	10.81	32.43	43.24	10.81
Cultural and Museum Studies	9.38	12.5	25.0	37.5	15.62

Confidence in Data Responsibilities After Study
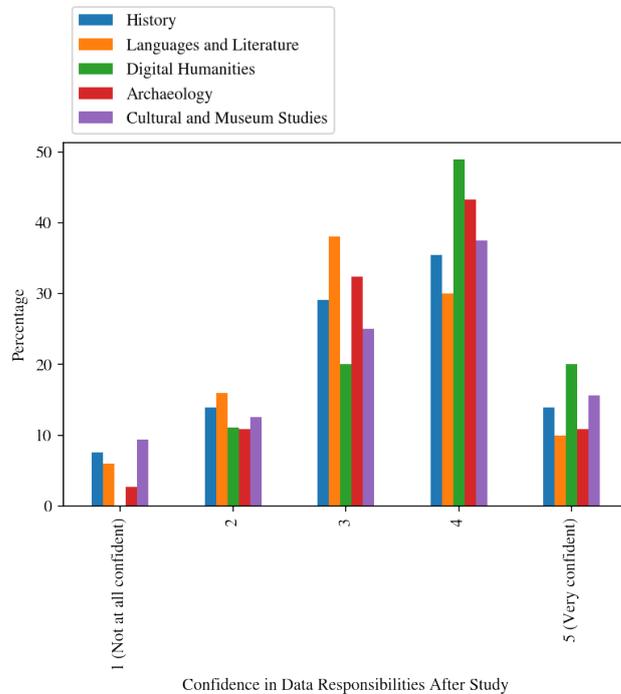
26.18 **Are you confident that you understand the policies and responsibilities related to research software that are placed on you during a project/study? (As a percentage of all respondents from the discipline)**

	1 (Not at all confident)	2	3	4	5 (Very confident)
History	10.53	19.74	32.89	25.0	11.84
Languages and Literature	10.2	22.45	38.78	22.45	6.12
Digital Humanities	0.0	13.04	30.43	43.48	13.04
Archaeology	8.57	14.29	40.0	28.57	8.57
Cultural and Museum Studies	16.67	3.33	26.67	36.67	16.67

62

Confidence in Research Software Responsibilities During Study
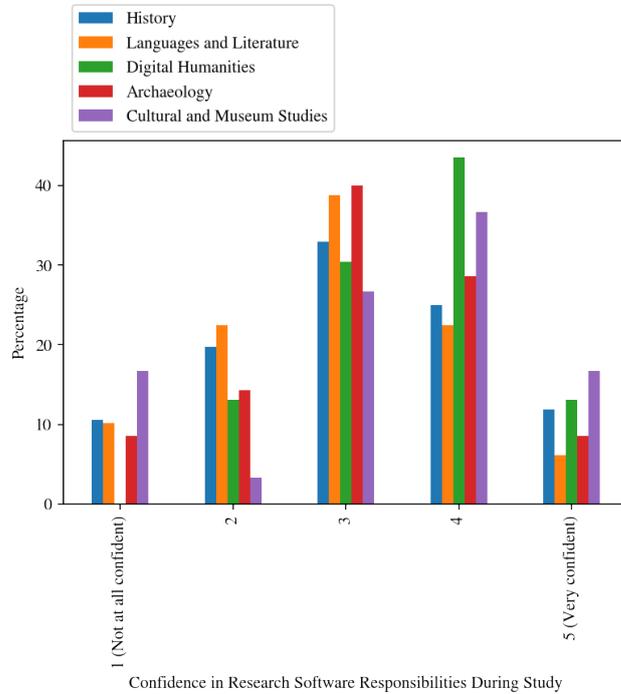
26.19   Are you confident that you understand the policies and responsibilities related to research software that are placed on you after a project/study has ended? (As a percentage of all respondents from the discipline)

	1 (Not at all confident)	2	3	4	5 (Very confident)
History	13.16	26.32	26.32	26.32	7.89
Languages and Literature	12.24	28.57	34.69	18.37	6.12
Digital Humanities	0.0	22.22	28.89	40.0	8.89
Archaeology	8.82	23.53	44.12	20.59	2.94
Cultural and Museum Studies	23.33	13.33	23.33	30.0	10.0

Confidence in Research Software Responsibilities After Study
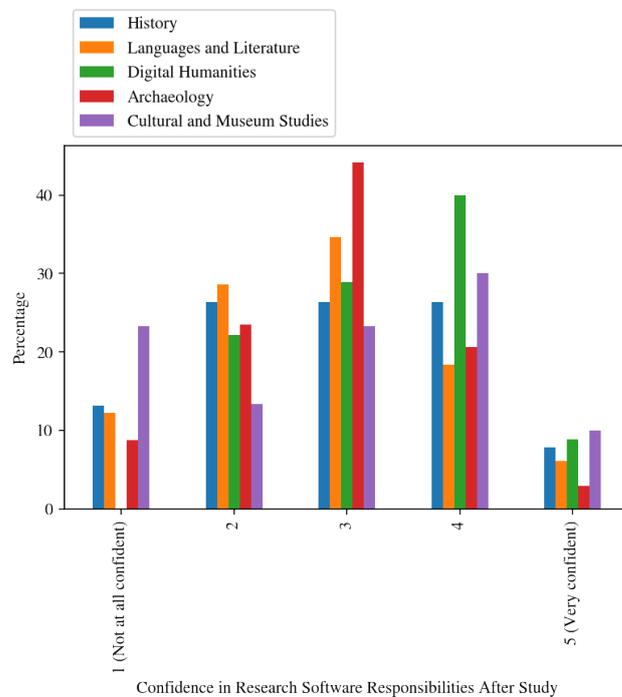
27 Analyses Grouped by Institution (Five institutions with the most respondents included)

Institution	N	Percent
University of Glasgow	15	4.66
University of Oxford	13	4.04
University of York	13	4.04
University of Liverpool	12	3.73
University of Birmingham	12	3.73

27.1 What is your career stage? (As a percentage of all respondents from the institution)

	Phase 1 - Junior	Phase 2 - Early	Phase 3 - Mid / Recognised	Phase 4 - Established / Experienced / Senior
University of Glasgow	60.0	0.0	6.67	33.33
University of Oxford	69.23	15.38	0.0	15.38
University of York	15.38	46.15	7.69	30.77
University of Liverpool	66.67	16.67	16.67	0.0
University of Birmingham	25.0	16.67	16.67	41.67
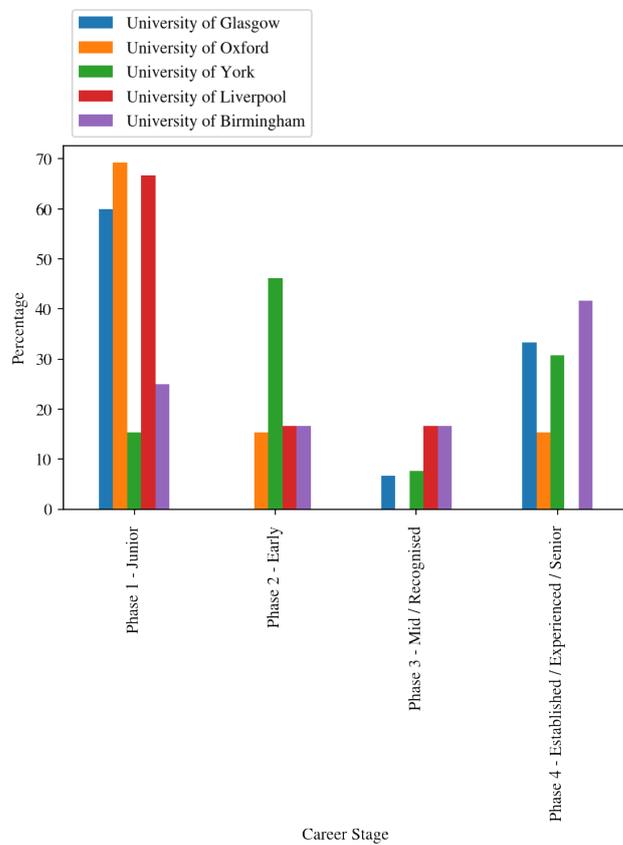
27.2 **Over the last five years, what types of research data have you or your team worked with? (As a percentage of all respondents from the institution)**

	Textual	Image	Audio	Physical artefacts	Video	Numerical
University of Glasgow	100.0	86.67	66.67	66.67	33.33	20.0
University of Oxford	100.0	76.92	7.69	53.85	7.69	15.38
University of York	92.31	84.62	46.15	61.54	30.77	76.92
University of Liverpool	80.0	60.0	40.0	30.0	10.0	30.0
University of Birmingham	91.67	91.67	66.67	33.33	50.0	16.67
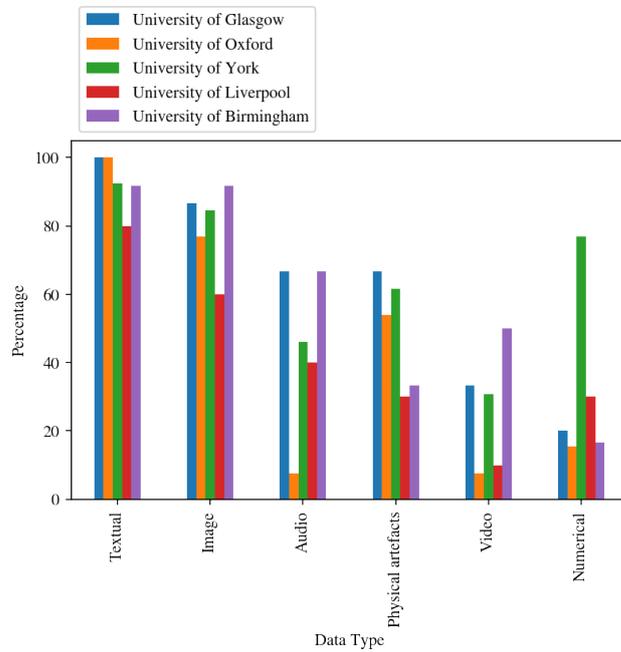
27.3 Over the last five years - once you have completed a research project/study - which of the following options have you used to store the data that were generated? (As a percentage of all respondents from the institution)

	In a shared area provided by your institution	On a laptop or desktop computer	In a shared area provided by yourself	In a data repository provided by your instution	In a data repository outside of your institution
University of Glasgow	86.67	66.67	26.67	33.33	20.0
University of Oxford	46.15	100.0	61.54	23.08	7.69
University of York	69.23	76.92	46.15	53.85	38.46
University of Liverpool	70.0	80.0	50.0	20.0	20.0
University of Birmingham	66.67	83.33	75.0	58.33	16.67

66
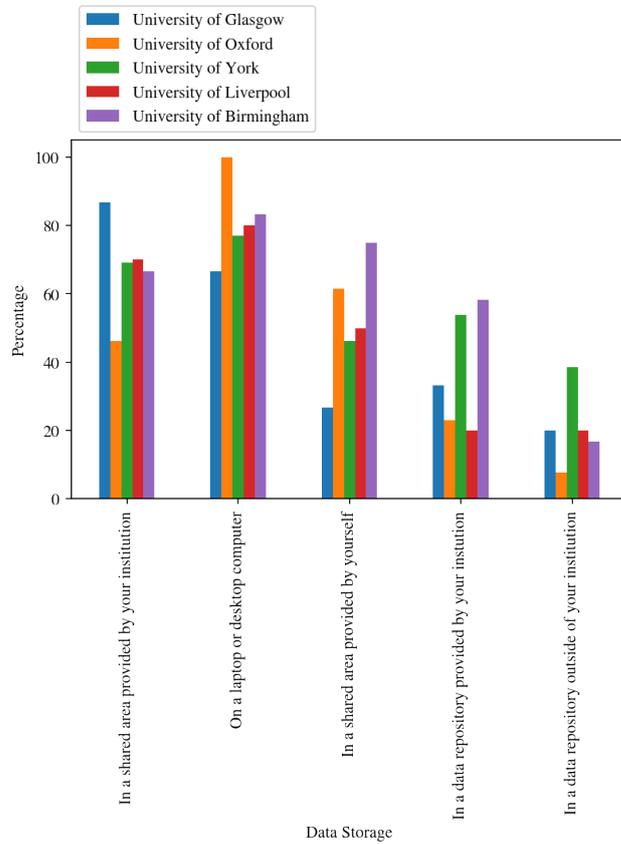
27.4 Which of the following options describes your typical approach to storing the data that were generated? (As a percentage of all respondents from the institution)

	On a laptop or desktop computer	In a shared area provided by yourself	In a shared area provided by your institution	In a data repository provided by your institution	In a data repository outside of your institution	Not applicable
University of Glasgow	20.0	6.67	60.0	6.67	6.67	0.0
University of Oxford	53.85	23.08	15.38	0.0	0.0	7.69
University of York	15.38	7.69	53.85	7.69	7.69	7.69
University of Liverpool	40.0	10.0	40.0	10.0	0.0	0.0
University of Birmingham	8.33	33.33	25.0	33.33	0.0	0.0

Typical Method of Data Storage
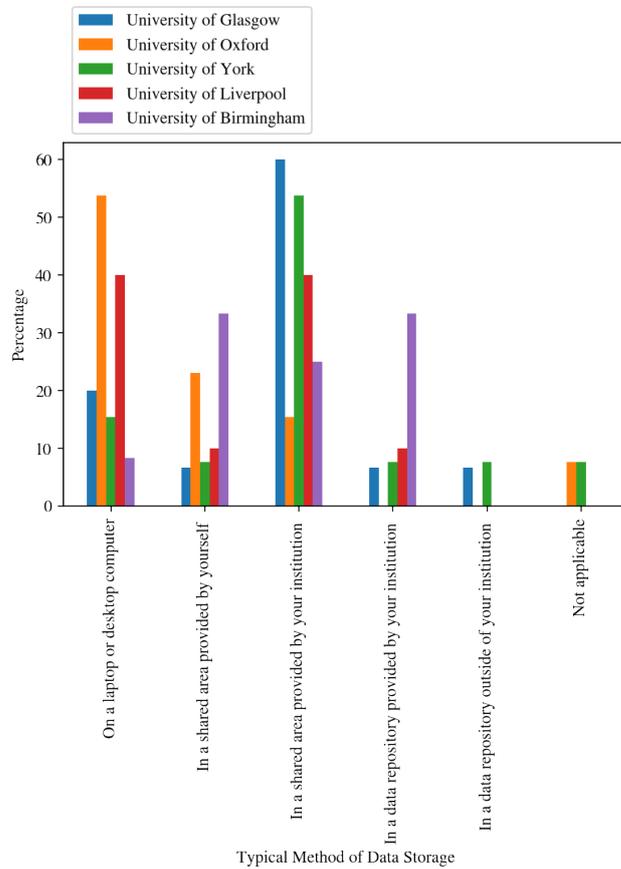
27.5 Data Sharing Frequency - Over the last five years, which of the following statements best applies to your research data? (As a percentage of all respondents from the institution)

	I do not have research data to share	I haven't shared my research data	I have shared my research data but I rarely do so	I occasionally share my research data	I regularly share my research data
University of Glasgow	6.67	20.0	26.67	26.67	20.0
University of Oxford	15.38	15.38	23.08	38.46	7.69
University of York	7.69	15.38	23.08	15.38	38.46
University of Liverpool	10.0	40.0	10.0	30.0	10.0
University of Birmingham	0.0	16.67	8.33	25.0	50.0

68

Percentage

Data Sharing Frequency
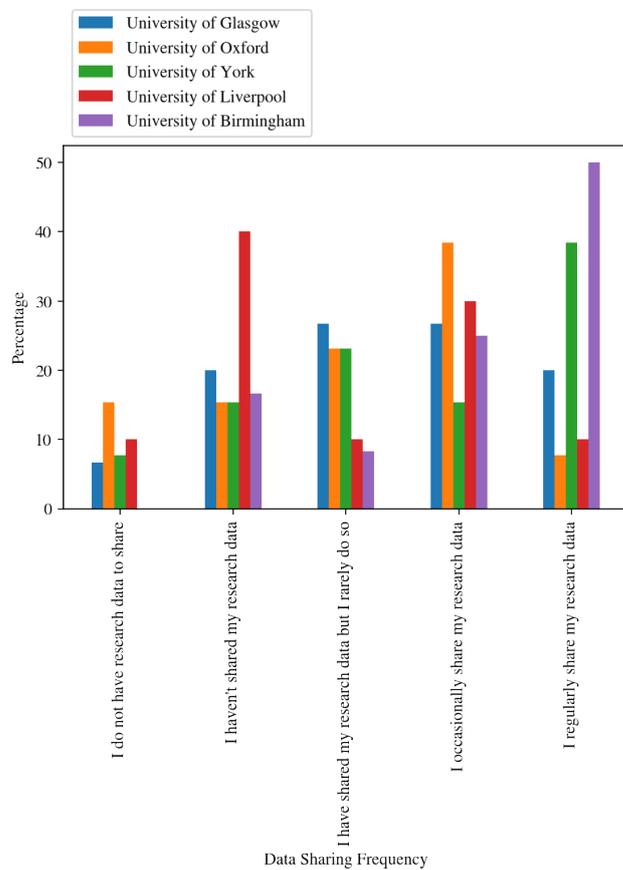
27.6   Over the last five years, what best describes your experience of data management plans on your projects? (As a percentage of all respondents from the institution)

	I haven't been involved in research that included a data management plan	In my experience data management plans are not followed	In my experience data management plans are followed in part	In my experience data management plans are mostly followed	In my experience data management plans are followed exactly
University of Glasgow	13.33	0.0	20.0	66.67	0.0
University of Oxford	76.92	7.69	0.0	15.38	0.0
University of York	15.38	15.38	0.0	61.54	7.69
University of Liverpool	60.0	10.0	0.0	30.0	0.0
University of Birmingham	8.33	0.0	25.0	66.67	0.0

69

Legend:
- University of Glasgow
- University of Oxford
- University of York
- University of Liverpool
- University of Birmingham

Y-axis: Percentage

X-axis categories (Data Management Plan Experience):
- I haven't been involved in research that included a data management plan
- In my experience data management plans are not followed
- In my experience data management plans are followed in part
- In my experience data management plans are mostly followed
- In my experience data management plans are followed exactly

27.7 Data Sharing Methods - Over the last five years, which of the following practices have been part of your standard research process? (As a percentage of all respondents from the institution)

	I license my data to allow it to be shared	I share my data with individuals/groups that have requested access	I create a unique identifier to make my data findable	I cite my research data in publications
University of Glasgow	13.33	60.0	20.0	73.33
University of Oxford	7.69	46.15	15.38	30.77
University of York	30.77	38.46	30.77	92.31
University of Liverpool	20.0	60.0	20.0	30.0
University of Birmingham	25.0	41.67	50.0	75.0
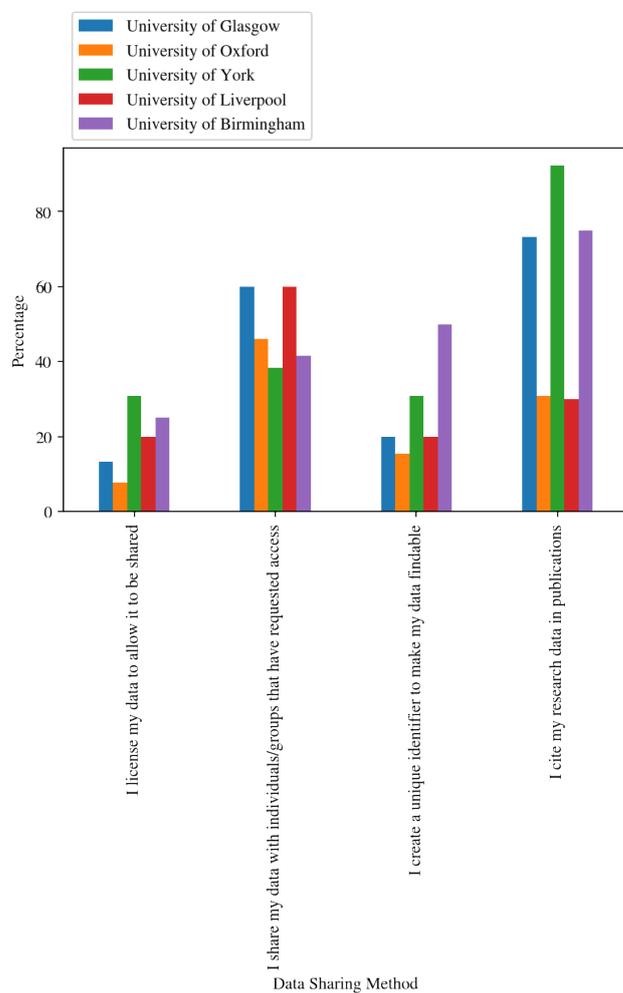
27.8 Research Software Use - Over the last five years, which of the following have applied to your research? (As a percentage of all respondents from the institution)

	I have not used research software	I have used research software provided by university	I have purchased specialist research software with research funding	I have used specialist research software which was provided for free	I have used research software that was developed for me	I have developed my own research software
University of Glasgow	20.0	66.67	20.0	20.0	13.33	0.0
University of Oxford	30.77	53.85	15.38	30.77	0.0	0.0
University of York	23.08	69.23	15.38	38.46	0.0	7.69
University of Liverpool	22.22	77.78	0.0	33.33	0.0	0.0
University of Birmingham	25.0	33.33	41.67	50.0	25.0	16.67

Legend:
- University of Glasgow
- University of Oxford
- University of York
- University of Liverpool
- University of Birmingham

Y-axis: Percentage
X-axis: Research Software Use

Categories:
- I have not used research software
- I have used research software provided by university
- I have purchased specialist research software with research funding
- I have used specialist research software which was provided for free
- I have used research software that was developed for me
- I have developed my own research software

27.9 Over the last five years, at the end of a research project/study, where have you typically stored the research software that was created? (N=39) (As a percentage of all respondents from the institution)

|	In a repository outside my institution	In an institutional software repository	On a laptop/desktop computer	In a shared area provided by your institution
University of Glasgow	50.0	50.0	0.0	0.0
University of Oxford	0.0	0.0	0.0	0.0
University of York	0.0	100.0	0.0	0.0
University of Liverpool	0.0	0.0	0.0	0.0
University of Birmingham	100.0	0.0	0.0	0.0

73

Research Software Storage Locations
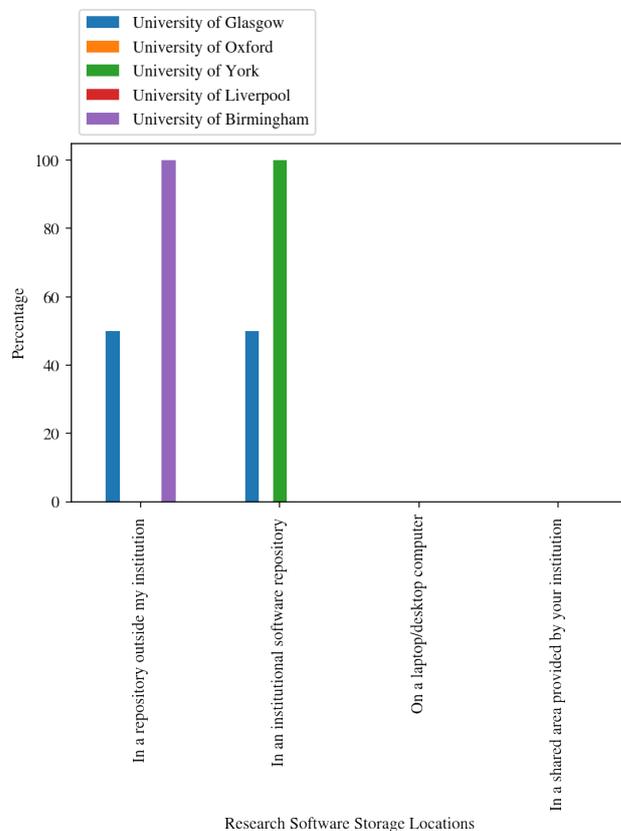
27.10   Over the last five years, which of the following practices have been part of your standard research process? (Research software sharing) (N = 39) (As a percentage of all respondents from the institution)

	I do not share my research software	I license my research software to allow to be shared	I share my research software with individuals/groups that have requested access	I cite my software in publications and other outputs from my research	I create a unique identifier to make my research software findable
University of Glasgow	0.0	50.0	50.0	50.0	50.0
University of Oxford	0.0	0.0	0.0	0.0	0.0
University of York	0.0	100.0	100.0	100.0	100.0
University of Liverpool	0.0	0.0	0.0	0.0	0.0
University of Birmingham	0.0	75.0	50.0	75.0	50.0

74

Methods of Sharing Research Software
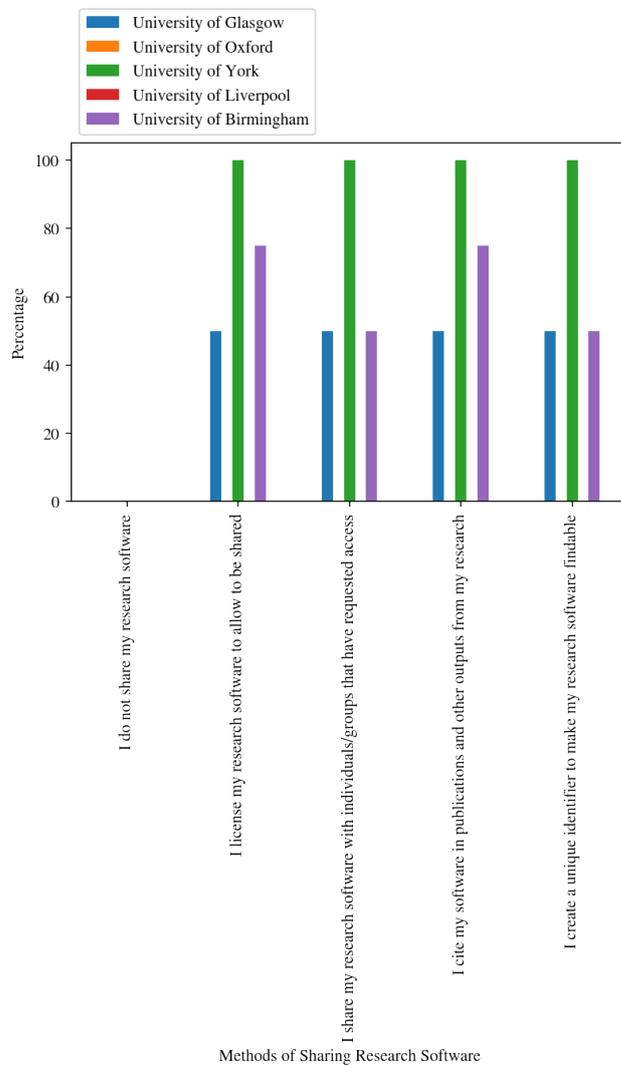
27.11 Who develops your research software? (N = 39) (As a percentage of all respondents from the institution)

	I develop my own research software	PhD students and post-doctoral researchers	Software developers from industry	Research Software Engineers from within academia
University of Glasgow	0.0	0.0	0.0	100.0
University of Oxford	0.0	0.0	0.0	0.0
University of York	0.0	100.0	0.0	0.0
University of Liverpool	0.0	0.0	0.0	0.0
University of Birmingham	25.0	25.0	25.0	75.0
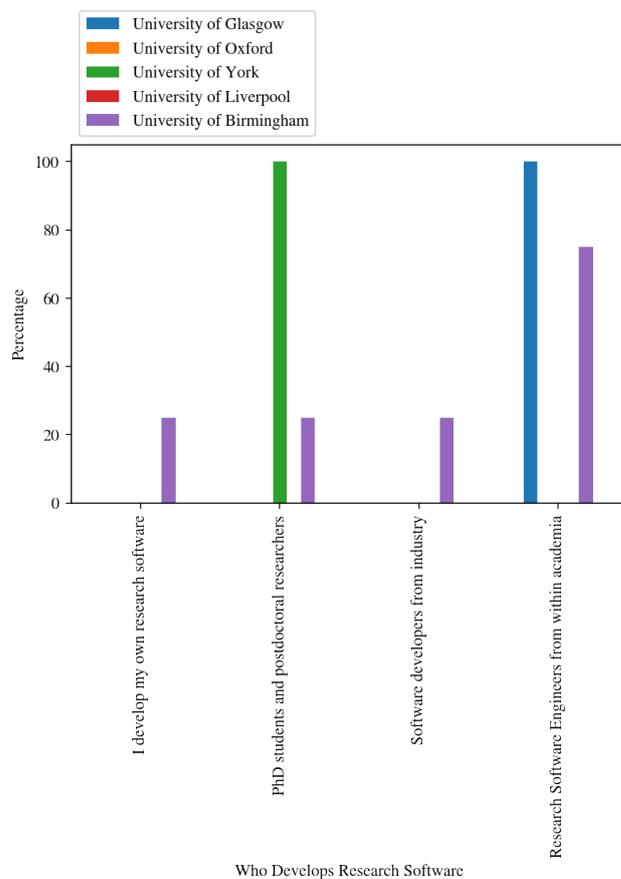
27.12 **Data Skills - Over the last five years, have you learned new skills or knowledge in relation to any of the following? (As a percentage of all respondents from the institution)**

	Data management/storage	Data analysis	Data sharing	Data collection
University of Glasgow	85.71	64.29	42.86	64.29
University of Oxford	18.18	45.45	36.36	45.45
University of York	76.92	76.92	53.85	38.46
University of Liverpool	55.56	77.78	33.33	44.44
University of Birmingham	72.73	63.64	63.64	18.18
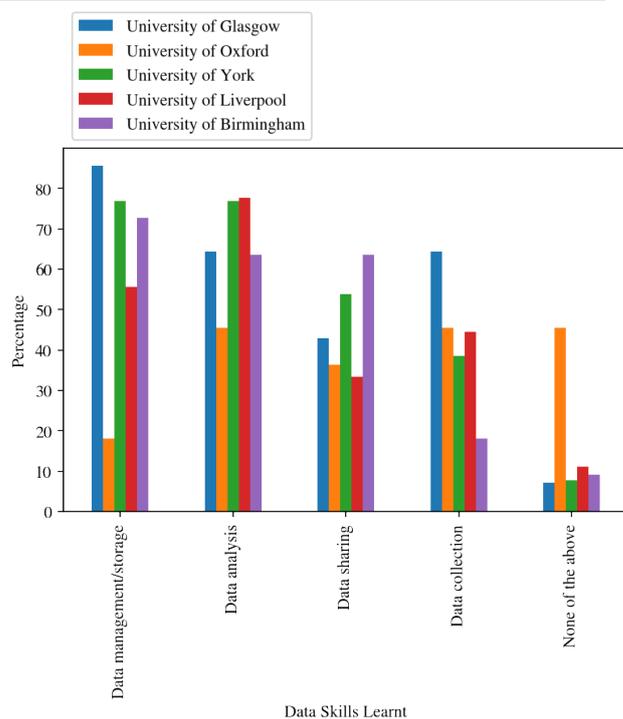
27.13 **Research Software Skills - Over the last five years, have you learned new skills or knowledge in relation to any of the following? (As a percentage of all respondents from the institution)**

	None of the above	Software sharing	Software development	Software requirements collection
University of Glasgow	71.43	21.43	14.29	14.29
University of Oxford	54.55	27.27	18.18	9.09
University of York	69.23	15.38	23.08	0.0
University of Liverpool	88.89	11.11	0.0	0.0
University of Birmingham	27.27	54.55	54.55	0.0
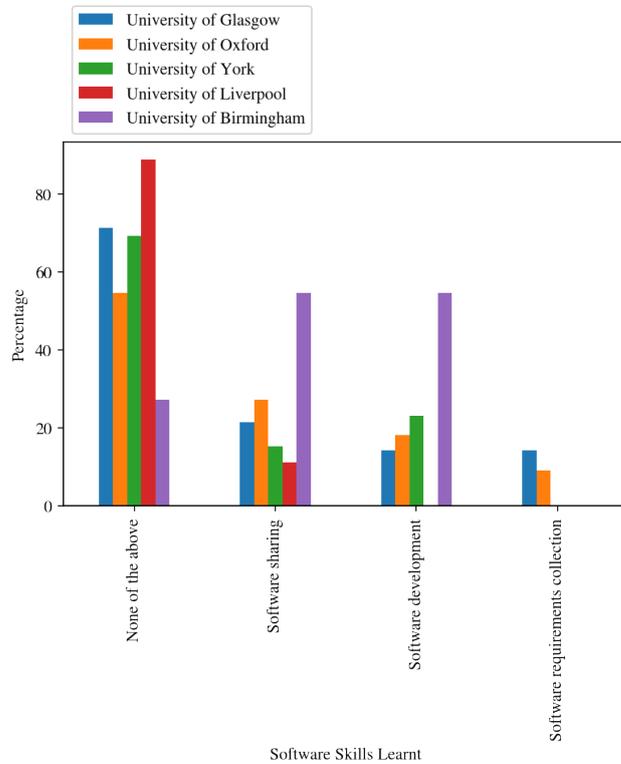
27.14 Where did you acquire your skills and knowledge about data and/or software? (As a percentage of all respondents from the institution)

	Learning on the job - self-taught	Learning on the job - peer taught	Training/advice provided in-house by my institution	Masters/Doctoral training	Training/advice provided externally
University of Glasgow	71.43	57.14	35.71	35.71	14.29
University of Oxford	81.82	27.27	27.27	54.55	0.0
University of York	100.0	53.85	53.85	46.15	30.77
University of Liverpool	88.89	22.22	44.44	66.67	11.11
University of Birmingham	63.64	63.64	36.36	18.18	0.0

Where Skills Learnt
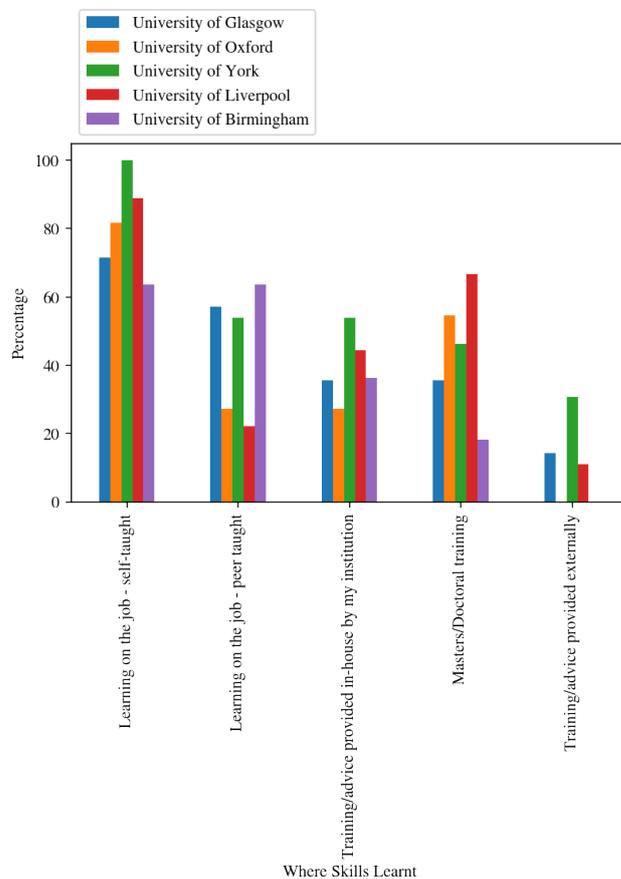
27.15 Do any of the following issues prevent you from gaining the skill you entered above (the main skill you want to improve)? (As a percentage of all respondents from the institution)

	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 with them	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	None of the above
University of Glasgow	21.43	50.0	21.43	14.29	28.57
University of Oxford	27.27	54.55	9.09	0.0	36.36
University of York	15.38	38.46	15.38	15.38	38.46
University of Liverpool	22.22	66.67	11.11	0.0	11.11
University of Birmingham	27.27	45.45	0.0	18.18	9.09

79

University of Glasgow
University of Oxford
University of York
University of Liverpool
University of Birmingham

Percentage

70

60

50

40

30

20

10

0

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 with them

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

None of the above

Barriers to Learning Skills

80

27.16 Are you confident that you understand the policies and responsibilities related to research data that are placed on you during a project/study? (As a percentage of all respondents from the institution)

	1 (Not at all confident)	2	3	4	5 (Very confident)
University of Glasgow	0.0	0.0	14.29	78.57	7.14
University of Oxford	10.0	20.0	50.0	20.0	0.0
University of York	0.0	0.0	15.38	61.54	23.08
University of Liverpool	0.0	0.0	66.67	22.22	11.11
University of Birmingham	0.0	0.0	18.18	54.55	27.27

Confidence in Data Responsibilities During Study
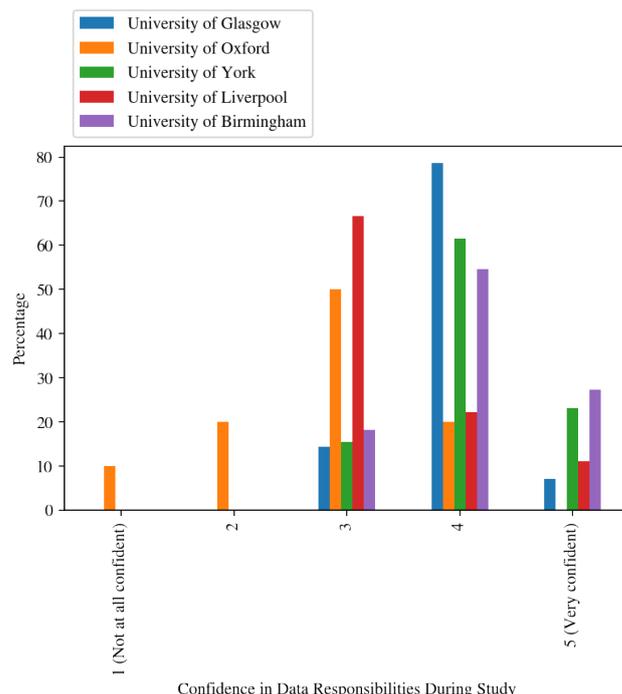
27.17 Are you confident that you understand the policies and responsibilities related to research data that are placed on you after a project/study has ended? (As a percentage of all respondents from the institution)

	1 (Not at all confident)	2	3	4	5 (Very confident)
University of Glasgow	0.0	7.14	21.43	57.14	14.29
University of Oxford	10.0	20.0	60.0	10.0	0.0
University of York	0.0	15.38	30.77	38.46	15.38
University of Liverpool	0.0	0.0	66.67	33.33	0.0
University of Birmingham	0.0	0.0	18.18	54.55	27.27

81

Confidence in Data Responsibilities After Study
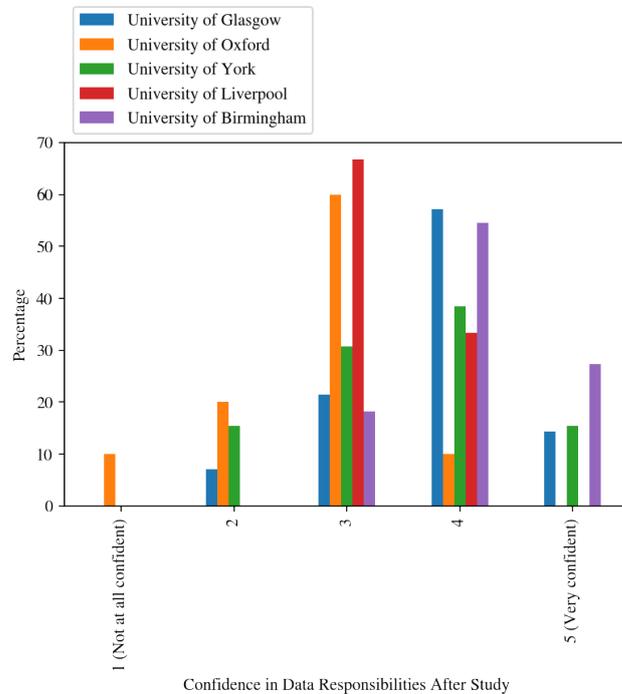
27.18 Are you confident that you understand the policies and responsibilities related to research software that are placed on you during a project/study? (As a percentage of all respondents from the institution)

	1 (Not at all confident)	2	3	4	5 (Very confident)
University of Glasgow	7.69	38.46	15.38	38.46	0.0
University of Oxford	10.0	20.0	60.0	10.0	0.0
University of York	0.0	16.67	25.0	41.67	16.67
University of Liverpool	0.0	22.22	55.56	22.22	0.0
University of Birmingham	0.0	27.27	27.27	27.27	18.18

82

Confidence in Research Software Responsibilities During Study
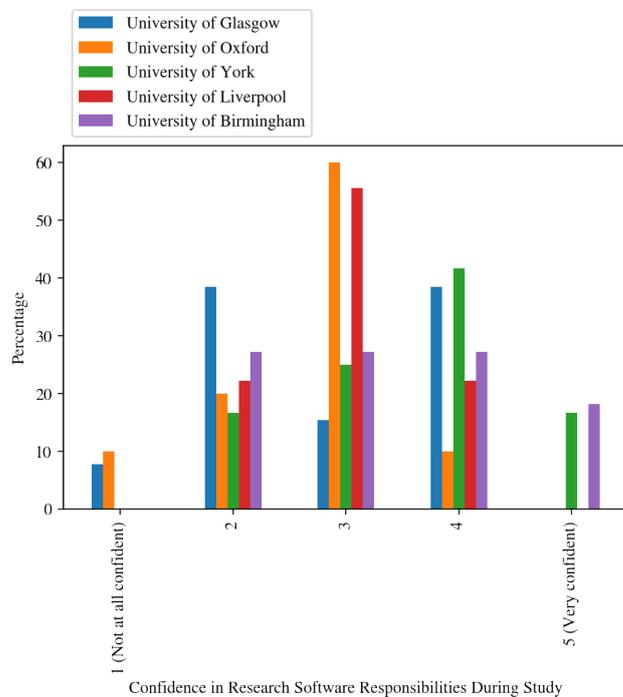
27.19 **Are you confident that you understand the policies and responsibilities related to research software that are placed on you after a project/study has ended? (As a percentage of all respondents from the institution)**

	1 (Not at all confident)	2	3	4	5 (Very confident)
University of Glasgow	7.69	53.85	15.38	23.08	0.0
University of Oxford	10.0	40.0	40.0	10.0	0.0
University of York	0.0	18.18	36.36	36.36	9.09
University of Liverpool	0.0	22.22	55.56	22.22	0.0
University of Birmingham	9.09	18.18	27.27	27.27	18.18

83

Confidence in Research Software Responsibilities After Study
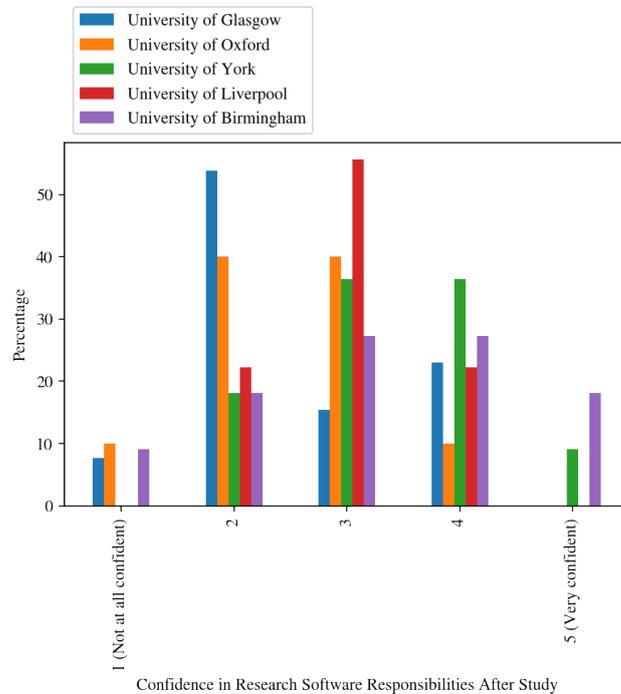
28 Are data-related practices different among respondents who work with physical artefacts? (N = 129)

28.1 Over the last five years - once you have completed a research project/study - which of the following options have you used to store the data that were generated? (As a percentage of respondents)

	On a laptop or desktop computer	In a shared area provided by yourself	In a shared area provided by your institution	In a data repository provided by your instution	In a data repository outside of your institution
Physical Artefacts	76.74	58.14	74.42	36.43	21.71
Overall	76.72	53.77	74.1	35.08	20.33

84

Percentage vs Data Storage bar chart with legend: Physical Artefacts, Overall
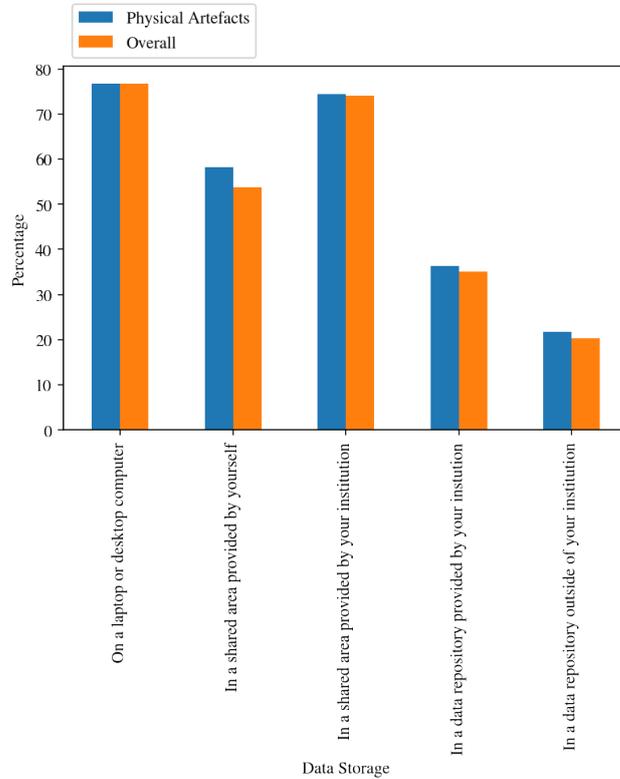
28.2   Which of the following options describes your typical approach to storing the data that were generated? (As a percentage of respondents)

	On a laptop or desktop computer	In a shared area provided by yourself	In a shared area provided by your institution	In a data repository provided by your institution	In a data repository outside of your institution	Not applicable
Physical Artefacts	26.36	20.93	31.78	8.53	6.2	6.2
Overall	26.23	17.38	38.03	7.87	5.9	4.59

Typical Method of Data Storage
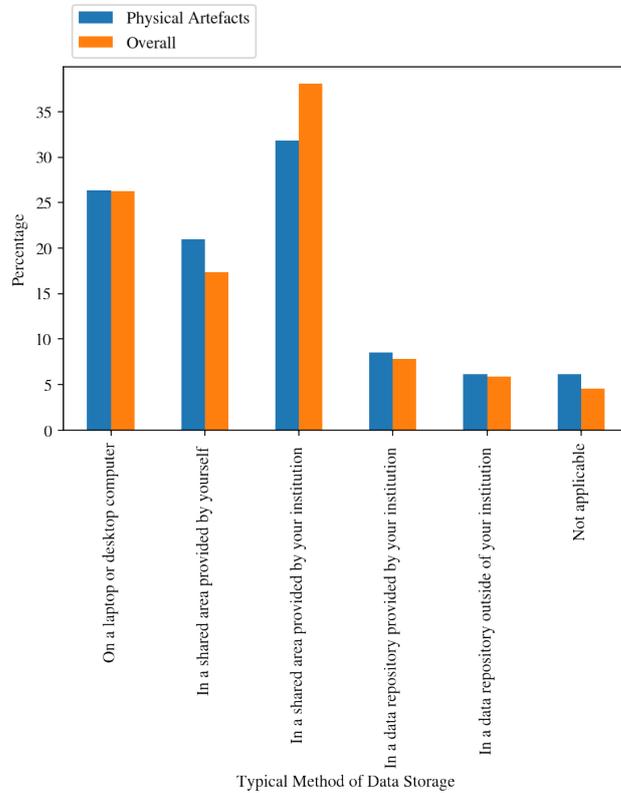
28.3 Data Sharing Frequency - Over the last five years, which of the following statements best applies to your research data? (As a percentage of respondents)

	I do not have research data to share	I haven't shared my research data	I have shared my research data but I rarely do so	I occasionally share my research data	I regularly share my research data
Physical Artefacts	5.43	16.28	19.38	27.91	31.01
Overall	7.54	22.3	18.03	24.26	27.87

Data Sharing Frequency

28.4 Over the last five years, what best describes your experience of data management plans on your projects? (As a percentage respondents)

	I haven't been involved in research that included a data management plan	In my experience data management plans are not followed	In my experience data management plans are followed in part	In my experience data management plans are mostly followed	In my experience data management plans are followed exactly
Physical Artefacts	30.23	6.2	18.6	40.31	4.65
Overall	32.13	3.61	19.67	38.03	6.56

Data Management Plan Experience
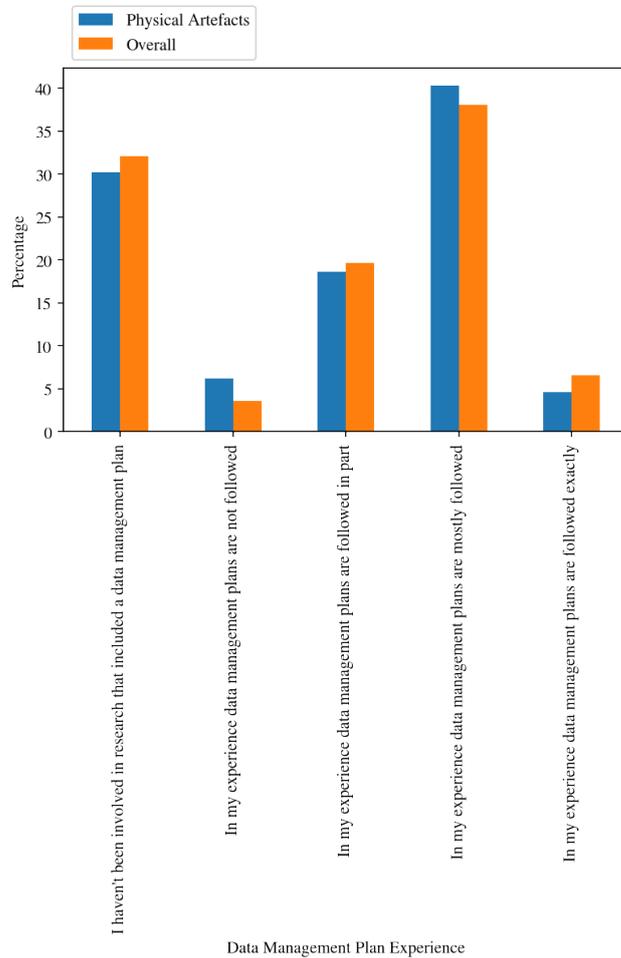
28.5 Data Sharing Methods - Over the last five years, which of the following practices have been part of your standard research process? (As a percentage of respondents)

	I license my data to allow it to be shared	I share my data with individuals/groups that have requested access	I create a unique identifier to make my data findable	I cite my research data in publications
Physical Artefacts	25.58	56.59	26.36	63.57
Overall	19.67	52.13	23.61	56.72

88
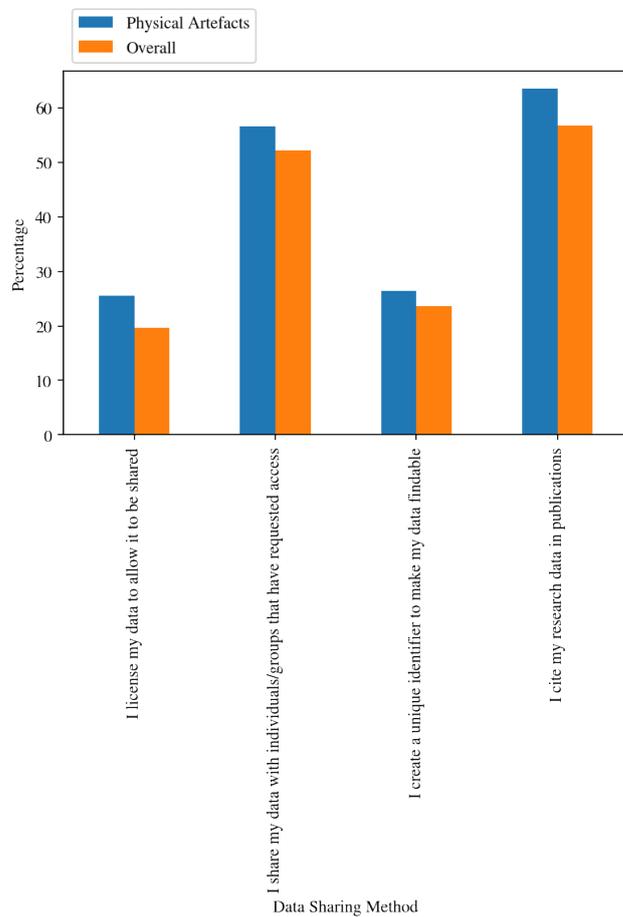
28.6   Are you confident that you understand the policies and responsibilities related to research data that are placed on you during a project/study? (As a percentage respondents)

	1 (Not at all confident)	2	3	4	5 (Very confident)
Physical Artefacts	2.59	4.31	23.28	48.28	21.55
Overall	1.84	6.99	26.1	44.85	20.22

89

Confidence in Data Responsibilities During Study
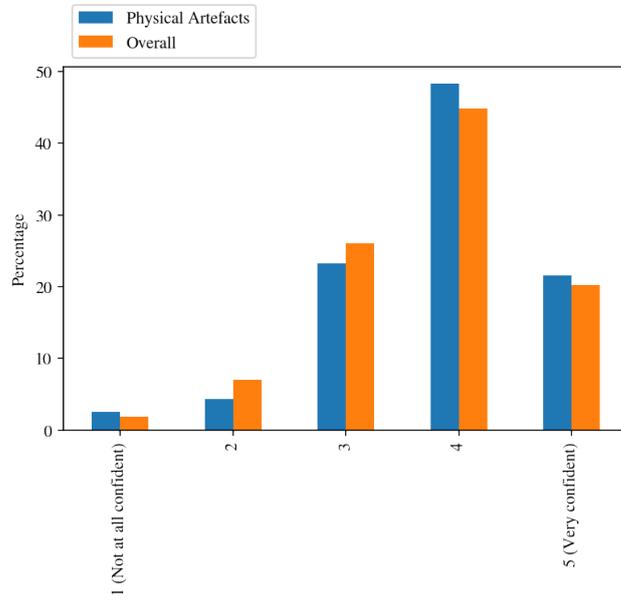
28.7 Are you confident that you understand the policies and responsibilities related to research data that are placed on you after a project/study has ended? (As a percentage of all respondents from the institution)

	1 (Not at all confident)	2	3	4	5 (Very confident)
Physical Artefacts	5.17	12.07	23.28	38.79	20.69
Overall	4.78	13.97	28.31	35.66	17.28

90

Confidence in Data Responsibilities After Study
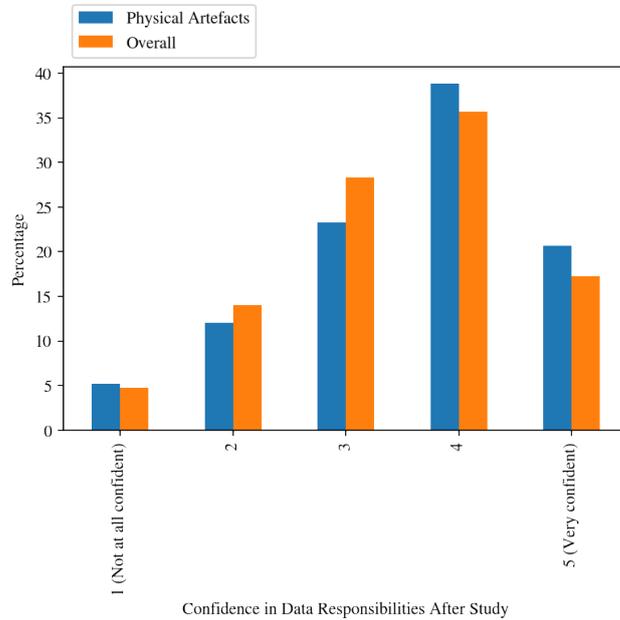
29 Are data-related practices different among respondents who work with numerical data? (N = 112)

29.1 Over the last five years - once you have completed a research project/study - which of the following options have you used to store the data that were generated? (As a percentage of respondents)

	On a laptop or desktop computer	In a shared area provided by yourself	In a shared area provided by your institution	In a data repository provided by your instution	In a data repository outside of your institution
Numerical	75.0	51.79	77.68	43.75	32.14
Overall	76.72	53.77	74.1	35.08	20.33

91
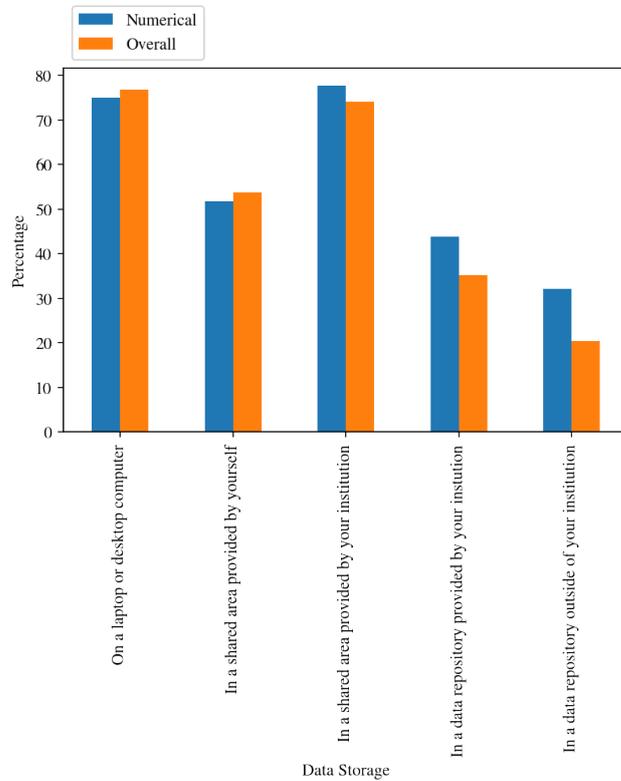
29.2 **Which of the following options describes your typical approach to storing the data that were generated? (As a percentage of respondents)**

	On a laptop or desktop computer	In a shared area provided by yourself	In a shared area provided by your institution	In a data repository provided by your institution	In a data repository outside of your institution	Not applicable
Numerical	16.07	16.07	43.75	8.93	11.61	3.57
Overall	26.23	17.38	38.03	7.87	5.9	4.59

92

Typical Method of Data Storage
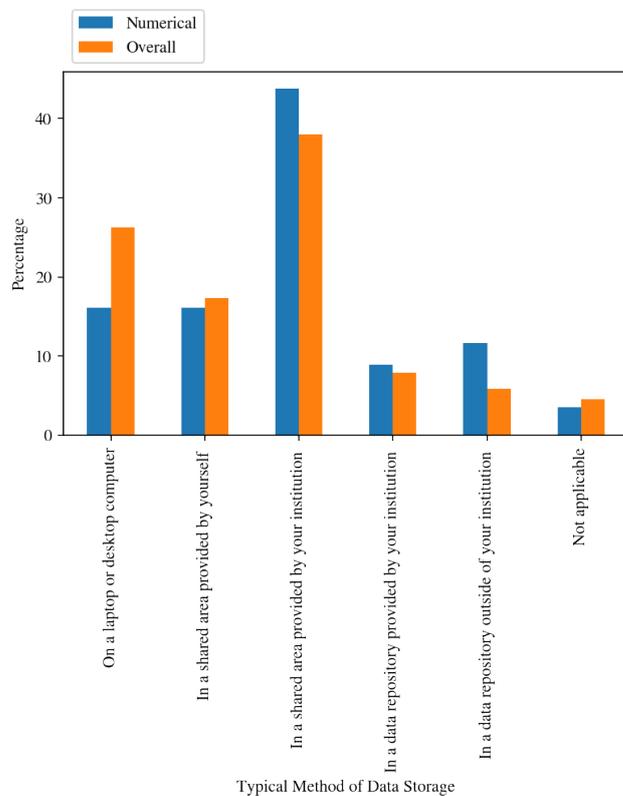
29.3 Data Sharing Frequency - Over the last five years, which of the following statements best applies to your research data? (As a percentage of respondents)

	I do not have research data to share	I haven't shared my research data	I have shared my research data but I rarely do so	I occasionally share my research data	I regularly share my research data
Numerical	0.89	16.07	17.86	29.46	35.71
Overall	7.54	22.3	18.03	24.26	27.87

93

Percentage vs Data Sharing Frequency

Legend: Numerical (blue), Overall (orange)

Categories:
- I do not have research data to share
- I haven't shared my research data
- I have shared my research data but I rarely do so
- I occasionally share my research data
- I regularly share my research data

29.4   Over the last five years, what best describes your experience of data management plans on your projects? (As a percentage respondents)

	I haven't been involved in research that included a data management plan	In my experience data management plans are not followed	In my experience data management plans are followed in part	In my experience data management plans are mostly followed	In my experience data management plans are followed exactly
Numerical	19.64	6.25	25.0	37.5	11.61
Overall	32.13	3.61	19.67	38.03	6.56

94

Data Management Plan Experience
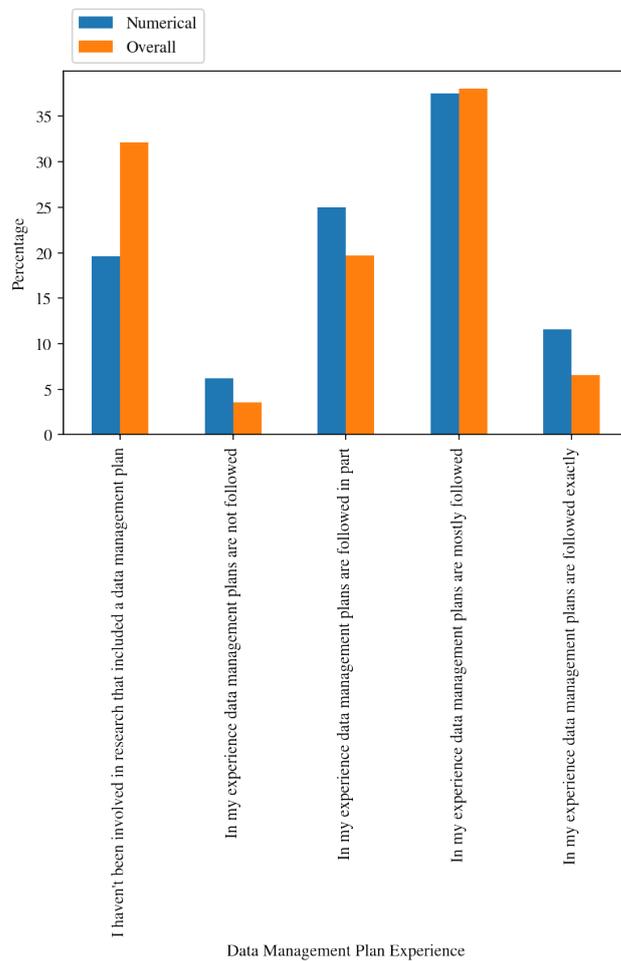
29.5 Data Sharing Methods - Over the last five years, which of the following practices have been part of your standard research process? (As a percentage of respondents)

	I license my data to allow it to be shared	I share my data with individuals/groups that have requested access	I create a unique identifier to make my data findable	I cite my research data in publications
Numerical	27.68	57.14	34.82	62.5
Overall	19.67	52.13	23.61	56.72

95
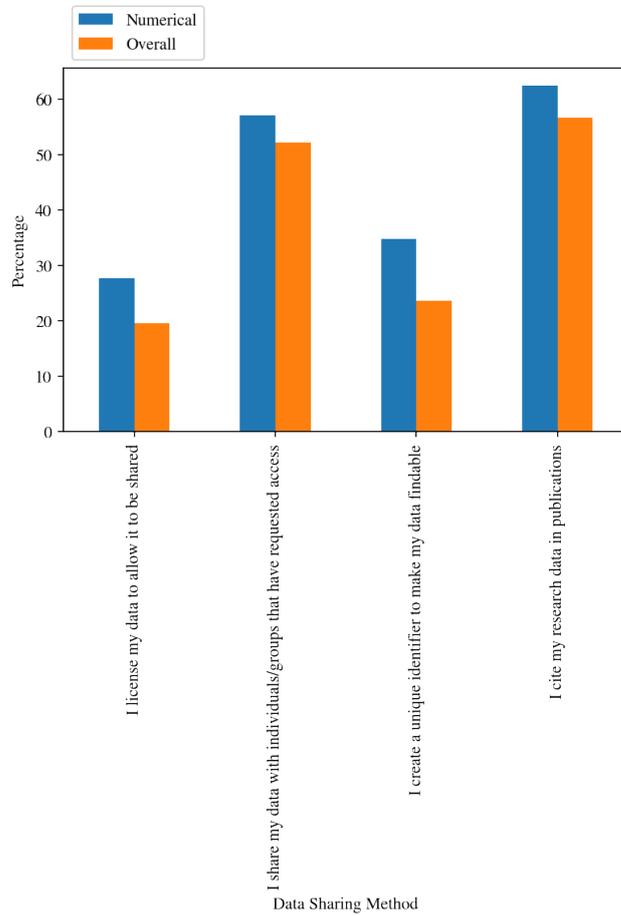
29.6   Are you confident that you understand the policies and responsibilities related to research data that are placed on you during a project/study? (As a percentage respondents)

	1 (Not at all confident)	2	3	4	5 (Very confident)
Numerical	0.0	4.67	28.04	42.99	24.3
Overall	1.84	6.99	26.1	44.85	20.22

96

Confidence in Data Responsibilities During Study
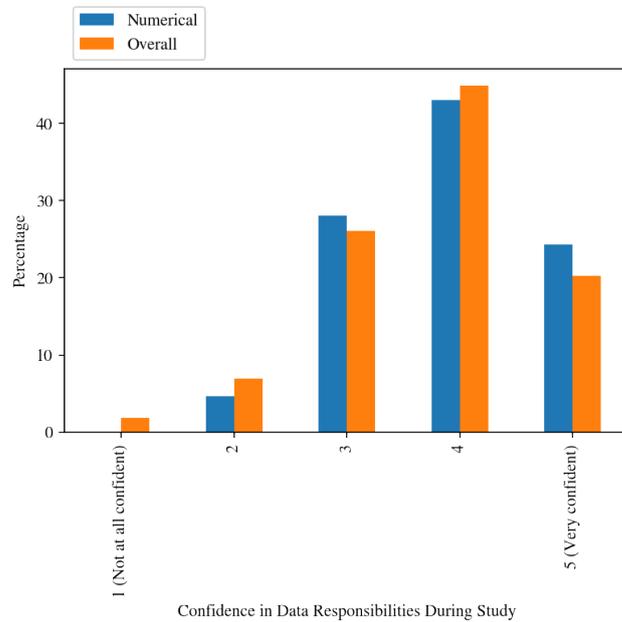
29.7 **Are you confident that you understand the policies and responsibilities related to research data that are placed on you after a project/study has ended? (As a percentage of all respondents from the institution)**

	1 (Not at all confident)	2	3	4	5 (Very confident)
Numerical	2.8	8.41	28.04	42.99	17.76
Overall	4.78	13.97	28.31	35.66	17.28

97

Confidence in Data Responsibilities After Study
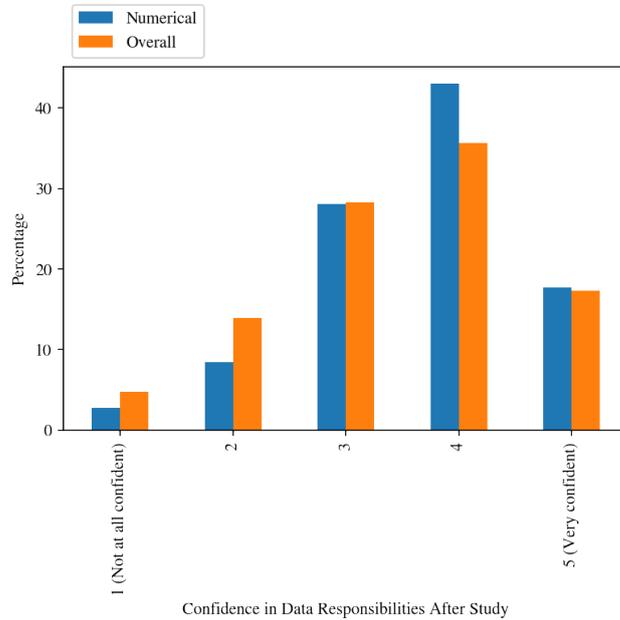
30   What are the disciplines and institutions of respondents who are involved in research software development?

Discipline	N	Percent
Digital Humanities	20	51.28
History	13	33.33
Information and Communication Technologies	9	23.08
Music and Visual Arts	8	20.51
Languages and Literature	8	20.51
Cultural and Museum Studies	6	15.38
Linguistics	6	15.38
Other	5	12.82
Library and Information Studies	4	10.26
Media	4	10.26
Divinity and Religion	3	7.69
Theology	3	7.69
Design	2	5.13
Archaeology	2	5.13
Drama and Theatre Studies	1	2.56
Classics	1	2.56
Dance	1	2.56
Philosophy	1	2.56
Development Studies	0	0.0
Law and Legal Studies	0	0.0
Political Science and International Studies	0	0.0

98

Percentage

Discipline

Digital Humanities
History
Information and Communication Technologies
Music and Visual Arts
Languages and Literature
Cultural and Museum Studies
Linguistics
Other
Library and Information Studies
Media
Divinity and Religion
Theology
Design
Archaeology
Drama and Theatre Studies
Classics
Dance
Philosophy
Development Studies
Law and Legal Studies
Political Science and International Studies
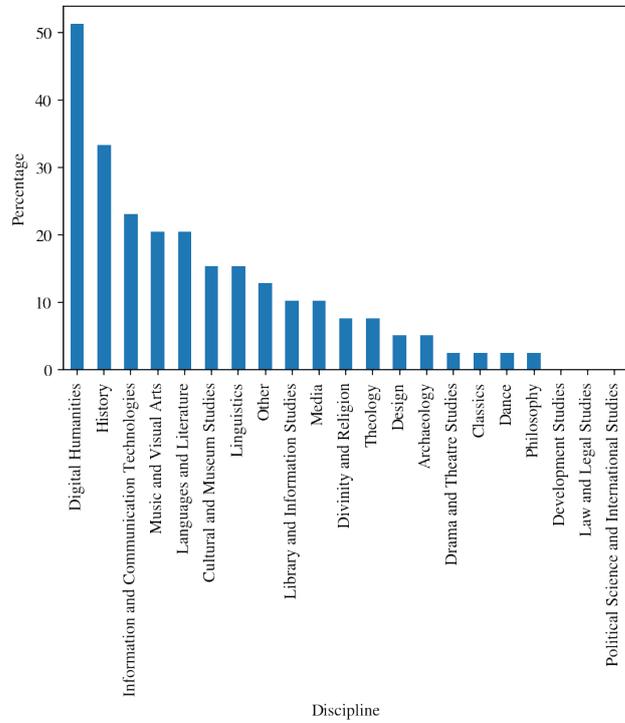
Institution of Respondents who Develop Software	N	Percent
University of Birmingham	4	10.53
University of Sussex	3	7.89
The Alan Turing Institute	2	5.26
Birmingham City University	2	5.26
University of Sheffield	2	5.26
University of Glasgow	2	5.26
The Open University	2	5.26
Durham University	1	2.63
King's College London	1	2.63
King's Digital Lab	1	2.63
Newcastle University	1	2.63
British Library	1	2.63
University of Kent	1	2.63
Aberystwyth University	1	2.63
University of Reading	1	2.63
University College London	1	2.63
University of York	1	2.63
University of Brighton	1	2.63
University of Nottingham	1	2.63
UAL Creative Computing Institute	1	2.63
Bournemouth University	1	2.63
University of St Andrews	1	2.63
University of the Arts London	1	2.63
Glasgow School of Art	1	2.63
The National Archives	1	2.63
Cardiff University	1	2.63
Imperial College London	1	2.63
Royal Central School of Speech and Drama	1	2.63

100
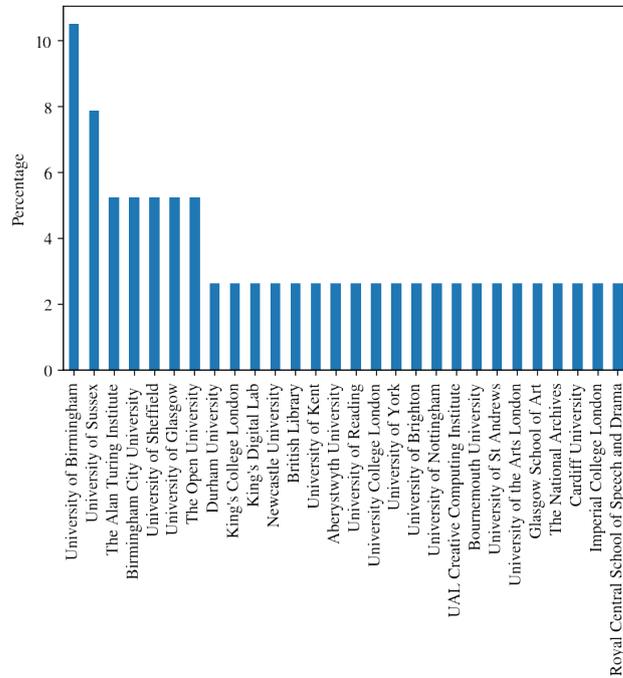
Institution of Respondents who Develop Software

31 What are the disciplines and institutions of respondents who typically store data on a laptop or desktop computer?

Discipline	N	Percent
Languages and Literature	23	28.75
History	22	27.5
Other	16	20.0
Music and Visual Arts	10	12.5
Archaeology	8	10.0
Design	8	10.0
Digital Humanities	7	8.75
Classics	7	8.75
Cultural and Museum Studies	7	8.75
Linguistics	5	6.25
Media	5	6.25
Philosophy	5	6.25
Divinity and Religion	3	3.75
Drama and Theatre Studies	3	3.75
Theology	2	2.5
Information and Communication Technologies	2	2.5
Dance	1	1.25
Law and Legal Studies	1	1.25
Library and Information Studies	1	1.25
Development Studies	0	0.0
Political Science and International Studies	0	0.0

101
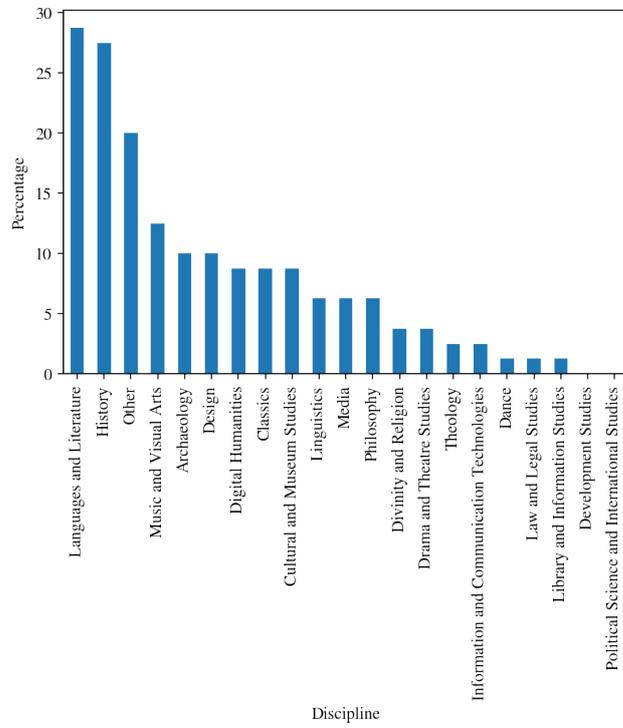
Institution of Respondents who Typically Store Data on a Laptop or Desktop	N	Percent
University of Oxford	7	8.75
University College London	4	5.0
Royal Holloway University of London	4	5.0
University of Liverpool	3	3.75
University of Glasgow	3	3.75
Loughborough University	2	2.5
Queen's University Belfast	2	2.5
University of York	2	2.5
University of Exeter	2	2.5
University of Manchester	2	2.5
University of the Arts London	2	2.5
Open University	2	2.5
Manchester Metropolitan University	2	2.5
The Open University	2	2.5
University of Cambridge	1	1.25
Victoria and Albert Museum	1	1.25
University of St Andrews	1	1.25
University of Bristol	1	1.25
University of Birmingham	1	1.25
Newcastle University	1	1.25
Northumbria University	1	1.25
Goldsmiths University of London	1	1.25
University of Aberdeen	1	1.25
University of Hertfordshire	1	1.25
University of the West of England	1	1.25
Ulster University	1	1.25
University of Warwick	1	1.25
Edinburgh Napier University	1	1.25
Royal Museums Greenwich	1	1.25
University of Leicester	1	1.25
University of Leeds	1	1.25
Univeristy of Leeds	1	1.25
Courtauld Institute of Art	1	1.25
Bath Spa University	1	1.25
Nottingham Trent University	1	1.25
Durham University	1	1.25
Imperial College London	1	1.25
University of Edinburgh	1	1.25
University of Liverpool	1	1.25
Lancaster University	1	1.25
University of Kent	1	1.25
University of Sheffield	1	1.25
Teesside University	1	1.25
De Montfort University	1	1.25
City University of London	1	1.25
University of Surrey	1	1.25
Northumbria University	1	1.25
Loughborough University London	1	1.25
The National Archives	1	1.25
Birkbeck College University of London	1	1.25
University of Lincoln	1	1.25
Liverpool John Moores University	1	1.25
Birmingham City University	1	1.25
University of Catania	1	1.25
University of Southampton	1	1.25

103

Institution of Respondents who Typically Store Data on a Laptop or Desktop
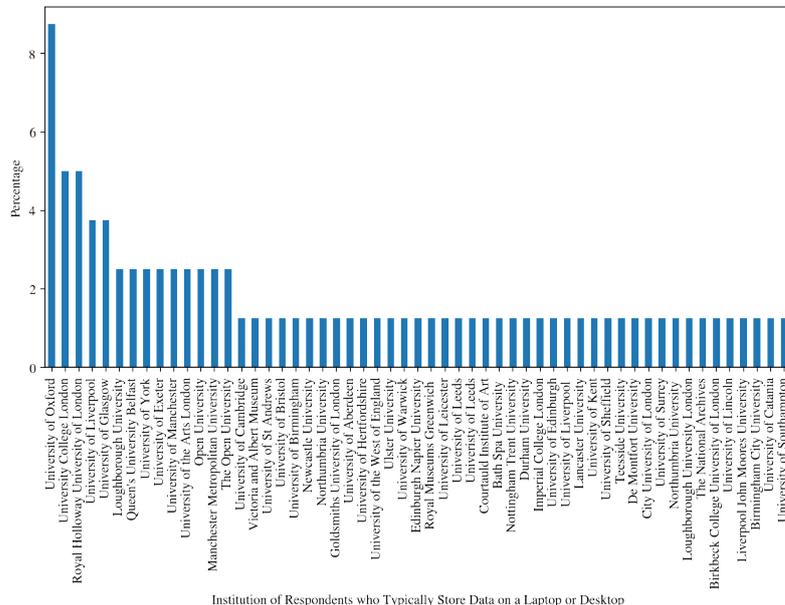
32 What are the disciplines and institutions of respondents who haven't been involved in research that included a data management plan?

Discipline	N	Percent
History	35	35.71
Languages and Literature	25	25.51
Other	20	20.41
Music and Visual Arts	10	10.2
Classics	8	8.16
Archaeology	8	8.16
Cultural and Museum Studies	7	7.14
Philosophy	7	7.14
Digital Humanities	5	5.1
Design	5	5.1
Library and Information Studies	5	5.1
Linguistics	4	4.08
Media	4	4.08
Drama and Theatre Studies	3	3.06
Information and Communication Technologies	3	3.06
Divinity and Religion	2	2.04
Law and Legal Studies	2	2.04
Dance	1	1.02
Development Studies	0	0.0
Political Science and International Studies	0	0.0
Theology	0	0.0

104

Percentage

Discipline

History
Languages and Literature
Other
Music and Visual Arts
Classics
Archaeology
Cultural and Museum Studies
Philosophy
Digital Humanities
Design
Library and Information Studies
Linguistics
Media
Drama and Theatre Studies
Information and Communication Technologies
Divinity and Religion
Law and Legal Studies
Dance
Development Studies
Political Science and International Studies
Theology

105

Institution of Respondents who don't use a Data Management Plan	N	Percent
University of Oxford	10	10.2
Birkbeck College University of London	6	6.12
Royal Holloway University of London	5	5.1
The Open University	5	5.1
University of Liverpool	4	4.08
University of Manchester	3	3.06
University of Cambridge	3	3.06
University of Edinburgh	3	3.06
University of Exeter	2	2.04
University of Liverpool	2	2.04
University of Glasgow	2	2.04
Durham University	2	2.04
University of York	2	2.04
University of St Andrews	2	2.04
University of Nottingham	2	2.04
Newcastle University	2	2.04
Loughborough University	2	2.04
University of Aberdeen	1	1.02
University of Antwerp	1	1.02
University of Kent	1	1.02
University of Westminster	1	1.02
University of Sussex	1	1.02
Bath Spa University	1	1.02
University of Birmingham	1	1.02
Liverpool University	1	1.02
Northumbria University	1	1.02
Univesity of Aberdeen	1	1.02
Glasgow Caledonian University	1	1.02
University of Cambridge	1	1.02
University of the West of England	1	1.02
University of Stirling	1	1.02
Edinburgh Napier University	1	1.02
Royal Museums Greenwich	1	1.02
De Montfort University	1	1.02
National Galleries of Scotland	1	1.02
Univeristy of Leeds	1	1.02
University of Warwick	1	1.02
University of Surrey	1	1.02
University College London	1	1.02
National Gallery	1	1.02
Lancaster University	1	1.02
National Museums Scotland	1	1.02
Teesside University	1	1.02
University of Leeds	1	1.02
University of Catania	1	1.02
Durham University	1	1.02
University of Leicester	1	1.02
School of Advanced Study, University of London	1	1.02
F1000	1	1.02
University of Leicester	1	1.02
Manchester Metropolitan University	1	1.02
City University of London	1	1.02
Open University	1	1.02
University of Strathclyde	1	1.02
University of Lincoln	1	1.02
Cardiff University	1	1.02
Birmingham City University	1	1.02
University of Worcester	1	1.02

106
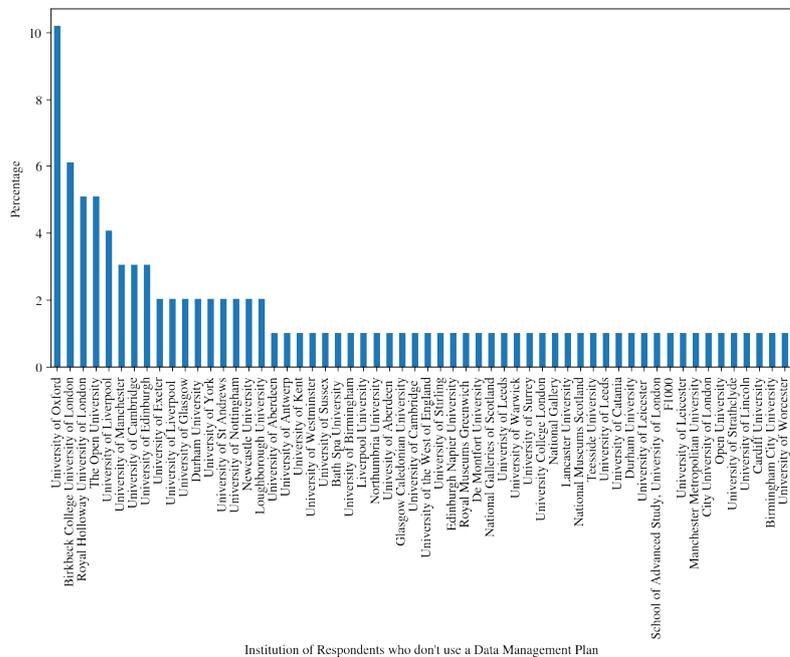
Institution of Respondents who don't use a Data Management Plan

33   **What is the spread of data skills learnt in the last five years - Have respondents who learnt one skill also learnt the others? (i.e. are the same people checking all of the boxes) Data skills included: data collection, analysis, management, and sharing.**

| Number of Data Skills Learnt	N	Percent
0.0	45	15.96
1.0	49	17.38
2.0	73	25.89
3.0	53	18.79
4.0	62	21.99

107
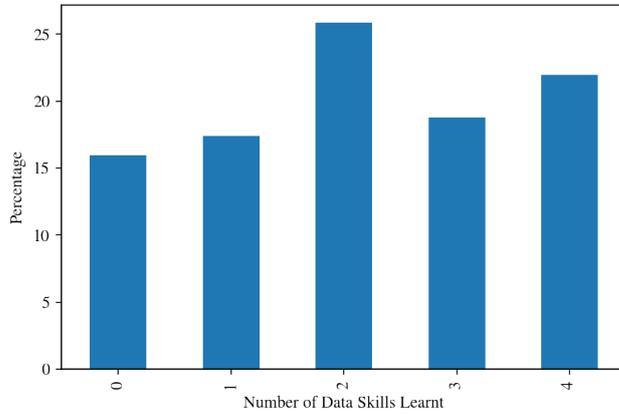
34   What is the spread of software skills learnt in the last five years - Have respondents who learnt one skill also learnt the others? (i.e. are the same people checking all of the boxes) Software skills included: software requirements collection, software development, and software sharing.

Number of Software Skills Learnt	N	Percent
0.0	165	58.51
1.0	69	24.47
2.0	34	12.06
3.0	14	4.96

108
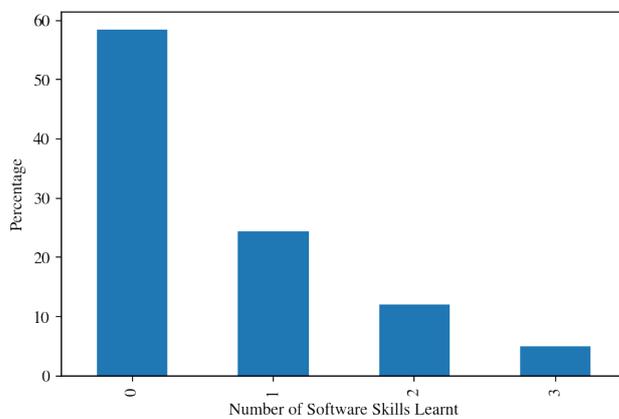
35   Do data-related practices differ among respondents who work in digital humanities (N = 54) and respondents who do not work in digital humanities (N = 280) ?

35.1   Over the last five years - once you have completed a research project/study - which of the following options have you used to store the data that were generated? (As a percentage of respondents)

|	On a laptop or desktop computer	In a shared area provided by yourself	In a shared area provided by your institution	In a data repository provided by your instution	In a data repository outside of your institution
Digital Humanities	73.08	55.77	88.46	53.85	34.62
Not Digital Humanities	77.47	53.36	71.15	31.23	17.39
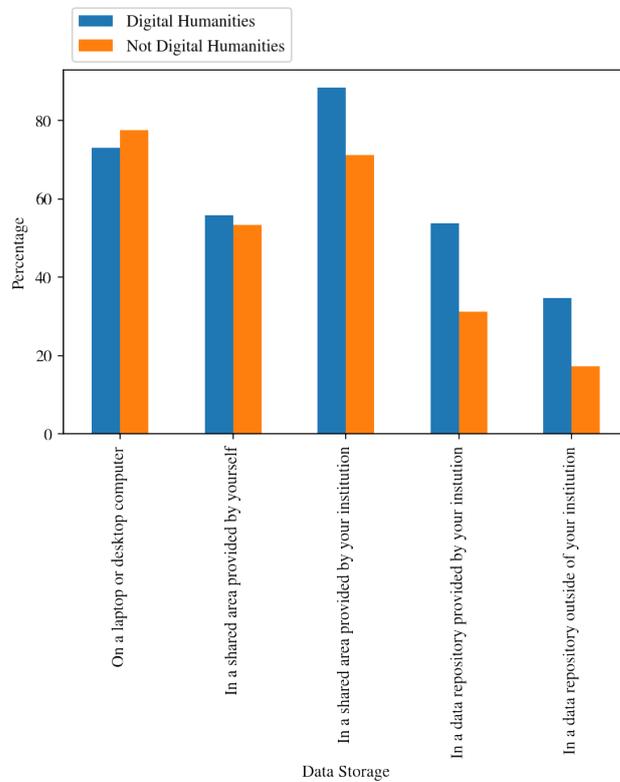
35.2 Which of the following options describes your typical approach to storing the data that were generated? (As a percentage of respondents)

	On a laptop or desktop computer	In a shared area provided by yourself	In a shared area provided by your institution	In a data repository provided by your institution	In a data repository outside of your institution	Not applicable
Digital Humanities	13.46	21.15	36.54	13.46	7.69	7.69
Not Digital Humanities	28.85	16.6	38.34	6.72	5.53	3.95
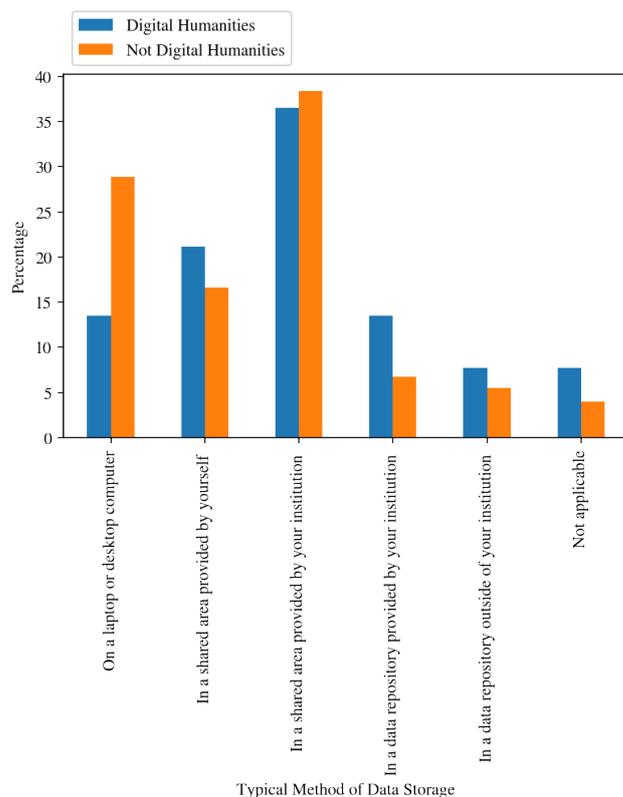
35.3 Data Sharing Frequency - Over the last five years, which of the following statements best applies to your research data? (As a percentage of respondents)

	I do not have research data to share	I haven't shared my research data	I have shared my research data but I rarely do so	I occasionally share my research data	I regularly share my research data
Digital Humanities	3.85	9.62	11.54	19.23	55.77
Not Digital Humanities	8.3	24.9	19.37	25.3	22.13

Digital Humanities
Not Digital Humanities

Percentage

Data Sharing Frequency

- I do not have research data to share
- I haven't shared my research data
- I have shared my research data but I rarely do so
- I occasionally share my research data
- I regularly share my research data

35.4 Over the last five years, what best describes your experience of data management plans on your projects? (As a percentage respondents)

	I haven't been involved in research that included a data management plan	In my experience data management plans are not followed	In my experience data management plans are followed in part	In my experience data management plans are mostly followed	In my experience data management plans are followed exactly
Digital Humanities	9.62	3.85	23.08	59.62	3.85
Not Digital Humanities	36.76	3.56	18.97	33.6	7.11

111

Data Management Plan Experience
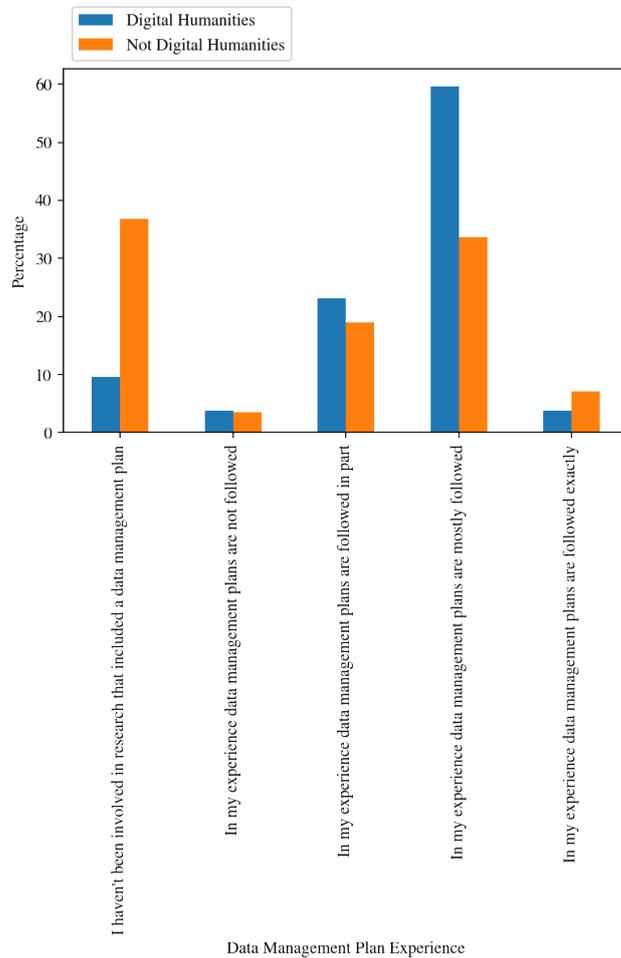
35.5 Data Sharing Methods - Over the last five years, which of the following practices have been part of your standard research process? (As a percentage of respondents)

	I license my data to allow it to be shared	I share my data with individuals/groups that have requested access	I create a unique identifier to make my data findable	I cite my research data in publications
Digital Humanities	48.08	63.46	51.92	67.31
Not Digital Humanities	13.83	49.8	17.79	54.55

112
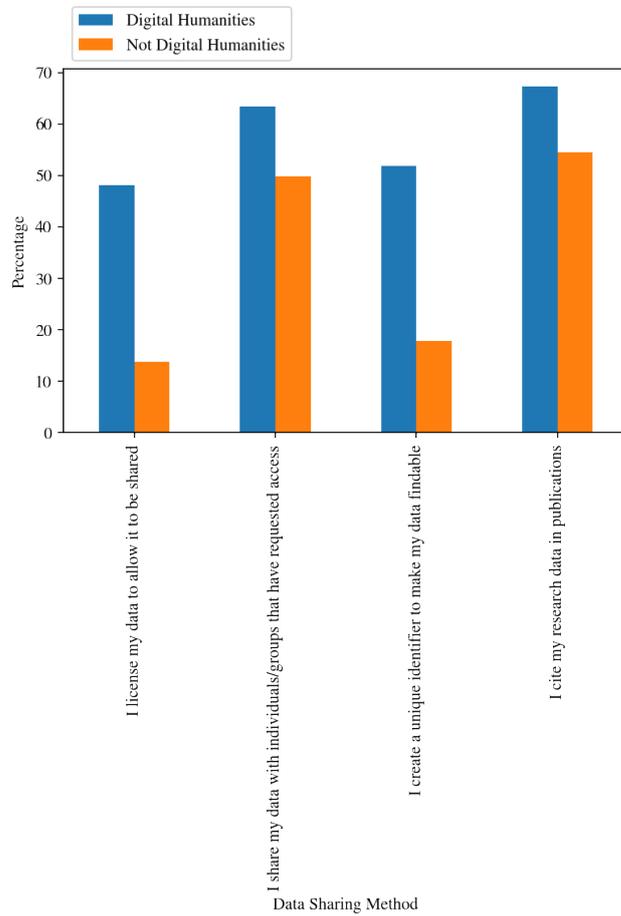
35.6 Are you confident that you understand the policies and responsibilities related to research data that are placed on you during a project/study? (As a percentage respondents)

	1 (Not at all confident)	2	3	4	5 (Very confident)
Digital Humanities	0.0	2.17	15.22	54.35	28.26
Not Digital Humanities	2.21	7.96	28.32	42.92	18.58
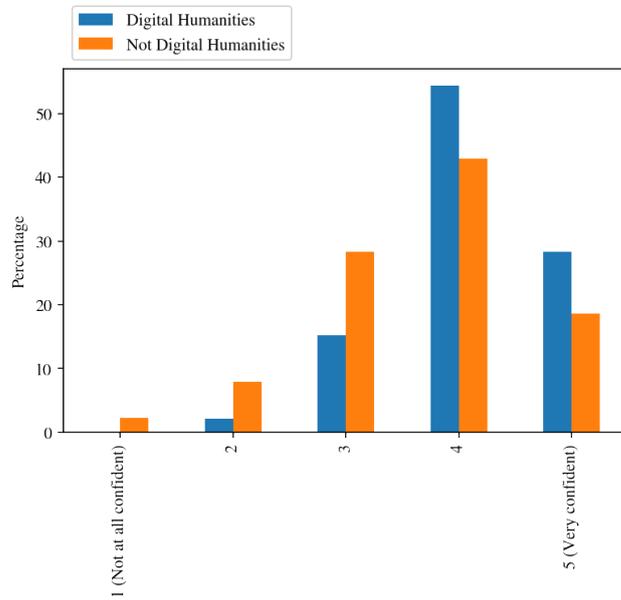
Confidence in Data Responsibilities During Study

35.7 **Are you confident that you understand the policies and responsibilities related to research data that are placed on you after a project/study has ended? (As a percentage of all respondents from the institution)**

	1 (Not at all confident)	2	3	4	5 (Very confident)
Digital Humanities	0.0	11.11	20.0	48.89	20.0
Not Digital Humanities	5.73	14.54	29.96	33.04	16.74

114

Confidence in Data Responsibilities After Study
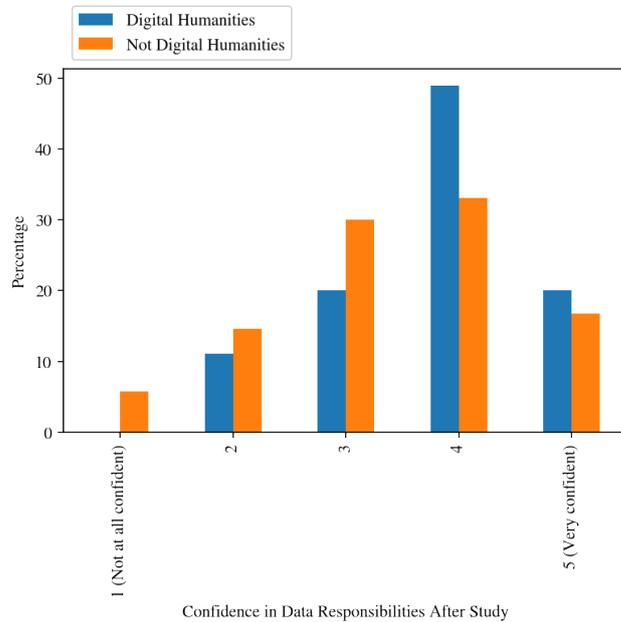
36 Other Responses

36.1 Other Disciplines

- Architecture
- Textiles and Fashion
- Sociology
- Architecture
- Human Computer Interaction
- conservation of paintings
- Chemistry
- Geography
- Health
- library and information science
- Creative writing studies
- Business and Management (International Business)
- Film Studies
- Critical Neuroscience
- architecture
- Art History
- Conservation Science

115

- Geography
- cultural geography
- Cultural Digital Heritage
- Indology, Somatic Practices
- Anthropology/ Ethnomusicology
- Education
- Electronic engineering
- Heritage Studies
- Conservation
- heritage science
- Creative Writing
- Geography
- Computer Science
- Statistics
- Archival science
- History of Art and Architecture
- geography
- History of Art
- Historical and Cultural Geography
- Art History
- Area studies
- Economics
- Human geography
- Geography
- Art History and Visual Culture
- anthropology
- architecture
- Cultural policy; creative industries; arts management
- History of Science
- Cartography
- Architecture; Archives; Architectural history
- Global mental health and bioethics
- interface of art history of science
- Cross-disciplinary - social science of music
- Architecture
- Architectural and Health Humanities
- Social Anthropology

- Heritage
- Human Geography
- Creative Writing
- Economics/Innovation/Policy
- Bioarchaeology
- Physics, Biology
- Conservation
- Business management studies
- Folklore
- Translation Studies

36.2   Other Organisations Applied to for Funding

- Erasmus+
- CHASE
- Jisc
- European Commission
- Royal Society
- CBA Yorkshire
- HEA
- Norwegian Research Council ; Australian Research Council
- NESTA
- RIBA
- Paul Mellon Centre
- Northern Bridge Consortium
- Arts Council of Wales
- Midlands4cities
- None
- Carnegie Trust
- National Endowment for the Humanities (US)
- Software Sustainability Institute
- Royal Society
- JISC
- The Textile Society, Heritage Crafts Association.
- NIHR
- JISC
- Midlands4cities
- My visa status precludes me from eligibility to apply
- Carnegie Trust

- RSE
- NIHR
- Carnegie Trust for Universities of Scotland
- DFG
- Heritage Consortium
- Paul Hamlyb Foundation; Calouste Gulbenkian Foundation; Leeds City Council; National Theatre; Leeds 2023; DCMS
- Paul Mellon
- NIHR, NIMH
- Nuffield Foundation
- Templeton Foundation
- Royal Society
- BFI
- Society of dyers and colourists
- UKRI

36.3   Other Types of Research Data

- Technical drawings analytical drawings
- 3d models
- Code
- Spectroscopic data
- remotely sensed satellite imagery
- Industrial textiles
- Brain Data
- Computer-aided design
- Live performance and workshops
- interviews, ethnographic notes
- singing and voice
- spectroscopic
- Live performance
- VR
- 3D Data
- 3D digital models
- Survey data (both quantitative and qualitative)
- Maps
- 3D laster scan
- mixtures of the above, location
- computer imaging
- transactional data, survey data, data derived from software APIs etc
- social media data
- Spatial

36.4 Other Methods of Data Storage

- Server
- Independent hosting environment
- Github
- Portable Hard Drive
- Personal memory stick
- Europeana
- Harddrive
- Github
- Not yet completed
- hard disk
- Still in research phase of project
- public archive
- in another institution where research grant is held
- USB memory stick
- RAID drives
- n/a (have not completed project in last 5 years)
- Physical data (not digital) in secure cabinets in my office.
- External harddrive
- Hard drive
- not completed
- flash drives
- GitHub
- portable drive as backup
- mainly physical hard drives
- Zotero
- Not yet completed research
- External hard drive

36.5 Other Methods of Obtaining Research Software

- I have developed open source research platform software for the community
- none of the above
- I have purchased specialist research software at my own expense
- I have purchased research software with my own money
- I write scripts using standard bash commands. What is this research software of which you speak?
- I have purchased specialist research software with my personal funds
- Paid myself for research software
- I received training to use specialist research software to supplement manual methods, but I ran out of time to use it
- Manual

36.6 Other Methods of Storing Research Software

- server
- Web-based

36.7 Other Methods of Acquiring Skills and Knowledge about Data and/or Software

- Na
- Na
- not applicable
- I haven't learnt any skills
- Previous computer science degree and professional work
- None of the above
- I have not done such training.
- Professional Software Engineer
- N/A but I teach myself - so long since I was a student
- Not learned these skills
- don't have any knowledge
- Not applicable

36.8 Other Barriers to Learning Skills

- N A
- too much focus on technical skills rather than methodological skills
- Lack of time to engage
- time constraints post pandemic
- No-one really takes ownership of this task
- I'm aware of training/learning opportunities, but I have *very limited* time to engage with them
- A mixture of the two options above
- No time to dedicate to this
- There are so many different opportunities and different methods it's confusing
- People keep asking me to fill in forms
- Lack of time due to overwhelming commitments
- It is difficult to apply and improve research skills that are not embedded in my current research practices
- Time
- I have searched for tips on Google
- no
- As far as I have found this information doesn't exist in one place.
- I have not made the time to engage with them
- I would like a helpdesk for when i get stuck but there is no such thing other than websites
- I think this is a matter of having funding available even moreso than having skills

- I haven't really investigated this though

- don't know

- I do make use of opportunities to develop these skills but there is always more to learn and sometimes there isn't enough time to do everything I would like to do.

- Not interested - I am happy to share data but no interest in learning skills to do it. I would rather stick pins in my eyeballs

37 Skills To Improve - Open Text Responses

37.1 Thinking about your current skill set, what is the main data or software skill that you would like to obtain to improve your research - All responses

- More skilled students and PDRAs so I don't have to do everything myself!

- Qualitative analysis of interview and video-based data

- continue to use present skills

- Online research tools (e.g. Twitter analysis)

- relational databases

- unsure

- High-level project management skills (PRINCE etc)

- I design for layouts

- Recording, transcribing, archiving, website scraping

- basic statistical analysis

- not applicable

- database creation

- Social media data scraping

- Excel and R Studio

- None

- None

- Social media scraping

- Data scraping and modelling

- Statistical analysis

- More regular practice with Python so that my laptop is better set up for using it

- humanities research methods

- image analysis

- Software development

- Greater data visualisation skills

- Improve use of qualitative data analysis tools

- digitization of material and long term plans and infrastructre for data managment and sharing

- Software sharing

- Database/archiving software

- Data collection

- Web design, Digital Humanities Text Encoding Initiative skills

- historical mapping

- I would like to see better institutional support and understanding of what we do and our data needs

- Packing software into libraries with standard testing procedures

- nothing currently

- Software development

- Efficient data cleaning/preparation

- Coding

- data storage

- Data analysis (eg using functions in a spreadsheet, use of R Studio

- We all use and think about data differently in our practice. As a practitioner, data is embedded in the work I and collaborators do. Recently, it has been led by data specifically. No software was used to analyse data but could be in the future. Software skill training in this regard would be helpful.

- Data management

- Participatory action research data analysis from multiple digital and non-digital sources. I have done much of my analysis by hand (reading, writing, reflexive journaling, coding and thematic mapping) because I have not found a single digital tool that will read diverse forms of data collection.

- Data collection

- Learning R, data scrapping, AI-based text analysis and prediction, AI-based archival data usage

- python and R

- Interpretative skills working with lidar and GIS

- I would like to improve more my skills on Data sharing and Data management/storage

- none

- Digital methodologies

- Data analytics and visualisation

- linguistic analysis

- Data analytics

- OBS Studio or similar screen casting tools

- I don't know

- graphical representation of data and historical sources

- Statistic software/Database software

- Greater familiarity with creative commons licensing

- Data sharing

- Software creation

- programming

- Data collection, such as web scraping

- 3-d modelling, better GIS

- Nvivo data management skills

122

- Database design and the web hosting of databases

- Using R package

- digital data collection, using R and Gephi, regression analysis in SPSS

- I'm not sure

- Matlab

- Improve my Python skills

- NVivo advanced training

- I can't think of anything I specifically want to obtain, in terms of skills.

- Software Engineering

- My research is practice led so I am not sure how to answer this question

- Unsure

- Use of research software to support data analysis

- More modern tools such as R

- Would like to be able to visualise/present the data more effectively

- Nvivo

- Effective data sharing - have explored this in principle but not yet in practice

- Database design and management e.g. SQL

- None

- none

- Qualitative text data analysis (e.g. NVivo)

- Data storage

- Qualitative content Analysis

- Im not sure

- None. My team do all this sort of work

- GIS

- Web design

- use of software. I do not feel that I urgently need it to collect and analyse my data efficiently. I'm considering the prospect solely because in job adverts, software knowledge is usually considered and favoured.

- Coding through R or Python

- Data storage; data sharing; network visualisation software

- AI application

- coding

- I think I am about to retire

- Data analysis

- digital mapping

- Frankly, having a better understanding of whether my research is generating a dataset and what I might do with it; otherwise these issues feel completely divorced from my research.

- I don't know what systems I am missing, so an introduction to new software available would help.
- With the phasing out of Zetoc, I need to learn new ways to conduct ongoing
- Literature reviews to keep up with developments in my field.
- Reference management
- Excel
- data analysis
- SQL, HTML
- Licensing
- N/a
- Programming Skills so that I can create new software or tweak existing software to better suit my situation.
- Too many things to list.
- Devops - there's no point in complex software if its complex to deploy
- javascript for visualization
- Have time to learn CIDOC-CRM which our new knowledge system is designed with
- data management and sharing on external repositories
- ArcGIS; R.
- Use of referencing/citation software
- understanding copyright rules better
- data scraping and rules about how to manage GDPR when it comes to audio and video
- network analysis
- Quantitative research for humanities
- Advanced statistics
- Data sharing tools
- GUI development
- Long-term preservation/sustainability of research outputs (data, user interface etc.)
- competence in R
- I need more time
- Database creation
- NVIVO
- Coding
- more knowledge about available software and their properties
- R
- Computational modelling in R and Python
- Combining multiple data sets; metrics, photographs, text- into a single catalogue of artifacts.
- None
- I would like to learn to use transcription services that help me process the documents I'm working with.
- I would like to improve GIS skills/knowledge that I already have

124

- shrugs

- I'd like to learn to build better databases, sql, filemaker, something that would allow me to build and query custom databases.

- Python

- Docker

- I would like to learn more about new ways of data visualisation.

- Better competence in use of Tableau

- Confidence in sharing DOI to better usable/accessible raw data (rather than architecture describing it)

- more understanding of data sharing software and proceedures.

- spreadsheet usage

- A basic course in besy practice for PhD researchers

- NVivo

- any software that helps me be 'metaverse' ready

- Data mining, web scraping, natural language processing, video mining

- Revive my quant skills.

- Photography editing (photoshop) comes to mind. Generally, I feel that beyond 3D modelling I do not know enough about the possibilities that might be relevant to my field of study. So it would be great to find out more about this.

- Would like to learn more software know-how, as I barely used it but feel it would be great for job hunt outside of academia

- Sharing data mutually with other researchers

- Skills related to software writing packages that will help when I get to the writing-up thesis stage

- Data visualisation

- Qualitative data gathering and analysis

- Visualization software for sharing my data

- I can't think of one - I also don't like to think of my research materials as data

- My research doesn't rely on specialist data software

- Better manipulation of PDFs

- Data analysis and visualisation

- Dealing with simple softwares (notion etc) currently

- Korsakow programming

- Improved skills in data analytics

- Training on how to make raw data useful for others outside my project and how to navigate permissions when data is from external repositories.

- Better approaches to linking and sharing data

- TEI XML encoding, photogrammetry, LOD, IIIF

- text analysis software

- I think I need a refresh/ introduction in all areas, but mainly on use of images and videos

- What data management software is out there for me (as a literary historian) as I generally haven't come across anything that I would actually use.

- how to better collect data and tools to analyze the data

- better Nvivo skills

- I am still at the stage of collecting data, when I start to analyse it in detail maybe using a programme like NVivo would be useful

- statistics

- data sharing

- Improve XML / TEI coding skills

- Software sharing (e.g. licensing, repositories, distribution, deployment)

- Web data collection

- using repositories for data and code on e.g. GitHub

- meeting needs of longer term storage

- Data collection - I want to be able to do more systematic social media research but have NO idea what tools are available to work on Facebook or Twitter. I am restricted to free platforms or platforms my (cash strapped) university subscribes to, and cannot afford personal subscriptions

- Full usage of GIS software

- Data collection

- machine learning for coding video data

- Correspondence analysis

- How to use automated analysis of large amount of data.

- AI to support qualitative data analysis such as sentiment analysis

- bibliography management

- Licensing; archiving

- More comprehensive understanding of NVivo

- none

- Using data analysis software

- How to access data from other projects

- Database and GIS

- To find and learn how to use effective transcription software

- data collection and data sharing

- text mining software and corpus linguistics methods

- database programming and management, back end, and front end presentation

- I don't know!

- learning new data analysis tools

- I would like to learn R and python

- Easily accessible information about options for achiving and sharing data.

- How to collect and share data effectively

- Software development

- not sure

- Coding
- Ability to analyse textual sources semantically using a computer more effectively
- Software to analysis digital audio recordings
- Use of Sketch engine to mark up texts
- Digital methods on social media
- Data sharing
- N
- learning how to use sharepoint
- Greater GIS skills
- Quantitative analysis
- Better understanding of the functionalities of Excel
- QGIS
- Extending my data analysis skills to cover Python and R
- Simple ways to meet GDPR regulations
- I would like to learn to write my own software for network analysis of data.
- I don't need data/software skills for my research
- Na
- I will at some point need to use archaeological sources, so I will both need to learn skills in obtaining data (some way of trawling through studies and databases) and create an efficient way of storing the data.
- TEI-XML
- Data collection: web scraping, corpus analysis, et
- Making online video
- more clarity on diffrent approaches and possiblities
- API use (I've previously worked with a postdoc on this but I don't have any funding at the moment, so ideally I'd be able to do it myself)
- Urgh. R and Python. But I'm too busy and tired. No time.
- Data sharing
- Film making and oral history collection
- none
- analysis
- Web scraping
- Data comparison
- Better use of R studio
- enhanced textual analysis techniques
- Unsure
- operating computer from the terminal, coding, integration of codes in different programming languages in one software
- Unsure

- competence in R
- Python
- Analytics/text mining
- programming/ text analysis
- Using new tools
- More effectively supporting research software I've made, which is used by others, so that it is maintained beyond the lifetime of the grant(s) that led to its development
- Using software (e.g. NVivo) for data analysis
- I feel I have the appropriate skillsets for my next research projects; though I would like to explore tools that help with image recognition/analysis - e.g. identifying very similar image compositions etc.
- Visual analysis tools
- would prefer not to need any. I like reading texts!
- Understanding how to carry out robust research using online participant pools.
- coding
- Using APIs
- AI
- m
- Data analysis
- GIS
- Analysis and manipulation of numerical datasets
- Python coding
- Referencing software eg zotero
- Using R
- Data analysis (statistics, visualisation)
- Digital theme-searching and web-scraping
- None - I have no interest in learning to do any of this. It's a job for a technician
- I have had access to everything I have needed
- I have everything I need
- Don't know
- Industry and commercial research data
- Data science software (R, Python, Matlab), data warehousing and exporting (SQL, PowerBi, Azure, AWS), data presentation and production tools (websites, Docker), code sharing and collaboration (Git, Github)
- Curating online exhibitions/websites
- zetell kasten
- Better ability to share work with the public - websites etc.
- NVIVO
- Using qualitative data analysis software
- Use of software to image data
- GIS mapping, participatory animation,
- n/a
- I'd like to be a better coder, but I'm not sure when I'll be able to take the time out to do it.

38 Are there any further skills that you would like to obtain to improve your research? - All responses

- digital humanities
- modelling
- no time to even think about the skills I dont have.
- Yes, there are 3d data acquisition tools which I would like to be more fluent with
- Coding (Python); sharing data (e.g. making open access)
- Data analysis
- Open Access data protocols
- coding: neural-net data analysis: understanding proprietary data-set formats
- Learn more languages (e.g. C++, Java, python)
- Desk top publishing (text/imaging) On line publication. Networking and collaboration..web based or via social media.
- Big data analysis
- Access to data analysis sites for designers.
- No. I have everything else I need.
- Intellectual property (of the software developed, as well as of the data collected and obtained during research)
- Technology Roadmapping
- Language skills
- Large data management
- creating DOIs; licensing and open access; repositories; data management after leaving university
- Statistical analysis
- Database design; web scraping; GIS skills
- scientific articles writing, Latex
- Statistics
- Researcher etiquette when working with human participants.
- theory building in qualitative research
- None
- Commercialisation; dissemination; data sharing
- time management to make research time
- Open Access
- data visualization
- A better understanding of the data management software available and when it is appropriate to use it.
- linguistic analysis
- Further Data Analysis
- Too many things to list.
- Community building around tools, software, data

- data management tools

- linked data practices

- Learning to access and manage excel spreadsheets; analysis of videos

- use of jupyter notebooks; sprql querying; coding

- no

- Best practice with research repositories - there are so many options - which one(s) to use??

- Anything to do with digital humanities would be of interest to me at this point.

- shrugs

- automated Javascript testing; alternatives to Selenium for functional testing

- Update data management skills since PhD qualification

- better understanding of ways of using software to compile and present data

- website creation

- Photography editing (photoshop)

- writing skills- layout and presentation of thesis work , how to insert images and create maps and my own diagrams.

- More Python programming (what we learn already is so great!)

- No, I just need time

- interview technique

- Languages

- social media - as an area to research and how best to engage with it

- data analysis; data representation

- Data collection techniques, including web scraping

- No

- I would like to be competent using Excel to manage projects

- 3-D cad modelling of buildings

- Better understanding of GIS

- No

- Statistical analysis software

- Better understanding of web-building; checking I properly understand GDPR and safe data-sharing; understanding better how to deposit data in public repositories.

- practical ethics training; practical compliance training;

- German

- Topic modelling; machine learning

- More practical things, ethos to edit film or sound

- CAD design

- Data management

- software development;python;APIs

- Better Medieval Latin and paleography; familiarity with Protein Data Bank analytical tools

- No

- Software

- How to negotiate with managers and IT depts who are resistant to change/technology and see it as a threat (security or otherwise).

- Ability to develop and synthesize abstract concepts

- GIS mapping, participatory animation

APPENDIX 7: CODING FRAME

Interview Analysis Coding Frame

ATTRIBUTES

- Career stage 1-4 + PRO

DATA

- Their research - current and history
- Understanding of term data
- Data processes (collection, analysis sharing, promotion)
- Data management plans

SOFTWARE

- Software use and tool/Proprietary/bespoke/open source software
- Software expertise (team, institution, personal, RSEs)

SKILLS (data and software)

- Skills gained how (community of practice, informal networks, collaboration, training)
- Barriers to skills acquisition – stuff that gets gets in the way....including of future skills training)

CURRENT SUPPORT

- INSTITUTIONS Data and software institutional support (and lack of it)
- AHRC

FUTURE SCENARIOS (AHRC policy interventions)

DIVERSITY

MULTIDISCIPLINARITY/COLLABORATION

COMMUNITIES OF PRACTICE