technopolis group

February 2022

Scoping the skills needs in the social sciences to support data-driven research

Final report

Version 2.0

Scoping the skills needs in the social sciences to support data-driven research

Dr Jo Ferrie, Dr Martin Wain, Dr Simon Gallacher, Dr Eleanor Brown, Rebecca Allinson, Dr Peter Kolarz, Prof. John MacInnes, Laura Sutinen, Rita Cimatti

February 2022

Table of Contents

Figures

Executive Summary

This study, commissioned by the Economic and Social Research Council (ESRC), aims to help identify the skills needs of UK social science researchers to carry-out data-driven research and fully exploit the increasing volume of large and complex data available for research purposes. Data-driven research is defined inclusively, covering all forms of data production and/or data analysis that informs and/or is used by social science researchers from archives to zines, inclusive of qualitative and quantitative approaches. Data-driven research produces a robust and rigorous evidence-base that can be opened to scrutiny, and that the researcher is confident in defending. Data-driven research has sufficient rigour to inform decisions within academia and beyond.

The study focusses on doctoral training and will be complemented by broader work, led by the ESRC, which considers skills needs at later career stages. It assesses core and advanced skills needs among social scientists, gaps in current provision, the structural barriers impacting the teaching or learning of data-driven skills and interventions and changes required in doctoral training, to address them. Our methodology used five phases of inquiry. A scoping literature and landscape review have delivered a foundation for primary data collection. This report is informed by 38 interviews with researchers at different career stages, a survey to Doctoral Training Partnership (DTPs) training leads and sense-checking workshops involving 14 'experts' in data-driven skills and/or data-driven skill teaching plus 11 ESRC-funded doctoral candidates.

Core training is defined as the teaching provided at Master's level and which is a pre-requisite for ESRC funded doctoral candidates. Advanced training is defined in two ways: specialised training which advances a researcher's capabilities towards being an expert user; and training that researchers' need at an advanced stage in their doctoral research projects. 'Gaps' are identified where there is: i) a lack of access to training in a particular skill, method or field, and ii) lack of knowledge to evaluate what skill, method, field or provider would best fill the gap identified.

The findings of this research are organised into two stages: a stage 1 'needs analysis', and stage 2 'doctoral intervention analysis'. For each stage, we present a summary of key findings, followed by a presentation of ideas on how to address relevant structural factors and barriers to data-driven skills learning (i.e. to 'close the gap').

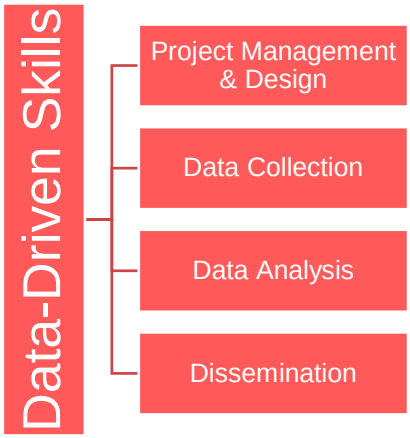
Stage 1: Needs analysis. What are the core skills needed by social scientists and what advanced training is needed and at what scale?

The championing of core broad-based training in the ESRC's 2015 Postgraduate Training and Development Guidelines has influenced practice across the sector and is impacting positively on postgraduate and early career scholars. The core requirements are extensive, and to train students in the skills, less time has been made available to engage with a deeper understanding of ontological and epistemological frameworks. Drawing from the interviews, workshops, and a survey to Doctoral Training Partnerships, it is evident that gaps exist driven by the separation

of skills training from conceptual knowledge and understanding. Participants in all phases of our inquiry reported the absence of a 'whole-project' approach to data-driven skills training that would help students understand better, the decisions (and consequences of decisions) they were making throughout their research training.

Analysis of all five phases of inquiry revealed four 'pillars' of learning that collectively produce confident data-driven researchers (see Figure 1).

Figure 1 The four 'pillars' of training required to produce skilled and confident data-driven researchers



The evidence informing this conceptualisation is presented throughout the data chapters (2, 3, and 4). While data collection and analysis currently dominate core learning in the UK, project management and design, and dissemination skills are less evident, and this is due to their absence in the core requirements in the ESRC Guidelines.

The Guidelines were published in 2015. In the last seven years learning from Q-Step,[1] a surge in literatures around pedagogy of training, perhaps triggered by the Teaching Excellence Framework (TEF) and a pandemic that has accelerated our use of digital spaces, permit scrutiny of what should be included as core requirements, and how core learning can be optimally delivered. In this context, and drawing from all five phases of our inquiry, there are skills needs that are not currently included in the Guidelines.

- **Project management and design, including data management** are fundamental data- driven skills. This includes managing resources, including time, and should include open access skills such as cleaning data files, fixing data inconsistencies, performing data enhancement, generating data in multiple file formats for preservation and dissemination and generating code lists. Allied to this, rules around intellectual property, safeguarding self, safeguarding data, and research integrity should be included as core training. Our research, drawing predominantly from the literature review, indicates these skills are most valued when a candidate can also demonstrate team-working, communication skills, strong knowledge of ethics and legal expertise

[1] See: https://www.nuffieldfoundation.org/students-teachers/q-step

- **Digital Skills.** This research identified gaps relating to digital skills in all four 'pillars' of data- driven knowledge production. There is no clear definition of digital skills found in the literature, and participants who contributed to this research noted the fluid and contested use of this term. Digital skills discussed here fall broadly into two camps: softer skills which reflect how we work, and communicate (from emails, to lecture capture, to dissemination strategies that use twitter); and research skills that align with methods/methodology (building apps, web-scraping, machine learning and big data analysis). Broadly, doctoral candidates appear to have an appetite for digital training regardless of the form of data they are creating/encountering. This is causing a capacity issue and doctoral candidates and recognised researchers may be best placed to deliver training in this area. Digital skills are valued globally and across all academic disciplines. There is a gap in the Guidelines around digital research design, and specifically the approaches to data collection that relate to data captured from digital sources including data ownership, curation and storage, and digital data analysis. The UK is at a tipping point where embedding such digital skills into doctoral training will be an asset for our UK-based scholars, and not doing this will cause a deficit in the UK's data-driven skill set within a few years. Unless this data-driven skill gap is closed, there may be an increasing divide within UK social sciences including UK doctoral candidates and those training overseas. Further, there is evidence that the digital divide exists across all research council domains, and in many competing nations

- **Dissemination skills** not currently covered by the Guidelines, and needed, include dissemination strategies and new pathways to publication including using digital platforms

As doctoral candidates move into their PhD, they seek advanced training. From the interviews and workshops, doctoral candidates and early-career scholars reported a lack of confidence in their data-driven skill needs, often starting their research without adequate specialised training. Advanced and specialised training provided through ESRC investments are highly rated but around half of researchers at all career stages were not aware of opportunities. Finding suitable specialised training was difficult, exacerbated by the PhD experience often being isolating and alienating from their peers and society. Doctoral candidates in the interviews and workshops reported the absence of peer networks and supervisors that were aware of specialised training opportunities and who could offer advice. Further, doctoral candidates experienced enormous pressure to appear as expert, particularly in data-driven skills as these were valued by employers.

Our research shows that many identified skills gaps are as valuable for non-academic career paths as academic career paths. Skills are further enhanced by collaborating with non- academic partners. Industry, civil and statutory sectors value working with doctoral candidates and provide vital learning spaces to hone argumentative writing approaches, write reports, and provide team working opportunities. These networks also have the benefit of growing during people's careers, providing mentors, guidance, and future research collaborators.

Stage 2: Doctoral Interventions. What interventions are needed at a doctoral level? What is the scale of change needed in the curriculum? More specifically, what is the gap between our current postgraduate research and development guidelines and the core training identified as being needed? Do we need to shift how we structure doctoral education to produce confident users of data-driven skills?

While ESRC can and should update its Guidelines to incorporate the gaps we have identified in core data driven skills, our research show that there are structural issues relevant to teaching and learning data-driven skills and a culture shift is required to deliver the scale of change necessary. These are summarised below:

- Core methods courses designed to deliver the Guidelines are often delivered at scale (100- 400), taught often by small teams of early career educators, and are designed to accommodate students with no prior learning. Examples were often given of Master's methods courses being copies of undergraduate courses, sometimes with the same teaching teams. Students include a large range of disciplines making it difficult to produce materials that 'connected' and for students with prior learning the content and pace can be frustrating
- Core methods course are spread over the course of a semester. Consulted doctoral candidates felt that being 'helicoptered' into methods learning once a week, over the course of a semester, hindered learning, and this group favoured condensed courses. A condensed presentation was also championed by trainers involved in interviews and workshops, as improving the learning experience when teaching data-driven skills. Further supporting evidence is available in the literature
- Core qualitative and quantitative training is delivered through separate courses. Despite the Guidelines not structuring requirements in quantitative and qualitative silos, 'core' training was delivered through separate courses, with each focusing on 'skill learning' in their assignments. As a result, learners struggled to conceptually understand the broader research design landscape, rather seeing quantitative and qualitative approaches being in opposition to each other. Conceptual learning means understanding the different forms of data, the context in which they were produced and the options available for their collection, collation, and curation. Data literacy is about knowing enough to know that there are multiple pathways towards knowledge production and why some pathways are less legitimate than others for a particular project or research question. While the 2015 Guidelines were designed to provide literacy, this is a gap that persists, and the separation of quantitative and qualitative approaches is part of the problem
- Adding digital skills onto a framework that separates quantitative and qualitative approaches will not sufficiently capture the specific skill gaps identified. Digital project management and dissemination strategies, digital data capture and storage and digital skills including new platforms (e.g. python) and new approaches (e.g. AI) require a learning experience that can examine all data forms. See section 3.3.1
- Anxiety about data-driven skills learning remains high and is exacerbated by structural issues such as large class sizes[2] and by the lack of specialised

[2] Characterised as usually 100-400 students, while disciplinary learning tended to range from 10-50 students

learning opportunities. Recent literature on pedagogy and data-driven skills with learning from the NCRM and Q-Step centres present strong evidence for personalising the learning experience and making class sizes commensurate with disciplinary classes. One potential pedagogical solution to improve 'connection' to the learning, is to create a 'flipped class'. Here the formal lecture or knowledge delivery mechanism is moved outside of the class session, and the 'homework' or application of knowledge is moved in

- Assignments have a skills-focus rather than a conceptual-focus and participants from all career stages reported that challenge-led learning activities and assignments would help students learn research design, project management and dissemination skills alongside data analysis skills
- The Master's dissertation was seen as an extension to disciplinary, rather than research methods training. Doctoral candidates in the workshop felt positioning it as a pilot to doctoral research would help them transition from Master's to doctoral work
- Quality methods teaching requires a more complex knowledge base than traditional lectures require, and universities are not acknowledging this in workloads, rewards, or recognition (i.e., promotion). Courses that host large numbers of students, and that use smaller teaching teams are creating profit for schools. From the interviews, workshops and literature, there is evidence that large classes that teach core data-driven skills learning tend to get poorer student feedback, which can lead to disinvestment from HEI managers

A culture change will require long-term investment using fees 'earned' by courses to support course teams to redevelop the learning experience in a sustainable manner. A major transition such as this may need front-loaded funding from the ESRC to trigger change. Providing teaching teams with time to invest in literature-informed development of learning materials is recommended to embed masterclasses, interactive activities, use of real-world data and challenge-led tasks into core courses. Strategies that foster team-working amongst students and post-course support were championed by doctoral candidates.

Change is also required in relation to advanced and specialised training. For the most part, the specialised training required to produce leading and confident data-driven skill researchers exists, but some shifts are required to make it visible and accessible. These are summarised below:

- Promotion of advanced and specialised training provided by ESRC investments could evolve to improve access. Cost was a major barrier to advanced data-driven skills training: cost of the event and cost of travel were cited. ESRC students can use their Research Development Support Grant, although they often have to wait to be reimbursed
- Supervisors need to be universally supportive of advanced training and also the role of placements in developing students' skills. Some doctoral candidates reported being discouraged from attending data-driven skills training. If supervisors endorse training, students are very likely to attend. ESRC-funded supervisors should be alert to opportunities for their students and mandatory supervisory training with a social science focus would close this gap. Around half of the doctoral candidates that contributed to this research did not get any data-driven skills training from their supervisor(s)

- There are structural issues that are limiting doctoral candidate's access to data-driven skill training. If advanced and specialised training is to be encouraged at universities, a sector- wide culture change is needed to value academic time being invested in training materials. A revisit to the Research Excellence Framework's component on the Research Environment could trigger a culture shift in this area by requiring Higher Education Institutions (HEIs) to report progress. The framework could also evolve to recognise broader publications beyond peer-reviewed articles including open data and learning materials

Intervention recommended: Re-imagining learning to ensure PhD graduates have the data- driven skills they need

Our recommendations propose a transition to new 'core' training and require collaboration between HEIs, Doctoral Training Partnerships (DTPs) and other ESRC investments such as the UK Data Archive, Centres for Doctoral Training (CDTs) and the National Centre for Research Methods (NCRM). The changes proposed will require a culture shift and re-engagement with data-driven skills training for many HEIs. New courses should remove unhelpful repetition, move to class sizes small enough to allow connections to be built between educators and doctoral candidates and this may mean the return to courses taught to disciplinary groups, rather than the pan-social science approach that exists in many HEIs. To move beyond the siloing of quantitative and qualitative methods, new courses should deliver a mixed-methods presentation that is intellectually demanding and focused on conceptual understanding. A mixed methods approach was championed by most participants and at all career stages. By removing repetition, the proposed courses have space for doctoral candidates to begin specialised training within their Master's year and this should improve engagement with core training, as well as reduce anxiety about their research challenges as students move into their doctoral research. To underpin these changes, the ESRC's Guidelines should make it clear that a broad-based approach to core training relates specifically to conceptual knowledge and data literacy. Skill development may be incorporated, but it is not the primary aim of the broad- based approach. The Guidelines should incorporate elements from the General Research and Transferable Skills Training (ESRC, 2015: 12-15) into their Expectations for Core Research Methods Training (ESRC, 2015: 7-10) using the four-pillar model (Figure 1). Re-framing the General Research and Transferable Skills Training as a Doctoral Education Framework will underpin the need for continued professional development during the doctoral years. In fact, a culture shift where data-driven skill learning is understood as career-long is needed.
Changes proposed to the Guidelines are understood as recommendations that will impact on the delivery of data-driven skills training for ESRC-funded students. To deliver a broader step change that includes all social science researchers, care will need to be taken by HEIs to extend opportunities at the implementation stage. We list six key recommendations below:

1. Core learning should begin at a more advanced level in the Master's. Training Needs Analysis (TNA), standardised across ESRC investments, should be used regularly and systematically to identify prior undergraduate and work-based learning, and require those without the required skills, to complete a foundation course in data-driven skills (condensed presentation taking

potentially of 1-2 weeks). The foundation course is skill-focused and aims to capture what is currently taught to undergraduate social science students in the UK. Skills include learning to use software for example, or data production approaches such as interviewing. Figure 2 features the proposed content of the course:

Figure 2 Proposed content of the foundation course

Foundation Course	Introduction to ethics
	Introduction to ontology & epistemology
	Levels of measurement & sampling
	Univariate & bivariate analysis with numeric data
	Data production including survey design, interviews, participant observation & focus groups
	Using documents as data
	Thematic analysis with discursive and narrative data
	Fundamentals of data literacy

2. A new core course should be developed featuring a condensed and mixed-methods presentation (potentially 3-4 weeks). Class sizes should be limited to 50 students, and this may mean moving students into disciplinary groups to complete their core training. Care should be taken to critically engage with substantive material and disciplinary ontologies, and this becomes easier where a smaller number of disciplines are represented in the class. Investment in teaching teams is needed to develop new and sustainable programmes informed by pedagogical learnings from Q Step and other initiatives. Respondents in all of the sense-checking workshops highlighted that uncritically appointing the same educators practiced in teaching siloed quantitative and qualitative materials risks perpetuating divisions rather than presenting the broad spectrum of data-driven skills and approaches to knowledge production. A culture shift of early career scholars with digital skills teaching established scholars may be required.
   This new course should be intellectually demanding at a commensurate level to other Master's courses. This is achieved by learning outcomes being less skill-focused, rather capturing reflexive and critical engagement with the strengths and limitations of chosen pathways to knowledge production. To support students in this learning, the core course should incorporate research design and project management learning, including dissemination strategies. Adding in these elements will help doctoral candidates make the most use of their Master's dissertation as a pilot for their doctoral research, giving them more time to practice their data-driven skills. Figure 3 outlines the potential content for this course.

It is encouraged that doctoral candidates complete a challenge-led assignment producing a portfolio of outputs that better capture a whole-project approach including design, project management data collection, analysis, and dissemination. Using real-world data is advised. From the literature review and participants through all phases of inquiry, using data that has a social justice foundation and/or data that has been produced locally enhances engagement

Figure 3 Proposed content of Core Training

Core Training - Emphasis on Conceptual Learning	Greater understanding of how epistemological choices impact on findings produced
	Recognise importance of data/project management and dissemination including research integrity, and demonstrating rigour as applied to traditional and digital approaches
	Awareness of and understanding in AI and machine learning as a research tool
	Promote using existing data such as surveys, existing repositories of visual and text data, social media and big data
Core Training - In Terms of Data-Driven Skills:	Appreciation of theoretical frameworks underpinning data and analysis
	Curation and analysis of numeric, text and visual data with appreciation for using archives and secondary data
	Application of data ethics inclusive of digital approaches, reflexivity and standpoint
	Fundamentals of coding - if not covered in foundation course
	Analysis approaches such as discourse analysis, phenomenology, and critical realism
	Maintenance of emotional well-being and the importance of and strategies towards safeguarding self
	Development of skills in digital data capture, curation, and storage, safeguarding digital data including awareness of rules around intellectual property
	Open data access issues and data cleaning skills, fixing data inconsistencies, generating data in multiple file formats for preservation and dissemination
	Building of dissemination and impact pathways
	Further practice of a wide range of analytical sofware (R, Nvivo, etc.)

3. Investment in active learning strategies, as opposed to more traditional didactic techniques such as lectures, is recommended. Examples of active learning exercises include computer demonstrations and simulations of statistical tests and concepts, using films to synthesise research findings and practice dissemination, critically appraising numeric outputs, reflective research diaries, practising analysing sample datasets in software packages and using podcasts of interviews with academics
4. To encourage access to advanced learning, a new specialised training module, drawing on learning available across ESRC investments should be linked explicitly to dissertation training. This will raise awareness of training

and opportunities to network with peers. This is a bold innovation and will take careful consideration for successful implementation.

The use of a foundation course creates a 20-credit gap in the Master's curriculum and it is proposed that students use this to build their own portfolio of training, drawing on four (delivering 16-32 hours of taught materials) courses that map directly to their dissertation and doctoral research plans. The courses could be delivered by their HEI, by their DTP or by other recognised educators including ESRC investments such as National Centre for Research Methods (NCRM) and the UK Data Archive. Students complete an assignment that draws together their learning from their specialised training into a single submission to be assessed by their HEI, and it is anticipated that their dissertation supervisor will be involved in grading this work. The submission could be an essay, an extension to the portfolio created for their Core Training module, or a refined research proposal.

ESRC funded doctoral candidates should work with their PhD supervisor, dissertation supervisor (if different) and DTP training specialist in developing their portfolio. The portfolio should include at least one advanced method or data-driven skill and at least one transferable or research development skill. The courses could be closely aligned, for example allowing a student to complete an introductory and then an intermediary course in the same data-driven skill

5. The TNA conducted at the point that studentship awards are decided should include an indication of what specialised training would usefully be completed before the PhD. For 1+3 students this is conducted before the Master's year begins. This provides the time required to allow DTP training leads and professional development leads to start planning their training programmes. It is critical that students are encouraged to revisit their choices at the end of the core training. Therefore, the specialised courses should be completed post-winter break assuming an autumn start to the Master's to give space for students to change their mind, and space for training providers to meet demand. For this to work optimally, all DTPs should use an agreed TNA template. The specialised training module is part of a strategy to increase awareness of the training available from ESRC investments that can support doctoral research

6. Building on current NCRM practice, all training provided by ESRC investments, and extending to UKRI investments, should share a typology that describes training level (basic/entry, intermediate, or expert/advanced). This typology should extend to cover two further categories: presentation and suitability. A regularly updated, curated and searchable source of information about training and development at ESRC and/or UKRI investments should be developed and promoted

1. Introduction

1.1 This study

This study, commissioned by the Economic and Social Research Council (ESRC), aims to identify the skill gaps that need to be addressed in order to prepare social science researchers in the UK to create and carry-out data-driven research. The study delivers recommendations for how these gaps can be addressed, specifically through postgraduate learning, and also with some focus on undergraduate and post-doctoral learning experiences.

The rationale for the study is to ensure that social scientists have the skills and capacity to fully exploit the increasing volume of large and complex data available for research purposes. Administrative and secondary data available for quantitative and qualitative analysis are increasingly open and accessible to doctoral students. Equally, advances in machine learning and artificial intelligence provide new opportunities to harness the big data produced by global corporations and social media. Understanding the implications of this fourth industrial revolution for social science research is vital and doctoral candidates will require skills their supervisors did not need. Social science research in the UK is increasingly understood within global initiatives, such as the influence of the UN's sustainable development goals as a tool for understanding 'grand challenges'.[3] There has been a significant increase in the scholarship of learning and teaching (SoTL) available to training educators and leaders, to consider afresh how data-driven skills are learned. This project aims to understand this wider context, to understand what data-driven skills are required by leaders in social science research. The aim is to capture the step-change required to provide world-leading learning opportunities for doctoral candidates with an eye to career progression. So, where possible, the study takes a life-course perspective, examining data-driven skills needs and provision of training and capacity development at different career stages. This learning is then used to better understand the changes that could happen at the doctoral level to produce successful career trajectories.

The term data-driven research is understood inclusively, covering all forms of data production and/or data analysis that informs and/or is used by social science researchers from archives to zines. Our definition is inclusive of mixed-methods, qualitative and quantitative approaches. There is more than a question of what is included, with considerations also of purpose and standard. Data-driven research produces a robust and rigorous evidence-base that could be opened[4] to scrutiny and that the researcher is confident defending. Further, data-driven research has sufficient rigour to inform decisions within academia and beyond.

Core training is defined as the mandatory training that ESRC funded doctoral candidates complete during their masters. At times we discuss the need for 'advanced' training. Advanced can be understood in two ways: the incremental approach to training where an advanced level requires completion of introductory

[3] See: https://sdgs.un.org/goals

[4] There are situations where research should not be open to scrutiny. The use of this term denotes a standard set, where rigour could be demonstrated.

and intermediate learning towards specialised knowledge; and advanced in the sense that the project has advanced, and new or different skills are required.
This research is timely. This academic year, 2021/22 marks 10-years of Doctoral Training Partnerships (DTPs, formerly Centres or DTCs), the ESRC framework for supporting social science doctoral candidates across the UK. This report is published shortly after the major Review of the PhD (Tazzyman et al, 2021) has made significant recommendations to transform doctoral education. Our work overlaps with their project and provides an opportunity to further contribute usefully to the positive development of social science postgraduate study and social science research.

The chapter will now outline the wider context of this work, by describing the training landscape.

1.2 Research Context

This section presents an overview of the PhD training landscape in the UK for social scientists and provides a description of ESRC funding opportunities for potential candidates. This context is presented prior to the conceptual and methodological underpinnings of this work.

The UK has been considered a global leader, delivering the highest standard of data-driven skills training for doctoral students (Royal Society, 2019:6), producing graduates 'valued for their depth of knowledge, critical thinking and research skills' (Tazzyman et al, 2021: 6) though the British Academy warned that the UK could not be 'complacent about its reputation as a world leader' (The British Academy, 2013: 10). Indeed, the British Academy questions whether the UK is in fact a world leader (The British Academy, 2016) and what gaps exist between our UK approach to data-driven skills and overseas educators.

The ESRC supports around 700 studentships through their DTP network annually and is considered 'the 'flagship' funder of UK social science PhDs' (Tazzyman et al, 2021: 16). These studentships are thought to represent around 20% of PhD funding available in the UK for social scientists (Tazzyman et al, 2021). Each DTP has a 7-year training grant that supports students who normally have an honours degree classification of 2:1 or higher. The student must have support from university supervisors, who are also often assessed as part of the studentship competition. From the landscape review, there are various forms of award, and some are student-led (student writes the proposal, supported by a supervisor); others are supervisor-led (supervisors write the proposal and then recruit a student); or collaborative (supervisors and non-academic partners co-write the proposal and then recruit a student). Further there are steer awards available in some DTPs to close identified gaps in data-driven skills. The three steers available are in advanced quantitative methods, datasets, and interdisciplinary awards. Before 2020, students needed to be based in the UK or EU for a full award, and now 30% can be based overseas (50% in Economics) prior to study. The DTPs aim to produce "highly qualified and rounded social scientists, equipped to meet contemporary economic and social challenges … bridging the gap from research to practice" (DTP website – Landscape Review).

Centres for Doctoral Training (CDTs) focus on a particular thematic area, and between 2017/19 funded around 30 PhD studentships. Two CDTs are listed on the ESRC

website. CDTs provide a rich environment to work in an emergent field. For example, the Data Analytics and Society CDT is based across the Universities of Leeds, Liverpool, Manchester, and Sheffield and operates between the social sciences and computing, mathematics and the natural sciences and strategically brings external partners including industry links together with students. CDTs are often linked with a DTP.

Currently a typical ESRC studentship lasts three years and covers fees and a stipend. This funding is available to students with a master's qualification and is known as a +3. The recent Review of the PhD (Tazzyman et al, 2021) has recommended extending this to fund 4-years of doctoral study to manage the additional pressures on post-graduate students in contemporary academia including managing visibility, publications, networking, and impact.

Currently core training is provided within a master's qualification provided by HEIs that make up the partners of the DTP and should be complete before the PhD research begins. If a student is on an ESRC PhD award, this is framed as a '1+' indicating an additional year of fees and stipend to support an additional year of master's study. Therefore, a studentship award that includes master's funding is known as a '1+3'. Applicants that can demonstrate completion (or near completion) of a masters that covers the training requirements in full can be recruited straight into their doctoral research and this is known as a '+3' award. Via the DTPs, the ESRC 'approves' masters courses that can demonstrate subject/discipline-rich learning and contain sufficient methods training to deliver the 2015 Guidelines in full. The Guidelines provide guidance on the skills that doctoral candidates should develop over the course of their studies. The guidelines stipulate that Doctoral Training Partnership (DTP) training, delivered with university partners, should address 'research skills', 'research methods', and 'broader capabilities.' The Guidelines were published shortly before and guided the transformation of ESRC's investment in doctoral education, from Doctoral Training Centres to Partnerships.

DTPs can make use of 'fractional awards': if a student has completed a masters with dissertation but not covered sufficient methods to meet the Guidelines, they can be awarded a small extension. For example, a +3.5 award would provide a 6-month extension to the PhD deadline and 6 months of additional fees and stipend to permit the student to complete core methods learning. In DTPs where fractional awards are not used, students with data-driven skills gaps are funded to complete a 1+3 even if they have already completed a masters with dissertation.

This context has informed the conceptual framework and methodology developed.

1.3 Conceptual framework

The Guidelines have been used as an anchor in this research. The anchor worked in three ways: Firstly it accepts the verdict from established and leading researchers contributing to this research that the Guidelines are considered to be the gold standard in core social research methods learning in the UK (supported below); secondly it gives us a framework to consider, critically whether the content does in fact deliver the data-driven skills required by social scientists and finally, it shows how this framework should evolve to inform learning experiences that can best

support students to close the gaps in their data and evidence building skills. As will be discussed throughout the chapter, the aim here is not to provide a one size fits all approach as this becomes a static version of best practice that soon becomes de-contextualised and stagnant. Instead, we outline the evidence around what works, and what principles may be applied to developing learning experiences for social science researchers.

The aim of the work was to develop an evidence base that supported recommendations to extensions or redirections of ESRC investments in doctoral education. Specifically, the study aims to set out how the ESRC can ensure that UK social science doctoral graduates have the skills required to undertake high-quality, impactful data-driven research. The study will inform the development of an ESRC capacity building strategy to support and facilitate data-driven social science research.

Conceptually, we wanted to deliver an evidence-base that captured the field of practice around data-driven skills as a primary aim (see Appendix C for definitions of key terms). By drawing on UK practices and global approaches we offer a relational analysis that allows the strengths and limitations of UK learning structures to appear. This then permits consideration of what skills gap exist and require attention. Through data collection and analysis, the concept of a skills gap has extended to include considerations of enablers (contextual or structural factors that help reduce the gap) and barriers (contextual or structural factors that contribute to the gap). Furthermore, we have evidence about what works and have added this to the framework as a useful steer for recommendations.

To move to a conceptual contribution that could 'close the gap', we returned to experts using workshops of doctoral candidates and data-driven skills trainers to challenge our emerging 'propositions'. Building up layers of evidence has allowed us to corroborate findings drawing on multiple phases of the research. The workshops added an additional layer of rigour, permitting us to 'sense-check' the findings and proposals towards closing the gap. Further it has allowed us to deliver recommendations which are feasible with resource and time investment; and which have been widely considered as capable of delivering a step-change.

In an attempt to future-proof this work, we employed a wider view of the lifecourse of researchers, i.e. skills requirements and opportunities to build capacity at different stages of a research career. To facilitate this, the study team has reviewed a number of comparable models for career stages and progression. These include the Vitae Researcher Development Framework (see Vitae 2019a and 2019b for example), which draws on European Commission work that sought to provide a clearer structure for researcher careers and provides a broader understanding of career development in the workplace,[5] and the typologies derived by Prof. Matthew Flinders (2020) and the European Commission's MORE3 project (2016). There is strong overlap between Flinders' findings (2020) and the MORE3 work, captured in Figure 4 below.

[5] See: https://www.vitae.ac.uk/researchers-professional-development/about-the-vitae-researcher-development- framework/developing-the-vitae-researcher-development-framework

Figure 4 Comparison of Flinders (2020) and European Commission (2016: 22-23) C MORE3 typologies of leadership.

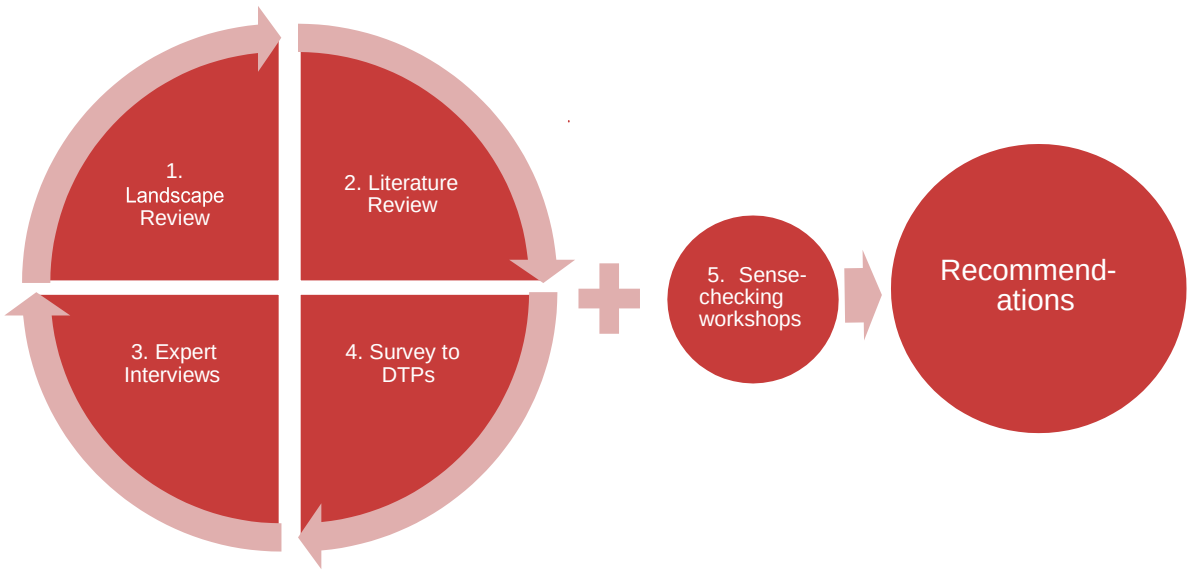
Flinders' Typology of Leadership		EC MORE3 Typology of Leadership	
Descriptor	**Definition**	**Descriptor**	**Definition**
Phase 1	Doctoral	R1: First stage researcher	Up to the point of PhD award
Phase 2	Postdoc	R2: Recognised researcher	PhD holders or equivalent who are not yet fully independent
Phase 3	University Scientist (lecturer/senior lecturer)	R3: Established researcher	Researchers who have developed a level of independence
Phase 4	Professor	R4: Leading researcher	Researchers leading their research area or field

The Flinders typology is useful and makes particular sense in the UK as it maps to professional status. The MORE3 typology is less tied to UK vernacular and so is principally used here in an attempt to make the report more accessible, and more obviously relevant to overseas users. The typology of first stage, recognised, established and leading researcher have been used to aid recruitment to the interviews. This was mirrored in the analytical approach. The data certainly shows different experiences in each of the career stages and there was also a contribution made from all participants about the 'needs' of other career stages. As a result, we preserved the career stage of participants during analysis and report this alongside quotations to provide context. The quotes and evidence we present inform the future of doctoral education pertaining to data-driven skills. To reach a sophisticated analysis, this report will lead with the thematic contribution and explore if and how different actors perceived issues differently.

1.4 Methodology and Research Questions

This report presents the analysis from five distinct research phases of the study (see Figure 5), which combine to provide a robust methodological framework: Literature review, a landscape review, interviews with experts and stakeholders (n = 38, see tables A2-A4 in Appendix B for more detail), a survey to DTP training leads (completed by eight of 14 DTPs), two workshops with doctoral candidates (n = 11) and four workshops with established and leading researchers (n = 14) to test recommendations. The workshop participants critically discussed propositions, a series of 'alternative' approaches to data-driven skill training available in Appendix D. The propositions were derived from analysis of data collected through the first four phases of data collection. Collectively the data permits critical inquiry into two broad areas of enquiry around data-driven skill gaps: i) The lack of access to training in a particular skill, method, or field where the presence of this opportunity would strengthen the researcher and the research they produce, and ii) The lack of knowledge to evaluate what skill, method, field, or provider would best fill the gap identified. This second point also relates to career stage, for example, the senior academic who has identified a skill gap but isn't aware of training opportunities for staff.

Figure 5 Methodological strategy to deliver evidence-based recommendations



Source: Authors' own elaboration

The five phases of inquiry were designed to deliver robust data in a limited time frame, maximising engagement with experts at all relevant career stages.

These areas of inquiry were identified at the start of the project, in partnership with the Steering Group and colleagues at the ESRC, and comprise the following:

- Stage 1: Needs analysis. What are the core skills needed by social scientists and what advanced training is needed and at what scale?

  This focuses strongly on the master's level and examines the core training set out by the ESRC's 2015 Postgraduate Training and Development Guidelines. There is a strong focus on how the Guidelines have been interpreted by Higher Education Institutions (HEIs) and educators and how doctoral candidates experience the learning produced. The chapter extends to consider specialised training gaps. A critical approach is used throughout to question whether training builds confidence in using data-driven skills. Evidence informing this stage is presented in chapter 3, which closes with some initial recommendations for how the Guidelines could evolve.

- Stage 2: Doctoral interventions. What interventions are needed at a doctoral level? What is the scale of change needed in the curriculum? What is the gap between our current postgraduate research and development guidelines and the core training identified as being needed? Do we need to shift how we structure doctoral education to produce confident users of data-driven skills?

The methodological approach to this study has allowed novel themes to emerge and a third consideration that has developed through the evidence has become a key focus of this report:

- What are the structural barriers impacting upon high quality teaching or learning of data- driven skills in the UK?

Evidence relating to stage 2 is presented in chapter 4. The chapter beings with scrutiny of institutional, academic-sector and inter-personal barriers to learning. It builds on the evidence, presenting further and consolidating recommendations for the Guidelines and closes with a re- imagining of data-driven skill training. Chapter 5, in addition to delivering a conclusion thinks beyond the scope of this project to how the evolution of the Guidelines recommended, can be implemented. Literature collated to inform this research is presented throughout. Chapter 2 provides a deep-dive into the literature around teaching data-driven skills to foreground established knowledge.

1.5 Glossary[6]

Figure 6 below presents an overview of key terms and acronyms used throughout this report.

Figure 6 Glossary of terms and acronyms used in this study

Term	Definition
AI	Artificial intelligence
BAME	Black, Asian and Minority Ethnic Group
CDT	Centre for Doctoral Training – in contrast to multidisciplinary DTPs, CDTs are thematic and pump prime the development and delivery of specialist training
Co-I	Co-Investigator, holds a share of a research grant
CPD	Continuing professional development
DTC	The former framework to DTPs, Doctoral Training Centres were multidisciplinary and offered a broad range of training across the social sciences
DTP	Doctoral Training Partnership – DTPs are multidisciplinary and offer a broad range of training across the social sciences. Often operating across a number of institutions (12 of the 14) using a distributed leadership framework.
ECR	Early-career researcher, usually temporary posts between a PhD and secure post
EDI	Equality, diversity, and inclusion
ESRC	Economic and Social Research Council
GDPR	General Data Protection Regulations
GT	Grounded theory
The Guidelines	The ESRC's Postgraduate Training and Development Guidelines, Second Edition 2015
HE	Higher education

[6] For consistency and ease of reading across reports, the glossary is informed by the recently published Review of the PhD and some definitions are verbatim (Tazzyman et al, 2021:4)

HEI	Higher education institution
NCRM	National Centre for Research Methods
OECD	Organisation for Economic Co-operation and Development
PhD	Doctor of Philosophy
PI	Principal Investigator, the grant holder of a research fund
Postdoc	Temporary position that allows a PhD to continue their training as research and gain skills and experience for an academic career
Q-Step	Q-Step is a 17-centre UK network of 17 centres that have focused on undergraduate social scientists and advanced the understanding and use of quantitative approaches to social inquiry.
RTSG	Research and Training Support Grant available to ESRC-funded PhD students and sometimes equivalent sums are available to students supported through other funders
SoTL	Scholarship of learning and teaching, referring here specifically to literature about learning and teaching data-driven skills and research methods
TNA	Training needs analysis
UKCGE	UK Council for Graduate Education
UKRI	United Kingdom Research and Innovation host 7 Research Councils including the ESRC
Notes on terminology: Throughout the report we refer to PhD students and alternatively as doctoral candidates. These terms are considered synonymous. We distinguish between doctoral candidates who are completing their master's qualification.	

2. Scoping teaching and learning in data-driven skills

Summary

The literature to support data-driven skills educators is significant and substantial and has grown considerably in the last 10 years. It is a rich source of guidance and inspiration for educators. The literature is clear: teaching data-driven skills is tough, and this has not been adequately recognised by many universities, with teaching being delivered often by time-poor junior colleagues. While data- driven skills often form core learning at undergraduate and postgraduate levels in the social sciences, the courses themselves do not attract senior management investment. For example, with the exception of Q-Step centres, methods learning is restricted to 1 or 2 courses through the entire undergraduate degree, and students perceive them to be 'added-on'.

Multi-disciplinary and large 'core' class sizes prevent or diminish the teacher's capacity to connect with students. Educators need both pedagogical knowledge and knowledge of data-driven skills. They also need 'substantive' knowledge, a sound understanding of the disciplinary home of the students in order to anchor learning in their disciplinary vernaculars and research traditions. This makes teaching data-driven skills well, a 'greater' occupation than teaching disciplinary courses.

Learning methods is more than how to 'do' data-driven skills: it extends to consider how to 'be' a researcher. This is achieved by disciplinary colleagues exposing their researcher identities and discussing the 'method(s)' behind the knowledge (move from what Durkheim said, to what Durkheim did). Students are more able to invest in data-driven skills learning when they see its disciplinary value.

Training that focuses on skills is needed, though this most usefully builds on a foundation of conceptual knowledge. Data-driven skills, particularly around numeric data, are difficult to learn. Skills should be developed with educators there to help support learning. Experiential learning and working directly with real data 'challenges' allows students to apply learning beyond the classroom.

Educators need to connect with students' aims and ambitions. Reflexive working enables students to imagine their own research path and imagine the researcher they want to be. Creating learning experiences where they can practice this, are essential to build confidence.

2.1 Introduction and purpose of this chapter

While this chapter does not directly address specific research questions, it is essential to foreground the conceptual understanding that underpins data-driven skills before turning to skill development and practice. To do this, we present a review of the pedagogical literature centred around teaching data-driven skills. The ESRC Postgraduate Training and Development Guidelines (hereafter known as the Guidelines), published in 2015, were developed during a period of new and significant investment in learning and teaching scholarship. There has been a steady increase in the quantity and quality of critical writing on, and research into

pedagogies for enhanced data- driven skills learning over the last decade which should inform the next iteration of the Guidelines. In line with Nind & Lewthwaite:

> We contend that building capacity in research methods requires building the pedagogic culture surrounding this field. (Nind & Lewthwaite, 2018b: 399).

Teaching data-driven skills is difficult (Williams et al, 2016), and so is learning about them. Building confidence in applying data-driven skills beyond the classroom (Howard & Brady, 2015) is tough. This chapter is dominated by the scholarship on teaching and learning (SoTL) literature around the pedagogy of data-driven skills, and will begin with how careful engagement with this work can add value to learning experiences.[7]

This deep-dive into the literature permits focus on what is known about creating effective data- driven skill learning experiences. It foregrounds the exploration of data-driven skills gaps to demonstrate where our contributors' experiences corroborate and deviate from the evidence presented in the wider literature. Literature drawn from SoTL also infuses the data chapters.

The following sections are dominated by the work of Nind & Lewthwaite (see 2018a; 2018b; also Kilburn et al, 2015; Nind et al, 2016, Nind et al, 2020; Lewthwaite & Nind, 2016) as their contribution to the pedagogy of social research methods is significant. This literature draws from an ESRC project running 2015-19 that took a global, full social science, methods-inclusive approach to better understand post-graduate training. This contrasts to other writing which tends to focus on individual practice and as such becomes limited to discipline, (Probst, et al, 2016; De Marrais, Moret & Pope, 2018) method, or stage of learning (introductory undergraduate; advanced post-graduate). The chapter will consider approaches to creating learning experiences perceived, by experts in the field, to work. The chapter, like the work of Lewthwaite & Nind across their publications, draws on literature that engages at the more micro level of a course or classroom, and explores:

- Barriers specific to learning data-driven skills
- Connecting with students
- Building capacity and using pedagogy explicitly
- Building conceptual understanding that bridges to disciplines
- Being a researcher and the value of practice

2.2 Barriers specific to learning data-driven skills

In terms of learning data driven skills, we found a significant literature around barriers and enablers to learning data driven skills where students are learning with numeric data (Hillman, 2014). Further, we found that some terminology, such as data analytics and skills, have been understood only in quantitative terms (DCMS, 2020; Lenihan & Witherspoon, 2020: for an exception see Lewthwaite & Jamieson, 2019). This imbalance may reflect the 'problematic' reputation of teaching quantitative methods in particular. This section will explore specific difficulties in teaching skills.

[7] We also mobilise quotes and insights from our primary consultation, where relevant

The impact of dividing skills into quantitative and qualitative silos is discussed briefly and returned to in 4.2 when discussing structural barriers to learning and teaching.

Within sociology, there was not just a division between quantitative and qualitative but a 'critical turn' away from using numbers. Payne (2014) argues that the division originally arose from a rebellion against the early social research focus on quantitative surveys and poorly taught quantitative methods. This account of a turn against positivism could be understood more optimistically as a turn towards critical research that asks 'why?' rather than 'what?'. There is of course a place for analysing numeric data in this critical approach. Once data- driven skill teaching was expanded and adopted by sociologists rather than statisticians, quantitative methods were relegated to an isolated course or module, often not integrated or taught by researchers with strong quantitative experience. With the emergence of methodologies such as feminist methodology, grounded theory, and phenomenology (and a resurgence of ethnography and archival methods), qualitative methods were increasingly emphasised, and sociology developed as more of a humanities than science subject. As a legacy of this era, there appears to be a common anxiety about quantitative methods among undergraduate and postgraduate students. While this research is linked to one disciplinary field within the social sciences, the rise in courses that span Schools and Colleges in order to standardise the student experience have not sufficiently overcome some of the prejudices of the disciplinary approach.

Current quantitative methods training at undergraduate and postgraduate level are not producing confident learners of, or users of data-driven skills that use numeric data. While Payne (2014) notes the important work of the Q-Step programme in UK undergraduate degrees, up-to-date information is lacking on how well, if at all, quantitative methods and engagement with pedagogical literatures around methods generally, are similarly integrated into postgraduate and professional data skills training. Specifically in terms of skills using numeric data, undergraduate social science students in Europe, North America and Australasia reach more robust quantitative skills than their UK counterparts (British Academy, 2016).[8]

Ralston (2020) found that some (undergraduate) social science students did not recognise the purpose or legitimacy of statistical methods in social research. This perceived lack of 'epistemological legitimacy', Ralston (2020) argues, is an important antecedent of statistical anxiety, suggesting that the problem lies with the historical epistemological opposition to statistical methods in sociology described by Payne (2014). In this regard, data-driven skills, and those using numeric data particularly, have historically, not been central to disciplinary teaching teams at the undergraduate level. Many barriers to learning social research methods (inclusive of quantitative and qualitative) have been identified including:

- Resistance by university managers to protect time within degrees to learn data-driven skills, on the basis that courses with less methods, are more attractive to students (see MacInnes, 2010; McVie, Coxon, Hawkings, Palmer & Rice, 2008; Scott Jones and Goldring, 2015)

[8] Note Q-Step centres launched between 2014 and 2015 and so this British Academy data could not capture the impact of the programme.

- Student feedback is seen as count data, rather than contextual with a careful differentiation between high value learning (see Ryan, Saunders, Rainsford & Thompson, 2014, make this case for social research methods), and enjoyable learning leaving methods courses with a negative reputation (MacInnes, 2010; Wathan, Brown & Williamson, 2011)
- The increasing importance of student feedback data can devalue courses that receive mixed reviews (Scott Jones & Goldring, 2014)
- Criticised for being skills focused (Morgan, 2007) or instrumental (Skemp, 1976) rather than being intellectually challenging or integral to substantive learning (Buckley, Brown, Thomson, Olsen & Carter, 2015)

There are additional barriers that link specifically to learning quantitative methods:

- Quantitative methods are perceived to be difficult courses (D'Andrea, 2006; Humphreys, 2006; Williams, Payne, Hodgkinson & Poade, 2008)
- Difficult because there is often a gap of a few years between learning numeric-based subjects at school and learning them at University (Hodgen, Pepper & London, 2010)
- Maths anxiety as a result of learning numeric-based subjects at school (Scott Jones & Goldring, 2014; Williams, Payne, Hodgkinson & Poade, 2008) and a lack of understanding that quantitative methods are not as difficult as Higher or A-level maths
- A conviction that numeracy will not be needed in the social sciences (with some exceptions in Economics, and to a lesser degree in Psychology and Political Sciences) (Williams, Collett & Rice, 2004).

These arguments combine to produce a 'deficit' view of learning data-driven skills. Importantly, Ralston (2020) notes the compounding factor that students may not see statistical methods reflected in publications or the 'real' world of research, reinforcing the perceived illegitimacy. This argument is supported by Linneman's (2021) review of 2804 journal articles. Self-efficacy and confidence were also noted as important factors relating to statistics anxiety (Ralston, 2020). Given the importance of developing at least a foundational and conceptual (what it can do) quantitative data skills (ESRC, 2015; Payne, 2014; Vitae, 2011), it may be important to engage pedagogical methods to support this, potentially via some of the co-learning and supported experiential learning methods described in section 2.6, below.

If current skills training is poorly perceived, a step change is needed to revive this valuable learning and enhance learning of data-driven skills.

2.3 Connecting with students

The Guidelines (ESRC, 2015) can be understood as a type of curriculum, but how this becomes a learning experience and how best to engage students is less well understood. Any investment in improving pedagogical practices ought to recognise the poor regard students still have for core methods training (see section 2.5). In their 2018a paper, Nind & Lewthwaite suggest a pedagogy of inclusion which aims to avoid an adversarial discourse with teacher versus complaining student; and encourage relationships as co-producers of critical learning, drawing on earlier work from Nind and colleagues:

> Pedagogy is 'the coming together of the teacher and learner and the production of knowledge is a political process with inherent implications for teaching practice' (Nind, Curtin & Hall, 2016: 33)

Here then, educators need space in their curricula to respond to the learners they're engaging with each year, having the chance to know them, and adjust the learning experience accordingly. Smaller class sizes work because they permit a connection to build between the educator and the ambitions and learning objectives of each student. As yet, methods pedagogy is not well understood in terms of being 'specified and handed on to others' (Nind & Lewthwaite, 2018a: 85). Rather, it champions a more careful, learner-engaged strategy of a living pedagogy that evolves with each cohort. This is about process, rather than best practice, or designing the perfect programme to reach all learners, presented 'on repeat'. This is more than discipline or content, but about standpoint. Diversifying a reading list to be inclusive of disciplines is insufficient to meaningfully engage at this level, for this would only reach the idea of an average learner, rather than actual individual students and their specialist interests, which of course, will change year on year.

One potential pedagogical solution is to create a 'flipped class'. Here the formal lecture or knowledge delivery mechanism is moved outside of the class session, and the 'homework' or application of knowledge is moved in (Wilson, 2013). With so many practiced in delivering online lectures, now could be an ideal time to practice this approach. Michaelsen, Knight & Fink (2004) describe the educator here foregoing a staged performance to act as a guide (see too Sweet & Michaelsen, 2012). Morris et al used this approach to teach research methods creating more space for team-based problem-solving during class time (2020). This approach has been linked with improved communication skills and peer networking (Huggins & Stamatel, 2015), and improved assessment scores (Chung et al, 2009; Heyborne & Perrett, 2016). It was also linked to improved preparation (Balan, Clark & Restall, 2015) which could in turn, disproportionality and positively impact on data-driven skills teaching where there is a wide range of abilities and prior-learning in the classroom (Nind, Kilburn & Wiles, 2015).

2.4 Building capacity and using pedagogy explicitly

Building capacity relates to the support required to help students appreciate the value of data- driven skill learning, and the investment required from them. Students are not necessarily encouraged to recognise the value of data-driven skills despite them being delivered through 'core' learning in most undergraduate social science degrees in the UK.

In the social sciences method and methodology may not be taught at undergraduate level and if they are, they are usually available only in 2nd or 3rd year which can leave students feeling that research skills are less important than their disciplinary classes. Published, peer-reviewed literature on the teaching of data-driven skills is very limited (Wagner, Garner & Kawulich, 2011; Garner, Wagner & Kawulick, 2009; Hesse-Biber, 2015) and a very small proportion of peer- reviewed articles describe their data-driven approach in accessible language or sufficient detail. A review by Kilburn, Nind & Wiles (2014a) highlighted these gaps in the literature, with existing research offering limited insight into how research methods are taught or how students learn best.

Lewthwaite and Nind (2016) contributed further insights through expert panel research with expert methods teachers with mixed content and pedagogical expertise from the UK, US, Australia, New Zealand, Africa, and Europe. This research critically transcends the quantitative/qualitative divide by seeking the expert view of scholars with an established reputation for delivering and writing about strong methods training. The study found that advanced social science data-driven skills training had often been developed by methodological experts without a teaching background. The authors argue that this resulted in training with high-quality 'skill' content but where engagement with pedagogic knowledge would enhance the learning experience. To provide a positive learning experience, teachers need to also draw on substantive content knowledge (tough if it's a multi-disciplinary class) and pedagogic knowledge. There were few examples of helping students visualise or understand their learning journey. In her analogy of 'trees' as a pedagogical device, Ferrie (forthcoming) uses the example of an oak tree to symbolise qualitative approaches with branches close to the ground that are metaphorically, easy to climb on to (although effort is required to climb up the branch and beyond the canopy to 'see' our research field). Learning can feel disrupted from one week to the next as the focus moves to another branch of the tree and students who appreciate this are less likely to be anxious that the course does not feel cohesive. In contrast quantitative methods is like a tall single-trunk pine tree, that needs learning in incremental stages (MacInnes, 2018): material must be learned in order, learners must trust in the educator (that the effort will earn reward): and can deliver a 'view' from the top that can 'see' a long way (though not very good for seeing grounded experiences). The different in learning styles required may account for the anxiety and resistance to learning (explored earlier in section 2.2).

The panel contributing to Lewthwaite and Nind's work (2016) highlighted that experience of 'good' methods training was an important foundation that experts drew upon to develop their teaching approaches. Lewthwaite & Nind (2016) examined the influences of best practice. In their work with leading experts, SoTL relating specifically to teaching research methods was a tool used, so too were each educator's training experiences. In particular some contributors discussed wishing to deliver positive learning experiences for students inspired by what worked for them as learners. Over time these method experts were able to accrue teaching expertise through experience, careful reflexive work, researching of teaching and learning and leadership roles in learning institutions. As data-driven skill academic teaching roles tend to be given to junior colleagues, a solution, other than decades of time, needs to be found to equip them with pedagogic knowledge. The strong emergence of SoTL over the last 5-10 years (Nind, Kilburn and Luff, 2015) provides an extended resource for data-driven skills educators to draw on in addition to their experience.

Methods teachers may need to have a research profile to be seen as a credible methods teacher. Nind & Lewthwaite (2018b) drew out this point when discussing the content knowledge methods teachers needed being both about their discipline (what their research is about) and about their method (how to find things out) and this 'dual' role, may make it more difficult to add a third realm to their practice, around pedagogy. It seems then that good data- driven skills teachers need to be experts in the general, where universities tend only to recognise and reward expertise in the particular. To impose another analogy, HEIs are comfortable with the concert pianist

and the individual ‘star’, and data-driven skills require leadership more akin to a conductor, able to help students understand the whole orchestra. This is echoed by Nind & Lewthwaite around the importance of disciplinary expertise and educators who have a very diverse class have a significant challenge to teach students well (2018b).

2.5 Building conceptual understanding that bridges to disciplines

Teaching social research methods is tough (Earley, 2014; Nind & Lewthwaite, 2018a). Learning experiences are different to those provided for substantive, disciplinary courses and students need to learn the language, the concepts, processes, and technical skills (Howard, & Brady, 2015). This is markedly different from disciplinary learning, but little is done to make this apparent to students (MacInnes, 2018) and this is helpful, for students to see the value of skill learning and underpinning conceptual knowledge. Nind & Lewthwaite (2018a) distinguish between teaching method (skills development) and teaching methodology (conceptual understanding including ontology and epistemology) and argue that as both are needed and present, time is needed to help students distinguish between the two. We present a number of reflections from consulted stakeholders and researchers to illustrate this point.

Theory and disciplinary knowledge are essential to support all other skills. Knowledge needs an anchor. Academics may take this for granted:

> ‘Too much focus on data and empirical production what does it mean? Most importantly having a shared appreciation of the theory and construct of the research, the question, and the justification for any methods developed.’
>
> Wider Stakeholder - Interviews

Forming a research question and understanding what evidence is required to deliver an answer, were considered key data-driven skills.

> ‘Data discovery is something useful for all researchers, having the skills to determine whether data could be useful, finding, accessing and understanding this, is essential.’
>
> Recognised Researcher - Interviews

Understanding what data has been used to date and how, involves strong engagement with the literature available. From an established scholar:

> ‘Have to know where data comes from, how reliable they are, have to be numerate and critical, understand what inferences can be made or note problems of the data.’
>
> Established or Leading Researcher – Interviews

Conceptual understanding had been disrupted by a strong focus on skill learning and students could be stronger in their capabilities to design research.

2.6 Being a researcher and the value of practice

Kilburn et al (2014) produced a framework that championed: making research visible; experiential learning; and reflective practice. Kilburn et al's (2014a) review appraised 24 peer- reviewed papers on pedagogy in social research methods from 2007-2013 and appears to be the most up-to-date review available, focusing mainly on formal teaching activities. This section will begin by introducing practice and then turn to using student's own data as a bridge to understanding the students' own research field and then move to the value of reflexive work.

Academic researchers are practitioners, though our research labour is often hidden from students as we focus on findings more than process within teaching. Kilburn et al. (2014a) describe the goal of *making research visible* as the process of making abstract concepts and aspects of the research process tangible and transparent through active learning. It's a useful strategy for moving beyond skill towards fluency. Active learning, as opposed to more traditional didactic techniques such as lectures, requires learners to engage in targeted activities. Examples of quantitative active learning exercises include computer demonstrations and simulations of statistical tests and concepts, critically appraising quantitative outputs, and practising analysing sample datasets in software packages (Kilburn et al., 2014a). Qualitative active learning exercises include, for example, using films to demonstrate aspects of qualitative research (Saldaña, 2009), reflective research diaries (Punch, 2012), or research enquiry and using podcasts of interviews with academics (Kilburn et al., 2014a). Teachers in Kilburn et al.'s (2014a, see too 2014b) work, emphasised the importance of striking a balance between didactic and active learning and providing learners with a strong foundation in key concepts before engaging in practical tasks. Teachers also emphasised the increased time needed to implement active learning, but advantages included opportunities for students to practise and receive feedback (confidence building) and explore methods without anxiety (Kilburn et al., 2014a, b).

There is here, a useful reminder that students don't learn 'researching' so much as they rehearse the practice of 'being' a researcher (or a political scientist, or economist or social historian and so on). The first step in achieving this understanding of practice, is to clarify that academics 'do' research and are producers of knowledge. Some students may consider lecturers as simply *re*producers: curators of knowledge presented via reading lists and course structures. This issue is exacerbated by large interdisciplinary methods classes where teachers feel less able to draw on their own research for fear students will criticise the course for favouring a single disciplinary approach. The interdisciplinary structure is preventing educators from utilising one of the most effective pedagogical strategies: bridging between taught classroom and the research field.

Kilburn et al.'s (2014a) review also valued experiential learning, whereby students collected and analysed their own data, or used real-world data to learn how to use a method in an approach similar to 'learning by doing'. Teachers argued that knowledge of experience cannot be taught in isolation or abstraction, so different pedagogical methods are needed to provide experiential learning, such as academic supervision or an apprenticeship-style model (Kilburn et al., 2014a). In Kilburn et al.'s (2014b) report on short courses in advanced research methods, teachers co-analysing data with learners was reported as a valuable way to support learning,

potentially providing a bridge between more structured active learning exercises and more independent experiential learning. What may be critical, is that students experience the whole project, rather than experience applying a skill, to understand better the consequence of using that skill, in that place and time, on the knowledge produced. The aim of this approach is to build confidence with the educator there as a safety net, and where the student can experience the complexity of the field. Teachers in Kilburn et al.'s (2014a) reviewed studies placed a high value on experiential learning, emphasising that experiences of 'real' data gathering, and analysis are essential and should be incorporated as core components of a curriculum. The dissertation delivered within a masters is an ideal space for this work to be performed.

Using expert interviews, focus groups, and video-stimulated dialogue with social science research methods teachers, Nind and Lewthwaite (2018) explored the challenges of methods teaching, focusing on postgraduate learners. Teachers discussed the importance of learners working directly with data as a way to connect them to research evidence and ideas (Nind & Lewthwaite, 2018a). Experiential learning can serve to spark learners' interest in data, thereby facilitating deep learning and understanding of the purpose and value of different methods (Nind & Lewthwaite, 2018a). Learner perspectives are also included in a study by Nind et al. (2020) that used diaries and focus groups with postgraduate students and early career researchers to explore their perspectives on learning research methods in the social sciences. Learners emphasised the value of creating their own data within a learning experience. While textbooks and written information were useful, they were not sufficient to equip students to use a method, and directed, guided experience was needed.

There may be a difference in how quantitative and qualitative learning experiences provide experiential learning. Qualitative classes prioritise data collection, such as conducting interviews (DeLyser et al, 2012) which in a semester-long course will reduce time to learn and do analysis. Rather quantitative courses prioritise analysis (Aguado, 2009), though have developed much over the last decade to provide real-world data (Adeney & Carey, 2009). There are practical challenges with using pre-existing or secondary data for teaching, as the data needs to be 'clean' and in a format that is digestible and manageable for novice researchers (Kilburn et al., 2014a, b). Here, involving students in these tasks of data cleaning and data management provides valuable experience. Teaching qualitative and quantitative approaches together provides learning opportunities that deliver a sound understanding of project management tasks.

To appreciate what data means, it is essential it is situated in context, including academic understanding, and of course understanding disciplinary context can help avoid mistakes:

> 'Must make sure that we don't repeat the 'positivist' mistakes of the 60's and put too much faith in data alone. Need to encourage researchers to remain open-minded about data in all forms.'
>
> Wider Stakeholder - Interviews

There is evidence that using real-world data that is also relevant to a student's experience can improve learning (Atkinson, Czaja & Brewster, 2006; Lindner, 2012), as can collecting their own data (Strangeld, 2013) and writing reflective journals about learning (Denton, 2018). We found no literature that explicitly evaluated experiential learning of project management (though there would be some implicit learning if conducting interviews or deciding which variables would work in a statistical model). Literature on project dissemination was limited to writing a report in an academic style (Shaw, et al, 2011; Eisenhart & Jurrow, 2011), rather than employing broader approaches designed to reach beyond the academy. As Kilburn et al comment, (2014a: 201) too much focus on doing may 'avoid the practical, pedagogical and ethical issues' inherent in research that produces a robust evidence base.

Reflection creates a learning space where students can assess their own development and future learning needs. Educators in Kilburn et al.'s (2014a) review incorporated reflection on the research process into their methods teaching, typically described as 'reflective practice' or 'reflexive learning'. This tended to involve personal journals or group discussions, whereby students were encouraged to reflect on their experiences of learning and conducting research (Kilburn et al., 2014a). Importantly, reflection was used mainly in relation to qualitative methods, for example, it was used to critically engage with positivism (Cox, 2012) or to better understand identity and how this informs practice (Thien, 2009). Reflection also appeared to be widely used in teaching ethnographic and narrative methods in order to encourage students to think critically about the research process and their own assumptions and expectations (Kilburn et al., 2014a). Reflective practice can be challenging where students are trying to learn to use a method while simultaneously questioning its relevance and validity (Kilburn et al., 2014a). Teachers in Kilburn et al.'s (2014a) study argued that incorporating reflexivity into methods teaching can benefit students by allowing them to share research experiences and develop a reflective, critical view of methods. This is supported by the Vitae Researcher Development Framework (2011), which includes critical thinking and creativity as key skills across the researcher lifespan. It is also lauded as a valued skill of non-academic employers (Carter, 2021).

Nind & Lewthwaite (2018a) interviewed leading methods teachers and found that many use reflection ahead of teaching in order to help them make clear their approach and standpoint, and to draw on their own research experiences. Kilburn et al (2014) did not find many examples of reflection (Delyser et al, 2012 were noted for their efforts) amongst quantitative methods teachers. There are examples in Lewthwaite and Nind's (2016) work. As all data-driven skills require a conceptual understanding of their capacity and limitations in producing knowledge, reflexivity should be considered a vital tool by educators and explicitly taught to students.

3. ‘Core’ and ‘Advanced’ data-driven skills needed by social scientists

Summary

Our analysis suggests that the ESRC’s Guidelines should evolve to make more detailed reference to project management skills. This includes open access skills such as cleaning data files, fixing data inconsistencies, performing data enhancement, generating data in multiple file formats for preservation and dissemination, and generating code lists. Allied to this, rules around intellectual property, safeguarding self, safeguarding data, and research integrity should be included in core training.

Our research indicates these skills are most valued when a candidate can also demonstrate team- working, communication skills, strong knowledge of ethics, and legal expertise. There is evidence that such ‘soft’ skills, in being described as ‘transferable’ have been interpreted as non-vital, and evidence is presented that these skills are critical for producing confident social researchers.

The Guidelines should make it clear that a broad-based approach relates specifically to conceptual knowledge and data literacy. Skill development may be incorporated, but it is not the primary aim of the broad-based approach. Students are not expected to be ‘skilled’ across all the areas required by the Guidelines. Rather, students need time to invest in specialist skills after core training. Time is needed to actively engage and practice the skills students intend to use in their doctoral research.

Digital skills should be considered as core skills for data-driven social science students. Specifically, training is needed around digital research design, data capture, curation, storage, and digital data analysis. Machine learning, AI, and digital skills are recognised as ‘must have’ for scholars working with numeric data and core training that prepares for specialised training in these fields will add value. There is not the same demand from established and leading researchers working with text or visual data although there is from doctoral candidates for training in these fields. Broadly, doctoral candidates appear to have an appetite for digital training regardless of the form of data they are creating/encountering. This is causing a capacity issue and doctoral candidates and recognised researchers may be best placed to deliver training in this area. The landscape review indicated that digital skills are valued globally and across all academic disciplines. The UK is at a tipping point where embedding digital skills into doctoral training will be an asset of our UK-based scholars, and not doing this will cause a deficit in the UK’s data-driven skill set within a few years.

Doctoral candidates need well-developed dissemination strategy incorporating open access data, digital and social media strategies and peer-reviewed journal articles to be academy-ready or industry-ready. Many doctoral candidates and recognised researchers felt that leading researchers did not appreciate this pressure.

The Guidelines should incorporate elements from the General Research and Transferable Skills Training (ESRC, 2015: 12-15) into their Expectations for Core Research Methods Training (ESRC, 2015: 7-10) using the four-pillar model (see executive summary, Figure 1). Re-framing the General Research and Transferable Skills Training as a Doctoral Education Framework will underpin the need for

continued professional development during the doctoral years. In fact, a culture shift where data-driven skill learning is understood as career-long is needed.

3.1 Introduction and purpose of this chapter

In this chapter we focus on the first stage of the study (needs analysis) and its component research questions:

- What are the core skills needed by social scientists?
- What advanced training is needed and at what scale?

This chapter focuses predominantly on core training and the master's experience as an interpretation of the ESRC's Guidelines, before considering advanced training needs. Recommended changes aim to produce more confidence among social scientists in their use of data-driven skills and rigour through their research. Where changes are proposed to the Guidelines, this is understood as recommendations that will impact on the delivery of data- driven skills training provided for ESRC-funded students. Regard for the wider student body is given where possible. To address these questions, this chapter will present the data around the current context and content of data-driven skills training, and will:

- Consider to what extent UK students are being trained to 'set the standard' of global excellence
- Will outline evidence that data driven skills gaps relating to project management, including research integrity, need to be embedded in core training
- Critically evaluate the current structure of core data-driven skill learning outlined in the Guidelines and value of advanced or specialised learning during masters funding
- Present evidence that digital skills could feature in core training not least, to underpin future advanced and specialised digital skill learning where relvant to a students' research aims and ambitions
- Deliver evidence that anchoring dissemination and well-being skills within masters courses will support students' learning.

This is the first of two 'data' chapters that foreground the findings from the five research phases. Each chapter is substantial and covers extensive ground. At the core code level, all of the data produced fits into one of two themes: i) 'What we do?' (i.e., content/context) and ii) 'How we do it?' (i.e., pedagogy/structure).

In presenting the evidence behind these key findings, the chapter will begin with examining the master's period, and core skills before turning to advanced or specialised study that is needed throughout the doctoral years.

3.2 Core skill needs

Social science doctoral candidates with ESRC awards have access to core methods training within their masters and all Doctoral Training Partnerships (DTPs) refer to the 2015 Postgraduate Training and Development Guidelines (ESRC, 2015) in full. The

Guidelines themselves list fundamental skills and concepts that were 'expected' to appear in core training and all DTPs interpret this as mandatory,

'We have assumed that 'expected' means 'required'

(DTP Training Lead – Survey; echoed by other DTPS).

As a policy document then, the Guidelines have heavily influenced practice across the sector and is impacting positively on postgraduate and early career scholars.

'A positive change in the last 5 years predicated by the ESRC – [PhD] funding criteria requires us to engage in core skills that PG students should come out with. That has been very effective.'

Established or Leading Researcher - Interviews

The areas covered by the 2015 Guidelines are broad, and the distinction between them is not always clear. For example, 'research methods' learning outcomes in the Guidelines include 'competence in understanding and applying a broad range of research methods (including quantitative, qualitative, and mixed methods) and the use of software for their application' (ESRC, 2015: 7). Such competency across a wide range of methods is delivered within a master's degree, typically meeting 60-credits of a 180-credit programme. Usually 60-credits are reserved for an independent project towards submitting a dissertation; leaving 60 credits for students to choose from disciplinary or advanced method courses. Typically, the core methods courses are divided into three 20-credit courses covering qualitative methods, quantitative methods, and research design. Not all students choose to complete their doctoral studies. This robust master's is one way of ensuring candidates leave with a valuable qualification. For EDI candidates, this reduces the 'risk' of committing to PhD study.

Most DTPs guide their partner institutions towards full compliance of the guidelines embedded within disciplinary masters, others provide masters with advanced methods options and others have created general social science master's courses taught between their partnership as a single provision for all ESRC funded students.

The success of the core training was evident among all participants:

'Doctoral preparation has improved the skills of postgrads notably. The 1+3 programmes have a good focus on skills development early on.'

Established or Leading Researcher - Interviews

Making training credit-bearing with assessments, and embedding it within a qualification (masters), helped students invest in the learning. The message that the training was integral to the funding was an incentive and recognised as such and valued by established and leading researchers and wider stakeholders. Now that a data-driven skill learning experience is accepted as the gold standard, it may be useful to think of the barriers to learning which go along with this model or have emerged through its application.

3.2.1 Project management, data literacy & research integrity

Principles of research design and data management are included in the Guidelines but project management is not explicitly stated as 'expected' learning. From the literature and other phases of inquiry, project management skills that are needed include managing resources including time, and should include open access skills such as cleaning data files, fixing data inconsistencies, performing data enhancement, generating data in multiple file formats for preservation and dissemination and generating code lists. Allied to this, rules around intellectual property, safeguarding self, safeguarding data, and research integrity are needed.

Evidence that **project management training is a data-driven skill gap** appeared through most phases of inquiry:

> 'For me, what stands out, I'm writing up my thesis now and what stands out for [training gaps] is project management. If I had had more training and more of a holistic approach to data analysis because there are so many decisions that you have to deal with. It's not a right or wrong answer it's just that you need to be able to justify what you did, and I think that's where I've lacked in confidence to justify decisions, because I feel like, I'm not properly trained in how to make those decisions.'
>
> Amelie – Workshops with Doctoral Candidates

This section outlines the evidence to support project management being a wide-spread data- driven skill gap requiring a structural response. Researchers must understand the whole project, and consequences of framing the research question a particular way, on the data found, and analysis used. Amelie's comments were supported by others and point to training required during the PhD though many championed earlier engagement during the master's too:

> 'People don't come to us saying that [they need training in project management] but I think that's an important part of any researcher's career really, to learn project management and data management absolutely. But to have a PhD project approved, they don't need to have a research or data management plan …. If you've not got your research question already and then you've got heaps of literature to wade through, that, the [investment in] research data management doesn't come until much later. Probably a bit too late.'
>
> Training Lead – Expert Workshop

Taking a whole project approach, was a skill many felt post-graduate students were not accessing during their core masters training or PhDs. This might relate to the dominance of a single project producing a PhD thesis, which provides experience but does not include time for reflection to consider what knowledge would have been produced had alternative decisions been made. One solution suggested was to use core training assessments to run smaller whole projects presented as a portfolio, drawing together learning from all courses. Students could write dissemination plans in order to critically assess which global management skills were required for any project and which skills were explicit to their in-depth study. This would then help

students decide what specialist or advanced data-driven skill training was required. Currently skills such as using software dominate at the expense of conceptual understanding.

> 'One area that [students] struggle is the data management. Skills such as data management and merging datasets are much more difficult, even for the [students] who have the advanced technique so to speak. Perhaps we don't spend enough time training in [project management]. So the basic understanding, instead of a great advanced technique, of 'this is my data, so I'm going to implement [data management plan].'
>
> Training Lead – Expert Workshop

A strong way of delivering this learning is to draw on personal experience. From Nind and Lewthwaite, this quote features Julia Brannen's contribution:

> People will always [draw on research experience], and it's a good thing that they want to bring their own research to bear in the teaching. (Nind & Lewthwaite, 2018a: 81).

Here, Brannen is talking, not about research findings, but about how research was experienced and illuminates the small decisions made at every stage of a project, that impact on knowledge produced. Exposure to the educators' reflexive work is vital, if the students are to engage themselves, and so build bridges between the classroom and their own research interests. Further, it is reflection and drawing on research-failures (as well as how research worked well) that distinguishes 'live-teaching-events' from textbooks. Learning how recruitment does not go to plan, how building rapport doesn't always work, how cleaning data takes weeks/months, are significant learning points that do not tend to appear in textbooks or peer reviewed journal articles. Many scholars who feel comfortable discussing failures in a face-to-face classroom are not keen on exposing their errors in a recorded session. For students to access the realities of research training, they need honest learning spaces. Without these, their own normal deviations from their research plan, played out during PhD research, will be understood as individual failures.

Within this call for greater skills around data-driven project management, evidence from the literature review, points to the need for particular skills in teamworking and communication, and for ethics and legal expertise. Five key actions were identified by the Organisation for Economic Co-operation and Development (OECD, 2020) to improve the skills and capacity of a digital workforce towards better data intensive science (inclusive of arts, humanities, and social sciences) that speak to improving project and data management skills:

1. **Enablers for digital workforce capacity development**

   Integrate digital workforce capacity development into broader science policy frameworks and actions

2. **Defining needs: digital skills, frameworks, and roles**

   Identify the key competencies, skills and roles required for data-intensive science in different contexts

3. **Provision of training**

Support training in foundational digital skills and more specialise skills for scientists and research support professionals

4. **Community development**

Support development of communities for new professional roles, learners, and trainers

5. **Career paths and reward structures**

Change academic evaluation and reward systems in order to attract and retain diverse digitally skilled staff

From OECD, 2020: 8

While the second and third action areas have obvious alignment to digital skills gaps (discussed in section 4.2.3), the fourth and fifth demonstrate the direction of travel towards a culture shift required to support doctoral candidates towards a whole project training model.

This would provide a strong foundation to support learning throughout the PhD and towards recognised researcher data-driven skills:

> 'Looking at jobs and even, if I were to apply for a grant, I have no idea around budgeting. I don't know how much things cost. At least having an insight into it, would increase my confidence moving forward.'
>
> Amelie – Workshop with Doctoral Candidates

To find an academic job, even a temporary one with a high teaching workload, applicants often need to demonstrate these skills:

> 'Although graduates sometimes secure permanent lectureships relatively soon after gaining their PhD, they are expected to develop a strong publication record and demonstrate that they can secure research funding, both of which require strong transferable skills.'
>
> Tazzyman et al, 2021: 26

Further this skill, of holistically being able to see the project, enables effective communication with researchers from different fields and increased opportunities to work effectively with inter/multi-disciplinary teams. Fluency about how knowledge production works within your own field is essential to then determine common ground and areas where negotiation is required.

Data literacy is another data-driven skill gap identified. Data literacy pertains to the conceptual knowledge of data production required to understand where one's own research 'sits', and this knowledge requires broader understanding of what is possible

in terms of rigorous knowledge production. The broad-based approach to the Guidelines (ESRC, 2015) aimed to deliver data literacy.

The Guidelines (ESRC, 2015) cover a comprehensive range of data-driven skills, methodologies, and method. According to contributors, the Guidelines had been interpreted as being more about delivering skills than concepts; and this was not possible across the range of items covered. To address the gap then, and produce confident and data-literate students, the Guidelines could evolve to emphasise the conceptual approach required in core learning. Certainly, students should not expect to be specialists in all the methods covered.

Conceptual learning means understanding the different forms of data, the context in which they were produced and the options available for their collection, collation, and curation. Data literacy is about knowing enough to know that there are multiple pathways towards knowledge production and why some pathways are less legitimate than others for a particular project or research question. So while the 2015 Guidelines were designed to provide literacy, this is a gap that persists. Data literacy was valued:

> ‘Technical skills around reading, understanding and context are important, it is important [as a qualitative researcher] to have the skills to read and understand quant[itative] papers - to develop critical thinking skills.’
>
> Recognised Researcher - Interviews

Here the link is made to strong critical thinking skills and linked to this is the principle underpinning research integrity of, if in doubt, return to the data:

> ‘Every [social scientist] needs to understand statistics, for better or for worse. All the core debates relate to the data. You would be completely disabled if you do not understand the data yourself, you do not need the same skills, but need to understand how it works overall.’
>
> Recognised Researcher - Interviews

This argument works equally well with traditionally qualitative approaches that also need to be conceptually understood and valued. In the workshop with doctoral candidates, students discussed the value of learning qualitative approaches to help them harness language for example, relating the usefulness of discourse analysis and writing short promotional statements on twitter. The separation of qualitative and quantitative approaches was seen to work against data literacy:

> ‘There's a lot of issues with just departments not talking to each other at all so like instead of like actually doing mixed methods it's like you put your qualitative hat on and then you take it off, then you put your quantitative hat on. And it's just like rather than picking and choosing what tools are relevant for you to answer the questions that you want to ask yeah it's like there's no like [conversation] and if you're doing anything interdisciplinary it's just a nightmare.’
>
> Kait – Workshop with doctoral candidates

Such data literacy was seen as particularly important given the rise in new forms of data. Relating to big data, some felt qualitative skills were required to critically evaluate the contextual value and determine reliability. Similarly social media data was used as a strong example of data that could be captured numerically but needed analytical approaches that could engage with sound, visuals, and words in order to fully extract the meaning. In turn, data literacy is the cornerstone of working in multi- or interdisciplinary teams.

The lack of adequate learning spaces around project management are allied to a **data-driven skill gap around research integrity**. Participants at all career stages and DTP training leads reported this gap. Some referred to reproducibility as a missing skill. Growing as a sign of quality in some disciplines (Political Science, Economics and Psychology), reproducibility is less easy to align with traditionally qualitative research. Here concepts such as authenticity, credibility and transparency are needed with increased demand reported by DTPs. Perhaps traditionally assumed to be delivered via supervisions, it seems clear that early career scholars are increasingly aware that this in an area where they require more support and training. One DTP delivered a 10-week Project Management course for all first years to learn together. The content met some of the gaps identified (particularly around Research Integrity) and successfully produced a strong peer network, or cohort of ESRC funded students.

Some participants did not describe their training gap as being about research integrity:

> ‘[There is demand for] 'softer' skills training conducting research into sensitive/emotionally demanding topics, managing mental health/wellbeing, managing change (including in the context of the pandemic), understanding the ethics of research, doing research online.’
>
> DTP Training Lead - Interviews

Yet these examples are very much about researcher positionality and how they safeguard their participants, themselves, and the quality of their work through the course of the project. There appeared to be an awakening to the importance of research integrity in the data:

> ‘Training from the [Office for National Statistics] on safe researcher training and statistical disclosure control is a skill that I didn't understand the importance of until I did my fellowship: what does private data consist of and how do you protect privacy within data’
>
> Recognised Researcher - Interviews

It was clear from the data that Researcher Development units in universities are taking this training seriously. Few established or leading researchers mentioned it as a skill gap, but DTPs, doctoral candidates and recognised researchers all signified its importance.

Project management should be part of 'core' training and assessed by a portfolio approach allowing students to reflect on, and defend their research strategy and identify future specialist learning needs.

3.2.2 Gaps relating to skills valued by industry

The chapter now extends the discussion of project management skills where that involves working with industry/policy actors or in industry/policy spaces demonstrating that these vital skills are equally valued by academic and industry career pathways. Around 42% of PhD graduates move into jobs beyond the academy (Vitae, 2020). While not directly data-driven, here we identify **skill gaps in communication and teamwork** that will help students build networks towards knowledge exchange and impact.

Recent research by the Chartered Management Institute (2018) revealed gaps identified by employers, as they assessed graduate students. Eighty-four percent were perceived to be unable to project manage; 86% did not have interpersonal and communication skills expected; 48% were not considered to have problem solving skills and 48% were not able to work as a team. While this data relates to undergraduate students, the isolating experience of a social science PhD may not help students develop these skills. As doctoral candidates aim for critical depth with their work, and a unique contribution to their field, they can feel isolated and alienated from their peers and wider society. In the social sciences, while students may share offices (disrupted by the Covid-19 pandemic), they rarely share topic. Further they are recognised and rewarded throughout their careers for solo-authored materials (East, Stokes & Walker, 2014). More likely, the students have good communication skills, but as they weren't explicitly valued within the traditional PhD apprenticeship-model, they did not feel confident in 'marketing' these abilities (Ferrie & Scott, 2021). Bringing these skills to the fore, through experiential learning will not just close skill gaps, it will improve students' awareness that skills have been developed. Broadly speaking, the social sciences struggle to communicate their value to industry. In turn, the term 'industry' is not well defined for social scientists and corresponds to an inclusive but nebulous idea of organisations that work to profit.[9]

A report by the Campaign for Social Science (Lenihan & Witherspoon, 2020) predicts that more and deeper cross-disciplinary working involving data-driven skills that span

[9] There is little information available on what the needs are of 'for profit' industry, or how these are the same or differ to the needs of government sectors, charity sectors or indeed, academic sectors. For example, of 23 industry case studies produced by the Campaign for Social Sciences (2019) 15 were derived from business and management and 4 related to economics. Only four case studies were drawn from disciplines beyond these two fields. In context, a recent review of 'pathways' across the UK DTP network revealed that Economics and Business and Management are represented in some way across seven of the 96 pathways in social science inquiry. Thus, there is a vast community of social scientists that are hidden to industry. One of the four case studies not based in business and management or economics, focused on the ethics and governance of artificial intelligence (Campaign for Social Sciences, 2019: 15), a vital area of research for industry and non-profit organisations.

across numeric and textual data will be required in the future. The report indicates that number and data skills are growing in value amongst all disciplines and enrich the qualitative understandings of individual, social, economic, and geo-spatial behaviour. The review of eight private sector companies found that most companies use social science knowledge and skill to understand the current and future behaviours of consumers and the constraints and opportunities of their market, deploying knowledge from economics, market research, psychology, political science, and geography (Lenihan & Witherspoon, 2020). The literature around the value of social science academics to industry is limited (Barnacle, 2005) and there have been calls to understand this better and for learning to trickle down to doctoral students (Borrell-Damian, Morais & Smith, 2015).

Filling this gap could utilise assignments within core learning. For example, Wollschleger (2019) redeveloped their (undergraduate) quantitative methods course in partnership with local non- profit organisations, embedding their real-world problems into class assignments. Engagement, learning and confidence in learning all improved according to class evaluations. Such experiential learning is a gateway to develop skills, and confidence in skills that are valued by employers generally. Embedding such strategies into methods training will allow universities to deliver against the UK Government's industry strategy which aims to 'help businesses create better, higher-paying jobs in every part of the United Kingdom (UK) with investment in the skills, industries and infrastructures of the future' (HM Gov, 2017: 12).

Wollschleger's (2019) approach is an example of 'mission' or 'challenge-led' learning and explicitly rehearses skills of value to industry. Students who can see the link between data-driven skills and a 'mission-orientated' (Mazzucato, 2013) approach will value working with organisations including for profit, policy makers or civil society (Ferrie & Scott, 2021). Here mission-oriented is very similar to the Review of the UK's (Tazzyman et al 2021) call for challenge- led experiences which could be found in placements, but equally built into 'core' course assignments. The workshop contributors supported this approach:

> 'I would love to study, like, you've been given a social problem and you go to investigate this as a group, as a class you would investigate this. And you'll have people who naturally within that who want to go an investigate it in certain ways and that teaches a lot about what do we understand about the world and, and how do we understand social problems, and how a thing can be analysed. And having those discussions and actually seeing where do these [approaches] come from, from your ontology? Your past? Your disciplinary perspective? How do you view the world?'
>
> Teaching Lead – Expert Workshops

From the landscape review, all DTP training leads reported increased investment in developing learning opportunities with businesses, from start-ups to banks and Governments; designed to encourage business acumen and entrepreneurial mindsets. People working in industry had been invited to facilitate training and there was widespread use of organisations like Skillfluence who specialise in bridging between academia and industry:

‘We promote collaboration with non-academic partners from the outset as part of the induction events … Training around development were devised to provide the skills necessary to achieve impact, such as policy internships, 'real-world challenge' events, a half day policy workshop with the Centre for Science and Policy Centre...For social science students [internships] can have very immediate and wide-ranging benefits to their studies as well as more lasting benefits in terms of confidence in their existing skill set’

DTP Training Lead - Survey

It is clear too that employers have valued working with students:

‘Employers generally value working with our PGRs, underlining the need for: ability to hone argument/writing ‘report-style’; also softer skills such as collaboration, team working.’

DTP Training Lead - Survey

Another spoke of their experience and what they continued to gain:

‘Former business colleagues from previous career stages are also helpful as mentors, guides, signposts...identify new data sources, suggest ways to analyse. Non-HEI contacts can really help.’

Recognised Researcher - Interviews

From more established scholars there was also recognition of the value of internships and other forms of collaborations, including an emphasis on the value of working with policy makers and third sector organisations:

‘We are also noticing students speaking more strategically about collaborative experiences or internships as learning spaces, which is great to see. We read the pedagogic literatures around industry needs. We involve industry/civil/voluntary sector partners in our training. We like to work with start-ups, they can be more receptive to us, and more attractive to students.’

DTP Training Lead - Survey

None of the established or leading researchers spoke of their own internships but two did reflect that this would be valuable for them to do at their current career stage.

The success of collaborations and internships may lie in the practicing of these skills alongside a new team of people, building confidence. It doesn’t seem that the interns are becoming better entrepreneurs or business leaders, so much as more confident researchers. On this basis, internships that are located in universities may usefully extend academic networks. Such placements need not be research focused but situated in professional services teams, exposing students to the leadership and management philosophies that ultimately structure academic practices.

To strengthen links to industry and to help students feel confident as they approach placements or jobs, it is recommended that communication and team-work skills are

embedded into ‘core’ training. Data-driven skills can be taught in ways that promote ‘finding out’ data as well as ‘how to share’ findings.

3.3 Advanced skills needs

The Guidelines state students should be competent across the core requirements. The evidence from the landscape review, interviews and workshops all indicate that competency is not achieved and that ‘core’ learning does not sufficiently ‘advance’ students’ capacity to manage, with confidence, data-driven skill learning. Students are not adequately prepared to manage their PhD research. In the context of this section, advanced is defined as skills and knowledge currently delivered after core training. Evidence presented supports the anchoring of digital, dissemination and wellbeing skills into core masters training, to provide an adequate springboard into further specialised training during the doctoral years.

3.3.1 Defining digital skills

Almost all doctoral candidates have a **data-driven digital skill gap**.[10] The evidence presented here will demonstrate that these are ‘must have’ skills and in demand across the research council domains and across industries. Digital skills have been termed the ‘fourth revolution’ as digital literacy is increasingly recognised as essential to industry, development, and employment (European Commission 2018a, 2020). There is evidence that digital skills should be reconsidered, from being a potentially useful skill to being a core skill for data-driven social science doctoral students. The OECD report suggests that digital skills should be taught at undergraduate level (2020: 21). Skills may be considered ‘soft’ (browsing digital literatures and filtering data) but are vital (data stewardship including data integrity, protecting personal data and reputation) for global citizenship and effective communication and collaboration (European Commission, 2018b). Even heavily computational disciplines (bio-medical informatics and physics for example) are likely to require additional training around safety, ethics and will benefit from up-to-date knowledge around coding strategies, or creating visualisations for example (OECD, 2020).

PhD students should have the opportunity to emerge from their studies with skills prized by the academy and by industry (Piatetsky, 2018). The big data and data analytics sector, requiring digital skills, make up the largest proportion of UK digital vacancies and is expected to experience the largest skills gap with 62% of companies requiring more digital capabilities (Mateos-Garcia et al, 2015). This is demonstrated in a secondary Nesta report (Windsor & Mateos-Garcia, 2015) which found that data-heavy businesses struggle to fill analyst vacancies. When it comes to digital data skills, industry is competing for (and if higher education sector leaders can promote the importance of) or co-operating with researchers with up-to-date skills. Industry values digital skills and are able to offer lucrative incentives for academics to move into the private sector and this is a global trend. This is particularly true for large tech companies that encourage researchers to publish

[10] While digital skills are core, there is also a need to increase specialised training provision in this area. This discussion is set out here in relation to advanced skills: Digital skills were ‘advanced’ at the time of publication of the Guidelines, but are now increasingly a ‘core’ gap

which used to be exclusive to universities (Royal Society, 2019). In this sense, we are at a tipping point where embedding digital skills into doctoral training will be an asset for our UK-based scholars, and not doing this will cause a deficit in the UK's data-driven skill set within a few years. In 2019, The Royal Society reported a 37% increase in the number of posts citing digital skills between 2013 (6.7 million) and 2017 (9.2 million) in the UK alone.

This section foregrounds evidence of the urgency to address this issue and considers what digital skills are needed in particular. This is useful to consider in the context of interdisciplinary learning (Campaign for Social Science, 2019), and for skills of value beyond the academy. Here skills needs are less about 'data' and more about 'digital' approaches to research design and analysis (Moltzou, 2019). Recent debates in the literature bring forward working with new forms of data, and this can be captured by the ubiquitous term 'digital' (Halford & Savage, 2017; Lewthwaite & Jamieson, 2019; Social Media Research Group, 2016; Tinati et al, 2014).

Digital skills are not well defined in the literature. There is confusion in how digital skills relate to data analytics, computational skills, quantitative skills, social media skills, and big data. This definition is broad. As the report focuses more on findings and draws evidence from our reviews and primary data, we will take care to ensure that the way others use the term 'digital' is made clear. Recognised researchers at the start of their careers reported in the interviews, that digital skills would be must-have skills for social researchers in the future and this view was supported by established and leading researchers. There was also a great deal of frustration around the definition of digital skills and how it mapped on to, or was distinct from computational social science, data analytics, quantitative methods, or data science. The lack of consensus around definitions is confusing:

'Really confusing, especially when coming from a qual[itative] background. Really not clear what they mean and whether opportunities bearing these labels are relevant/accessible to me. Some titles are immediately off-putting (data analytics) - but may be relevant and useful and accessible.'

Recognised Researcher - Interviews

And from three different participants:

'Terminology has not served us well. AQMeN did some work on this - looking at the overlap between data science, computing science and other areas with social data science in the centre of a Venn diagram. That seems right. Digital social science may be an even better term. Almost certain that digital humanities have leapt ahead of social science in re-positioning their discipline and taking on new training and development.'

Established or Leading Researcher - Interviews

'[Terms like Digital Skills] It is confusing and sometimes about marketing.'

Established or Leading Researcher - Interviews

‘It’s a mess. The lack of clarity is possibly deliberate.’

Wider stakeholder - Interviews

Indeed the 2021 analysis of training needs data within Scotland’s DTP revealed a strong overlap in demand for training:

‘Students may be confused about the differences of some key terms (how are digital methods different from computational methods? And from analysing social media?). Of students identifying a training requirement in quantitative digital methods: 81% also requested training in social media; 70.2% requested qualitative digital and big data; 61.9% requested big data; 57.2% requested programming; 54.7% requested computational social science and 53.6% requested machine learning.’

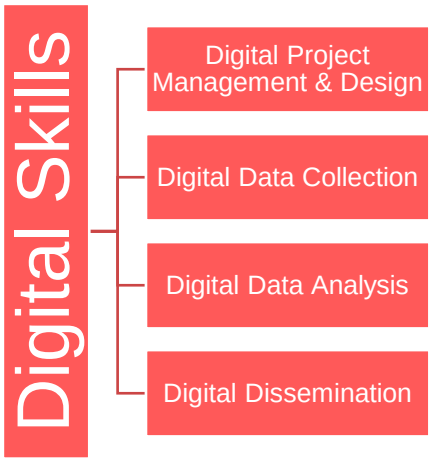
Excerpt from TNA analysis for the Scottish Graduate School of Social Sciences, 2021

Of particular note here is the large overlap then between students wanting ‘quantitative digital methods’ and ‘qualitative digital and big data’: at 70.2%, there is a strong argument that the division between quantitative and qualitative is an unnecessary and less than useful, construct. Requests courses at the doctoral stage may suggest that demand is for specialised training (advanced towards expertise) though students learning core data-driven skills will find it useful to see how their learning could build towards courses in AI and machine learning. Moving forward, training providers need to be transparent about what is being delivered: social statistics remain the same, even in a digital context, and does not necessarily become ‘data science’.

To provide some clarity, our definition of digital skills articulates with the life-course of a project (see Figure 7, below). The understanding of this term, and its deconstruction into four key areas is informed by the evidence collected for this project and key literature and draws from the work of Zauner (2021), an intern with the Scottish Graduate School of Social Sciences.

Digital project management relates to digital research responsibilities and includes access to equipment and software, data safety and research integrity training. Critically, it includes digital ethics which has seen a growth in engagement as many qualitative researchers in particular, have re-imagined their projects digitally during the pandemic. It extends to include digital copyright and health and well-being safeguarding with digital data/platforms/communication. Here skills around management software (such as Zotero, Mendeley and Endnote) and knowledge of communication software (such as Zoom, Teams, Trello, Mentimeter and Kahoot) are recognised as valuable. In this field there is also growing recognition that there are gaps in knowledge, for example, some postgraduate students are used to working with handheld devices but need training in how to work with desktop computers. It can be difficult to source training in less fashionable software such as Excel and operations such as producing Gantt charts (Zauner, 2021).

Figure 7 Working definition of digital skills



Source: Authors' own elaboration

Digital data collection refers either to the collection of data already existing in digital worlds (for example, web scraping twitter or harvesting big data) or the production of new data gathered through digital platforms (for example digital ethnography or developing a survey for use on Facebook). There is recognition that data understood as 'big data' is not exclusively quantitative, rather the rich textual contribution of twitter (for example) and rich visualisations of Instagram (for example) or performances of tik tok (for example) require curation and analysis approaches consistent with a qualitative approach. Traditional qualitative approaches such as archiving have particular relevance to digital data collection.

Digital data analysis includes the use of software to perform analysis (for example, NVivo, Maxqda, R, Python, see Dowling and Wilson, 2017) as well as the application of analytical approaches that work with non-digital data: for example, the application of discourse analysis to data drawn from twitter; the use of phenomenology to engage with the lifeworlds developed through Instagram or structural equation modelling to explore the factors that enable some people to emerge as 'you tube stars'.

The fourth element of our definition, digital dissemination, considers the skills needed to communicate and disseminate knowledge; towards knowledge exchange and potentially impact. From running a twitter campaign (Dowling and Wilson, 2017; Rainford, 2016), potentially linked to improving metrics around peer-reviewed published articles; to developing an app that transforms research into a provision. This definition enables the deconstruction of two dichotomies that too often restrict academics: academy/industry and academy/general public. Further digital dissemination can be understood as the creation of virtual communities that enable students/colleagues to share the joys and frustrations of research careers, potentially reducing isolation (Bennett and Folley, 2014, Rainford, 2016) and potentially increasing competition/feelings of failure.

To help clarify the data-driven skill gap, digital skills will be examined separately from open data requirements. The section closes by considering how to build capacity in this field.

Understanding what researchers, who are postdoctoral and beyond, require in terms of data management and digital skills is one way of working-back to think about post-graduate data- driven skills training, however, the literature is limited. Jisc's (2018) Digital Capabilities: The Six Elements has informed our definition of digital skills and covers:

- Digital Identity and Wellbeing
- Information Communication Technology Proficiency
- Information, data, and media literacies
- Digital creation innovation and scholarship
- Digital learning and self-development and
- Communication, collaboration, and participation.

A 2013 NCRM report from Roberts et al. used a series of seminars to explore digital methods (i.e., broad data skills involving digital tools) capacity building for researchers across career stages, from PhD students to experienced academics. The report emphasises the need for social science to adapt and embrace digital methods to ensure the field remains culturally significant and able to respond to relevant contemporary issues. Researchers involved in the seminars highlighted the importance of early career researchers receiving appropriate training, resources, and assistance to enable them to reconcile new data approaches (e.g., methods for working with social media data) with existing methodological traditions. The report also raised the value of encouraging interdisciplinary work to make best use of new digital methods, preferably embedding this at PhD level to prepare students. Data archiving and ethics explicitly relating to digital data, were noted as potential challenges requiring careful attention and training, in addition to a need for new data cleaning methods and algorithms to manage Big Data. Bone et al. (2016) similarly argue that there is a need for new skills in social sciences to enable researchers to work with online data, but they also note the challenging 'lag' between technological advancements and both methodological possibilities and the ethical policies of institutions.

A 2020 review (Jisc, 2020) revealed that training around using software and data integrity were available to post-graduate students but fewer (43%) reported access to digital health and well- being training. Zauner (2021) reported that developing digital skills was the third highest priority for doctoral candidates after specialist methodologies and project management. Domain A1 (knowledge base) of the Vitae Researcher Development Framework (2010) includes 'information literacy and management', which includes skills such as using IT (e.g., spreadsheets and databases) and knowledge of data security. IT literacy and digital technology skills are also included within 'academic literacy and numeracy'. While the Framework has value as the primary UK-wide template for researcher development, it is (of course) not specific to social science. In addition, the framework does not appear to be used routinely beyond the PhD stage. For example, in the Vitae (2019a) Careers in Research Online Survey, only 11.1% of researchers (postdoctoral and beyond) agreed or strongly agreed that they used the Vitae framework to support CPD activity, and 43.7% strongly disagreed (n=6422). As the framework was created in 2010, it also lacks detailed recommendations on what constitutes essential digital skills in contemporary social science research or for industry.

While not specifically related to a method or methodology, there has been a similar call from the European Union to increase social science capacities to improve social and economic development with investment in digital skills within doctoral training (Kottmann, 2011). Elsewhere doctoral candidates are accessing advanced digital data-driven skills learning and some shift in the structure of UK training will be needed to keep pace with this, particularly if a new approach fully considers equality, diversity, and inclusion. Boud & Lee (2008) writing from the U.S. capture a shift, from focusing on the quality of research amongst doctoral candidates to thinking about the quality of training they received. Triggered by the availability of big data, this shift called for greater understanding and use of quantitative methods specifically (National Science Foundation, 2011) in the social sciences. The impact of this has filtered down to undergraduate courses in the U.S. and students are encouraged to take extra courses and credits, often during their summer breaks (a trend noted in science and technology disciplines by Kilburn, Nind & Wiles, 2014b). This then places pressure on the individual to do extra to address a structural issue (lack of capacity for advanced digital methods amongst social scientists) and is likely to lead to inequalities (some people need to work and earn during the summer months, people with impairments may need to rest, people with dependents may be needed at home, people who can't afford course fees can't take advantage).

Demand for digital skills rose sharply during the pandemic and was reported as a growth area by most DTPs as doctoral candidates reinvented their research projects to continue to collect primary data and avoid face-to-face activities:

> 'This last year demand for digital mining has grown and I suspect it would have done without Covid. The move to online learning was managed swiftly with no decline in the amount of training offered, but with an increase in digital skills learning (using smartphones within research; data mining and social media; institutional ethnography online, for example).'
>
> DTP Training Lead - Survey

The rapid rise in digitalisation was happening before the pandemic and has accelerated as the research workforce moved to online working. Mons' (2020) estimation that 5% of researchers need digital skills sufficiently advanced to deliver data stewardship is, due to the acceleration of the pandemic, already an under-estimation. University research methods courses in the UK championed digital workarounds for primary data collection and while some students and scholars will be keen to return to traditional approaches, there is an opportunity to build on this new wave of capacity around digital methods.

Most participants in the interviews cited machine learning and skills around artificial intelligence (AI) as essential for future data-driven researchers but gave little explanation of why or how traditional approaches were inadequate. Thus, machine learning and AI are additional to traditional approaches, and how they might be incorporated, along with digital skills generally, becomes the focus of this section.

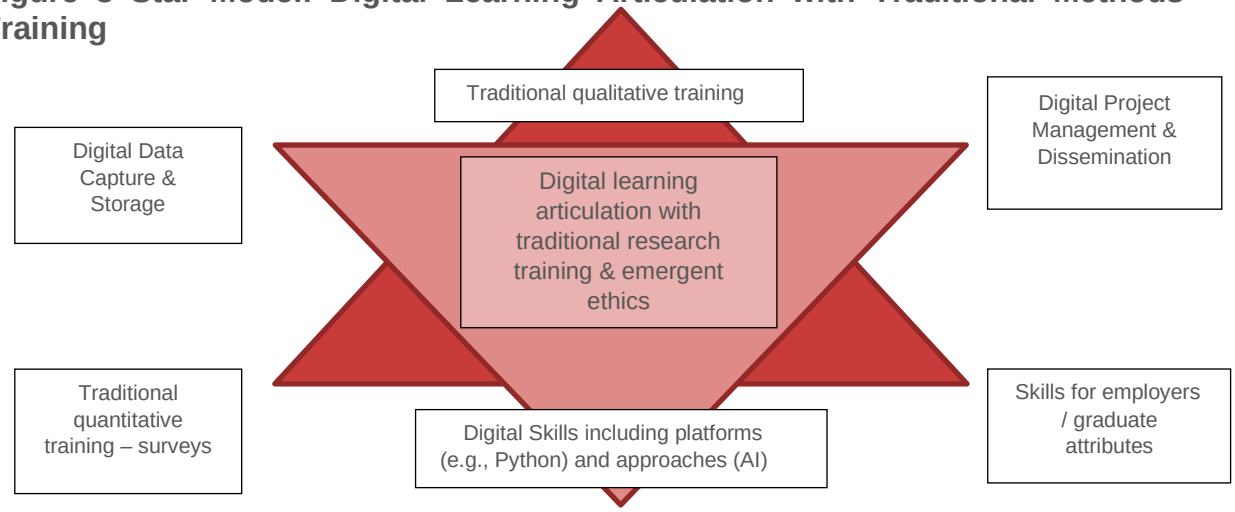
Given our findings emphasise the value of digital skill learning, it is useful to consider how digital data-driven skills could articulate with traditional methods. While digital literacies and skills have been recognised as integral to doing research and a

necessary part of the post-graduate students' toolkit, (Henderson et al, 2016; Littlejohn et al, 2012), how it should be embedded in research learning experiences is less well explored (Fielding et al, 2019). Here we consider how digital data-driven skills could articulate with traditional presentations and present our 'Star Model' (Figure 8, below) to capture these ideas.

The 'Star Model' (below) presents two triangles that work together to produce a 'star'. One triangle symbolises traditional data-driven skills learning, the other represents digital skills. The traditional training triangle has three points: qualitative, quantitative, and transferable skills and is influenced by the Guidelines and evidence presented in this report, of how the Guidelines have been applied by HEIs. Using text or visual data is well captured by the Guidelines and includes skills such as interviewing, observing, transcribing, and running focus groups. Students engage with introductions to alternative epistemologies towards advanced analysis, for example, training in discourse analysis, phenomenology, and process tracing. This training is broad-based, and many doctoral students would seek advanced training in approaches to analysis, often from specialist experts, for example ethnography training from anthropologists. Such training is valued and valuable, and expertise gained is used to produce world-leading, cutting-edge research, and should not be lost.

Quantitative training is another corner in the traditional triangle and is also well captured in the Guidelines. It covers data cleaning and linking, building expertise using a software, and moving through chi-square, t-tests to simple linear regression. Stronger courses will focus on interpretation of findings, but generally advanced training is provided through distinct training opportunities and the landscape review revealed that specialist training around social network analysis, structural equation modelling, difference of differences and cluster analysis 'sell-out' quickly and are well attended. Such approaches are key to meeting society's grand challenges and should not be lost.

Figure 8 Star Model: Digital Learning Articulation with Traditional Methods Training



Source: Authors' own elaboration

Many universities have tried to capture the skills employers are looking for, within a concept of 'transferable skills' or 'researcher development' and can be found within

DTP advanced training, representing the final point of the traditional triangle. The skills offered here tend to relate to professional practices and examples include time management and report writing. There is also training delivered on software, dominated by training in Excel (following feedback from industry) and in recent years training in Python. The emergence of digital skills allow work to build on these foundations but the popularity of courses such as excel, tell us that not all industrial partners are evolving in line with technological opportunities and for this reason, this traditional training will still be needed.

In an early proposition in the sense-checking workshops there were protests from doctoral candidates and experts because surveys and secondary data analysis were not explicitly included in new propositions of core training:

> 'I was really fascinated with the secondary data analysis and that's … I would like people to be more informed about it because I was fascinated about how much is stored somewhere in the depository, and [students] will never hear about it!'
>
> Rehana – Workshop with Doctoral Candidates

Another student agreed:

> 'For me, I found it super useful going int the UK data archive because you can actually find something which might also be relevant conceptually, in terms of your own research, that you can use them to practice the skills on data that might be similar to data that you're going to collect at some point.'
>
> Kait – Workshop with Doctoral Candidates

It is perhaps because of the usefulness of the traditional approach that digital skills are taking some time to appear in core training provision. Meeting the 2015 ESRC Guidelines in full requires 60-credits of training, and there is little capacity to increase training without diminishing disciplinary learning.

There is though, a danger in continuing to structure training on the traditional approach. As the Star Model shows, digital skills sit upon traditional approaches, but not exactly in line with them. The digital triangle includes data capture, curation and storage, advanced analysis and project management which capture the skill gaps identified earlier in this chapter.

Digital data capture is one 'point' of the digital triangle and includes sourcing or mining data produced by apps, big data, social media data including Facebook, twitter, and Instagram. Included here are data-driven skills around curation and safe storage of digital data.

All social scientists need training in advanced analysis, and this could include training in coding and new software such as R, Stata, Python to the level where complex and sophisticated visualisations can be produced. This point on the digital training could also include training that builds directly from approaches traditionally considered quantitative such as machine learning and the aligned field of artificial intelligence. It

may be that a traditional qualitative learner is keen to enter this field, but they fundamentally need the traditional quantitative skills in order to access it.

The final point on the digital triangle relates to project management and includes forming communication or dissemination strategies that exploit social media platforms. This could include twitter campaigns or developing an app for an industrial partner. Digital dissemination skills are placed on the Star Model between traditional qualitative and traditional skills for employers it is fair to say that students who use numeric data and skills may also have an interest and much to contribute to this field. It is placed where it is however, because of its heavy use of words and symbols as a communication device and thus draws stronger links to the narrative foundations of qualitative approaches.

In developing an evolution to post-graduate education, and equally to the educational opportunities academic and non-academic people have access to, it is clear that traditional approaches are not outdated or redundant.

As a final point, ethics should not be situated in any one area, but explicitly engaged with across this Star Model of training. As there is still work to do to thoroughly embed ethics in quantitative training, it is recommended that ethics training be foregrounded as essential as these new training fields emerge.

Including digital skills in core learning could effectively join up traditionally opposed quant/qual learning spaces. Further this could add value towards interrogation of ontological issues (is social media reality? How is this reality socially constructed?) and to issues of quality relating back to research integrity (exploring credibility, validity, reliability, reproducibility, transparency, and a consideration of how these 'standards' relate to the data, each other and can be defended). This will advance conceptual understanding and data literacy and give space for students to 'advance' in a method they plan to use in their research, giving them greater facility to build confidence in their specialism.

There is demand for digital data-driven skills. Recognised, established and leading researchers who used traditionally quantitative approaches reported that machine learning and AI as well as R (and to a lesser degree Python) were essential and some established scholars had sought training in these fields, suggesting it may not be something they felt confident learning themselves. The 'newness' of the approaches meant they may have felt able to admit they needed training; or that not being trained was not an option.

Early career scholars and the reports from DTPs around student requests, suggest that early career scholars think similarly about digital skills that align well with traditional qualitative approaches, and are seeking training. Textual analysis of social media data, digital ethnography and analysing visual data were all reported as areas where demand for training was growing. None of the established or leading researchers reported wanting, needing, or having done training in these forms of digital research. Further focus on this area will reveal then if digital skills will be 'must have' for researchers using forms of qualitative data.

Figure 9 Landscape Review – Examples of UK investments in digital skill training

UK investment in digital data-driven skill training		
Digital Economy Crucible, Swansea University The Crucible is an EPSRC-funded leadership development programme for early career researchers (1-8 years post-PhD) from all disciplines and from all UK universities to develop skills surrounding digital economy research. This initiative is part of CHERISH Digital Economy Research Centre at the University. It is an opportunity to develop career- altering skills surrounding digital economy research, network, and collaborate to achieve societal impact and economic gain. The programme has nurtured more than 100 researchers from 35 different UK universities.	**Cambridge Centre for Data-driven Discovery (C2D3)** Based at the University of Cambridge, the Centre works with groups of researchers, departments and other interdisciplinary research centres, strategic research initiatives and networks to promote collaboration, share ideas and scope out key research questions in data science and to support the funding proposals. Research workshops are an important part of the centre's programme. These range from internal half-day meetings to multi- day conferences with external speakers and delegates.	**Humanities & Data Science Group, Alan Turing Institute** The group aims to strengthen relationships and build collaborations at the intersection between data science and digital humanities. It was set up to highlight the role that can be played by the institute in the area of digital humanities by demonstrating that data science research can answer questions relevant to the humanities and vice versa. The group's activities include meetings, workshops and research projects with emphasis on AI and machine learning.

While DTPs and the social sciences are awakening to demand for digital skills, there are examples of investment in training opportunities and Figure 9, above, outlines three examples from the UK. These do not focus exclusively on doctoral researchers, and so illustrate a wider context of targeted investment in later-career stages where relevant.

There is growing expertise in digital skills in the UK and significant investment in specialist training. This investment is not exclusive to the social sciences or targeted at doctoral researchers. As per the OECD (2020) report, even if social scientists are not wanting to use digital data, they are still likely to need some digital skills, at least around communication and data storage. This moves the discussion towards open data.

3.3.2 Knowledge and awareness of open data

Social scientists need greater knowledge and awareness of **open data and open access: this should therefore be considered a data-driven skill gap**. Open science aims to enhance access to data-driven research by investing in the infrastructure that supports information and communication technologies and tools. As reported in the OECD (2020) report, the use of digital data and an open science approach increases transparency and can more easily facilitate reproducibility. For these reasons, the scientific breakthroughs and new knowledge gained is equally open to scrutiny and can thus be verified to produce rigorous and robust evidence. The OECD report argues that though responses to the pandemic have accelerated the use of digital data and reinforced the importance of data-driven open science, it has also exposed gaps in the skill sets of researchers and an infrastructure that needs greater investment. A 2016 report demonstrated below average numeracy in England, relative to other OECD countries, defined in the report as difficulties with basic quantitative reasoning (OECD, 2016). As well as upskilling in digital research design and data capture, investment is needed in curation and storage, new pathways to publication and rules around intellectual property, research integrity and reuse of data:

> 'Digital technologies and access to data are driving change, but human digital capacity and skills are likely to be the critical determinant of scientific success in the future.' (OECD, 2020: 7)

This issue is a concern to doctoral candidates:

> 'How cool would it be if you can teach people how to get their own data? There is interesting data out there, but sometimes, it's not easy to get those data. It's not that they are all on UK archives, or just like sitting somewhere. We have to do quite a lot of work with administrative data, you need to scrape those data from websites or you need to talk to people to get a data sharing agreement. So, it's not just about 'oh let's do a fancy cool thing about social medial data' it's more about empowering people to look for their own data. It's important to collect the right data in the first place. You can do all kinds of fancy modelling, we can do the most advanced statistical techniques, but if your data is shaped like … Need to know what is good data, and what is clean data.'
>
> Christian – Workshop with Doctoral Candidates

Established researchers also noted the need to educate students on contemporary challenges in curating data:

> 'It's copyright. How do you use copyright images that range from maps to images of anything? And on the other hand, how do you set the current copyright for whatever open research you're going to create? There is a lot of misconception, I think still, there is very much a shared idea that if you can find something online, it's free to use.'
>
> Training Lead - Expert Workshop

Van den Eynden and Corti (2017) discuss the skills needed in data publishing in social sciences across the researcher lifespan, for example in the UK Data Archive and the ReShare self-deposit data repository. The authors argue that being able to self-publish data is now an essential skill and is likely to become an increasingly important task. Many of their recommendations align with research integrity issues and it is clear that doctoral students will be under much greater pressure than their supervisors were, to make transparent their research design and share their data. For example, the published data can be used to support learning and teaching via secondary data analysis. Where data processing procedures were formerly carried out by in- house data processers prior to digital archiving, this task now falls to the researchers themselves. Researchers may be required to undertake tasks such as checking data files for inconsistencies, carrying out data enhancement, generating the data in multiple file formats for preservation and dissemination, and generating code lists. Each task requires a particular set of skills; for example, data enhancement may include converting interview transcripts to standard UK Data Service templates or converting data files to preservation formats. The authors highlight that the skills needed by researchers to use data repositories and prepare data for publication require training, which should ideally be embedded in undergraduate and postgraduate research methods training. This further supports the idea of project management and dissemination training being given equal weight to data collection and analysis.

3.3.3 Capacity issues

As many established academics focused capacity on teaching during the pandemic, their own research has been impacted. Thus, we may have many who have supervised digital research, and who may have made the turn to using digital approaches in their own research but who have not had opportunities to engage themselves in digital research or training. This may also be due to the 'deficit approach' (Scott Jones & Goldring, 2015) that describes universities only investing in staff training if someone has no skill, but rarely investing in enhancing skill or to learning something new (such as teaching with digital data). This argument was supported by our workshop with doctoral students: half do not receive data-driven skills training from their supervisors. From the interviews, we learned that recognised researchers were employed for their digital skills and anticipated that they would receive further training from PIs and Co-Is. A gap emerged as PIs and Co-Is could not deliver and this was most linked to emerging software like R and digital skills:

> 'For older staff, they do not have the need to learn the same skills. They often have research students to do the more contemporary things. Coding wasn't a big thing, so the people who have lectureships or permanent jobs do not necessarily need to learn those new skills. They can benefit from the skills of younger staff. As senior staff you provide the more theoretical support, rather than skill.'
>
> Recognised Researcher - Interviews

This recognition of the pattern was endorsed by many established and leading researchers. Some were critical of this trend:

‘Older stage career people often ignore things that they don't understand and are inaccessible to new methods.’

Recognised Researcher - Interviews

This gap in skills knowledge presumably leads to gaps in research integrity, with junior researchers still awakening to the consequences of their micro-decisions around data collection, data management and analysis and established scholars missing the micro-detail required to focus on standards. To be clear, most participants regretted this state of play:

‘There is a reality in academia that younger researchers will be more up to date with techniques. Many of the more advanced researchers will rely on post-docs or others to do this kind of data work. They wouldn't do that under any other circumstances. It would be good if they had the skills to provide support and understanding. Students will write the programme, but it would be good to have the supervisor provide feedback. But many do not have that kind of experience to support students.’

Established or Leading Researcher - Interviews

Potentially behind this structured approach is fear of admitting that there are gaps in knowledge, or fear that they won’t be able to learn a new method easily. Some frustration was evident amongst the early career participants:

‘One of the most obvious challenges is that there is a pretence, a cloud of dishonesty and [established researchers] are afraid to say they do not understand something. People do not want to ask questions and look foolish. There needs to be a way to drop the façade. People need to be able to admit what they do or do not know.’

Recognised Researcher – Interviews

While some felt that incentivisation should include training requirements in grants and funding applications, where leading scholars had to state the strength and limit of their expertise and how they would train the junior members of the team; others resisted this. Time pressures, as a related issue, are discussed in section 4.2.2.

Early career scholars are challenged by an inequality in skillset and an expectation that they can manage, without support or guidance, to deliver. There is a tension for early career scholars too, where the available and fragmented research positions are not able to contribute notably to career development. Another early career participant made a series of significant points:

‘Perhaps funders should make skills updates a requirement. That would concentrate minds. Skills development needs to be a part of career development - some academics in [principal investigator] roles might have neglected this and it does not help newer researchers when this is the case.’

Recognised Researcher - Interviews

Closing the digital gap requires a culture shift of the early career scholar teaching the established, and this provides doctoral students with rare opportunities for sector-impacting leadership. Unless this data-driven skill gap is closed, there may be an increasing divide within UK social sciences including UK doctoral candidates and those training overseas. Further, there is evidence that the digital divide exists across all research council domains, and in many competing nations (OECD, 2020 cite US/bio-science, Barone, Williams & Micklos, 2017; Australia and bio-science, Lonie & Francis, 2017; and Japan/Information Sciences, Kurata, Matsubayashi & Takeda, 2017). Universities still do not place sufficient value on teaching contributions however and a post-doctoral student would undoubtedly earn greater recognition and reward investing their time in peer-reviewed outputs than developing and delivering high- quality and transformative training.

Investing in digital learning experiences and the early-career scholars that can best produce them, could fuel this culture change towards greater recognition of quality teaching (DORA, 2012). The ESRC could consider an intervention. One solution might be requiring post- doctorates with ESRC funding to convert their cutting-edge methods skills into new training as part of their commitment. Alternatively, principal investigators of major research investments could be invited to develop dissemination pathways straight to doctoral researchers that again, showcase innovative approaches to data-driven skills.

It is worth noting that ‘ECR need training budgets too’ (Early career scholar - Interviews) a view that was supported by established and leading researchers:

‘Methods keep moving on. I don't do my own numbers anymore, have technically skilled PhDs and postdocs. Last time I did a refresher was 2017. Don't know if there's anything that takes people at mid-career to start thinking about quant[itative] or what the incentives are. Or refreshers on new techniques for people with initial exposure to tool up. Not sure what the demand would be either - starting again is a big ask for mid- career researchers. I don't know how to use R for example.’

Established or Leading Researcher - Interviews

This leads to a challenging consideration: if leading researchers with significant research funding and awards don’t have the skills to deliver their projects, who is teaching these skills to the next generation? Are trainers acknowledged for essentially closing the knowledge gap and driving forward the UKs capacity to lead and innovate? Several established scholars wanted to upskill and be part of the solution, and this will be discussed further in section 4.2.2 around knowledge and awareness of specialist training.

3.3.4 Dissemination skills

Dissemination strategies were identified as a data-driven skill gap and an area identified by doctoral candidates particularly. Data-driven skills, capable of sector-changing rigorous and robust evidence is complimented by the skills to reach

multiple audiences and disseminate effectively. These skills are vital for doctoral candidates seeking to build a career in industry or the academy.

Doctoral candidates were keen to develop writing skills to reach academic audiences.

The landscape review revealed that literature reviews were considered a core research skill overseas, particularly in German master's provision where the dissertation was often devoted to a systematic literature review. A contributor to the workshops reported that this form of master's dissertation was restrictive in one sense, discouraging the practice of data-driven skills. The advantage of this approach was that around 80% of students reportedly published (Training Lead – Expert Workshops) from their master's work, gaining significant skills around dissemination.

Doctoral candidates also sought training in writing for different audiences:

> 'I think we need to learn how to write journal articles. Even if you don't want to stay in academia, even if you're working in an organisation, you need to know how to write.'
>
> Amy – Workshops with Doctoral Candidates

One example of how dissemination could be incorporated into a master's curriculum, or used as a doctoral intervention, was developed by Nairn (2020), focusing specifically on peer-review article writing. In discussing collective editing pedagogy, Nairn, (2020) introduced the value of writing workshops that unpack the labour involved in writing well. One student observed:

> '… seeing how we can actually tear apart an entire paragraph from different people and then put it together as a group, I think showed how much work is needed to really get something from an idea into a publishable version (female, post-PhD)' (Nairn, 2020: 889).

The teamwork involved in elevating writing is demonstrable in this quote. Moving writing from an individual into a community practice permits, for those who need it, circumvention of the supervisor as the people traditionally charged with outlining the 'rules' of writing for publication (Aitchison & Pare, 2012; Kumar & Aitchison, 2017).

There was significant support amongst our participants for using a peer-learning network to gain experience of the peer-reviewing process:

> 'If we were given a platform, where an ESRC student wants to send out the paper for publication and they just put it [on the platform] and say, if anyone has time/ or wants to have a look and just give constructive feedback, it would be highly appreciated and it will again, help both ways, because I think it's helpful for us to receive feedback on our work, but also for us to give feedback to someone else.'
>
> Amy – Workshop with Doctoral Candidates

This comment followed a much earlier conversation about the emotional resilience required for peer-review publishing:

‘Perhaps we could also work at trying to change the ‘tradition’ it seems of reviewers going beyond constructive criticism and being downright mean! It concerns me that my supervisors have had to chat to me about emotional resilience around what people are going to say about my research and that some of it may not actually be helpful but hurtful.’

Lucy – Workshop with Doctoral Candidates

This is a critical point. Barriers to sharing work limit the potential of UK social sciences to learn and develop. While peer review articles may still dominate as a ‘gold standard’, the promotion of this work via social media can significantly increase metrics (downloads of the article, which in turn push up the impact factor of the journal) which can be translated into promotion, recognition and reward. Increasingly then career development hinges upon public engagement:

‘In terms of research dissemination, I think the thing that I would find more valuable is disseminating research outside of academia. For example, even through Twitter. I have little confidence in posting anything in Twitter, but I am aware that Twitter tends to be a platform and a very useful space for academic conversation … I feel like, it seems silly to ask for help writing like a simpler piece of writing, but it is actually quite complicated when you boil down such complex topics in that way.’

Clara – Student Workshop

Knowledge exchange and impact is understood broadly as raising awareness beyond the university and harnessing social medias will help scholars establish the foundations of such work (Thrift, 2008; Universities UK, 2009). Incorporating such knowledge into training will give student greater control over their visibility and in turn their career. Doctoral candidates with limited time, are increasingly concerned with making themselves employable and confused about how best to invest their time:

‘We need to prioritize where to put our energy. I’m, at the moment, struggling with if I need to prioritize writing a peer-reviewed article or develop an interactive website with all the PhD data that I have created. Which one do I go for? Which one is the best for my career? It seems the peer-reviewed article is much more valued in academia than an R-Shiny dashboard but, I would find the R-Shiny dashboard much more exciting because it’s a way for me to disseminate my research beyond academia as well as being important for academia.’

Amelie – Workshop with Doctoral Candidates

The significance of peer-reviewed articles is founded, in part on the Research Excellence Framework (REF), which is a gateway to reputation and resourcing for UK HEIs. REF’s definitions of ‘what counts’ do not extend to open access data, though critically, this is required to fuel future research. Responding to the propositions delivered to the workshops, one contributor argued:

> ‘The one thing I saw missing from the [propositions] was ‘open’. I didn’t see anything there around making sure that resources are published as open educational resources. Particularly given the limitations on all our times, we want to make sure that courses can be reused and contributed to across the field, particularly when we’ve got new subjects coming up. Openly available data that can be used, is I think the big thing that we see on our list of questions [from educators]: which is the best data set of teaching techniques?’
>
> Training Lead – Expert Workshop

Doctoral candidates recognise the need for extending what counts as a publication, and this was shared as a principle by the established scholars who contributed to the workshops. It may not be understood by everyone in the sector and the REF could influence the culture change needed to recognise an extended portfolio of valued contributions from academics.

Doctoral candidates are aware of and feel pressure to perform a range of dissemination activities. Training that helps them build a strategy will aid them to make their data-driven work visible.

3.3.5 Well-being and emotions

This final section relates to ‘being a researcher’ and identifies **researcher safety, particularly around preventing emotional harm, as a data-driven skill gap**. Here it is not the skills *per se* that doctoral candidates seek training on, but how to apply them in the field. This may not be evident to data-driven skills learners that are used to classroom-based teaching. It may be more useful for advanced researchers who are doing their PhD fieldwork. Anchoring some learning in ‘core’ courses will help students know what training to look for as they advance in their fieldwork and need to find specialist training. Engaging with data is rarely neutral and developing skills to manage the emotional labour, safeguards researchers.

Researcher safety and well-being came through strongly as a theme in the DTP survey and early career interviews. From one of the doctoral candidates at a workshop:

> ‘The one thing that I really wish there was more of was looking at ethics and looking at doing sensitive research, because I think the ESRC really pushes this idea that they want to have students that are doing things which are going to make an impact and like a lot of big social issues, they want proposals and they include those but then you have this cohort who are all researching really sensitive and hard hitting and like, I don't know, just it sometimes … actually quite horrible topics that are really important but it just doesn't feel like there's any preparation or support for that, and you know, some people might have backgrounds in and working with people who have some of the issues or in the situations that they're researching. There's so many mistakes that could be made and that could really impact both the researcher and the participants and this isn't even

like an advanced training module or anything on ethics or sensitive research.'

Lucy – Workshop with Doctoral Candidates

Most DTPs had responded to requests for training around well-being and gave some insight into why this may be needed now. Apart from specific resources around the pandemic, most well- being training responded to emotionally difficult research work that dealt with sensitive issues and/or research with people facing (often intersectional) social injustices. There was strong evidence of cross DTP initiatives (often between research councils) and with academics working closely with researcher development professionals and third sector organisations that promoted positive mental health. Therefore, this issue is sector-wide. It may be more pertinent to social scientists because of the emphasis on lived experience and themes relating to structural, political and embodied violence.

This is not an area that is captured in the 2015 Guidelines and a revision might include more explicit reference to researcher safeguarding and well-being. That supervisors don't seem to have been a significant source of support here, (according to interviews with recognised researchers and the workshops with doctoral candidates) and that established and leading researchers were less likely to discuss this issue suggests that projects or PhDs may have changed. The environment itself is more pressurised and so more emotionally demanding. Doctoral candidates are concerned that their investment in research may not result in employment. This is an increasingly urgent issue for all, particularly for non-traditional scholars who do not have the social and economic safety nets of others with affluent or understanding family. There is also a rise in such students having lived experience of the social injustices they research which in turn amplifies the emotional experience of their research. In any case, researcher safety, safeguarding and preventing emotional harm should now be understood as a fundamental aspect of project management and recognised through core training and re- visited as a doctoral intervention to support students advancing through their fieldwork.

3.4 Appraisal of the ESRC Guidelines: The gap between current postgraduate research and development guidelines and identified skills needs.

The Guidelines, produced in 2015 were a sound response to Savage and Burrows' (2007) critique that postgraduate students lacked a sufficiently broad-based capacity in, (Biesta, Allan & Edwards, 2011) and knowledge of methods. They argued the broad-based approach would better permit defence of methods selected (a common theme in PhD methodology chapters and interrogated during the viva, the PhD oral examination) and allow critique of methods they had not used in their doctoral work. The UK has used this model of core methods that aim to deliver literacy and offer a broad-based education, for the last 10 years. As stated, all ESRC funded studentships require core training to be completed. This research has demonstrated that while broad-based skill learning happens, the conceptual understanding is not sufficiently established.

There are a number of gaps identified through our five phases of inquiry, in the current way that core training is delivered in relation to data-driven skills. Specifically it is advised that the Guidelines evolve to present the 4 pillars of data-driven skill (see

Executive Summary, Figure 1) and ensure transferable skills such as project management and dissemination are explicitly included in core training. It is also advised that as far as possible, new training incorporates as many digital examples into learning particularly around digital ethics, safeguarding, data mining and data storage (see section 3.3.1 and Figure 7). The Guidelines should also make it clear that a broad-based approach relates specifically to conceptual knowledge and data literacy. Skill development may be incorporated, but it is not the primary aim of the broad- based approach. Students are not expected to be 'skilled' across all the areas required by the Guidelines, but to have sufficient grounding to knowledgably pursue specialised training.

The Guidelines should incorporate elements from the General Research and Transferable Skills Training (ESRC, 2015: 12-15) into their expectations for Core Research Methods Training (ESRC, 2015: 7-10). Included in the 'research methods' learning outcomes in the current Guidelines, data management is defined as including skills such as cleaning and preparing data for analysis, coding and manipulating data, securing data storage, preparing data for archiving and dissemination, and safe data disposal. However, there is some degree of overlap with what are termed 'general and transferable research skills', which the Guidelines suggests should be available during the doctoral research years rather than as part of core learning completed during the masters' year. These skills include computing skills, understanding ethical issues, engaging with users and maximising impact (i.e., dissemination and knowledge exchange), and understanding research and intellectual property rights. Technology has increasingly become an inseparable part of research (from word processing, and publishing to secure data storage), and the digital skills relating to data management are central to conducting data-driven research. Whilst the Covid-19 pandemic has accelerated the need for digital fluency around project management, data collection and dissemination, there is no clear evidence to help understand how the degree of digital skills required may vary depending on a researcher's career stage, specialist area, and research methods.

It will be useful to re-frame the General Research and Transferable Skills Training as a Doctoral Education Framework to underpin the need for continued professional development during the doctoral years. Though changes to core training suggested here will expose students to advanced data-driven skills, it is not the intention to signify that specialised learning is completed during the masters. In fact, a culture shift where data-driven skill learning is understood as career-long is needed.

Possibly due to the age of the ESRC Guidelines (2015), digital skills are not strongly highlighted and limited to broad competencies such as 'computing skills' and 'using a computer-assisted data analysis package'. There is therefore a lack of clarity on what constitutes essential digital skills and how these should be addressed in both doctoral and professional research training. As with research methods, it may be unrealistic to expect students to achieve competency in a wide range of digital tools, such as software packages for data analysis and online data sharing platforms. Section 4.3 consolidates these findings with those from chapter 4 to deliver clear recommendations.

3.4.1 Student capacity for identifying gaps: Reflections on training needs analysis (TNA)

To help **avoid data-driven skills gaps, students need guidance** from those who understand their research and have a general understanding of research design, to help them consider which specialist data-driven skills they will need to invest in learning. The TNA could evolve to help close this gap.

Within the Guidelines (ESRC, 2015), the ESRC championed the use of TNAs. These are annual (or more frequent) reviews of training need to help doctoral candidates consider how to usefully acquire skill and invest their limited time, well. The next section evaluates these as a device for helping doctoral candidates access data-driven skills and makes recommendations for how these could evolve, and this in turn, will impact on the Guidelines.

The Guidelines (ESRC, 2015) are difficult to fit within a 60-credit framework of methods courses. In this regard, learning is not individualised and thus is analogous to a level 1 undergraduate curriculum. As students move to doctoral supervision, they encounter a very individualised learning experience. The doctoral candidates involved in the interviews and workshops did not always feel that they had developed the skills to adapt to this individualised way of working and were not used to having a high degree of autonomy.

A masters that can bridge the structured masters learning and bespoke feel of the PhD would help students transition and build confidence. In a similar way to Manathunga's (2019: 1231) perspective of supervision, an approach that considers the past, present and future time in considering data-driven skills learning, may have advantages. The training needs analysis (TNA) that are conducted as candidates apply for ESRC doctoral funding currently take into account prior post-graduate learning (a completed masters for example) but could acknowledge the past, undergraduate and work-based learning to a greater extent. This could effectively 'excuse' students from learning they would find repetitive. An early TNA that considers gaps relative to the Guidelines as well as what specialist data-driven skills a student plans to use in their doctoral research (the future) could lead to learning these skills during the master's year, (the present). The TNA would prepare students well, to understand better their role in managing their education as they emerge from taught structures to independent study.

This idea comes with some risk. Students may be unable to reliably foresee what methods training they would benefit from. At the start of their masters, students would have written a research proposal that will include an overview of methodology. In the US, students often develop their proposal during their doctoral programmes. In turn, the US funding lasts longer, though completion rates are lower (around 56% social science in the US (Sowell, Zhang & Redd, 2008) versus around 75% in the UK (Clarke & Lunt, 2014)). It is unclear how much time is 'saved' by UK students completing their proposal in advance. It remains unexplored, whether the un-resourced labour of producing a proposal deters non-traditional students from applying.

As a proposal has been produced by 1+3 students, in theory then, they should manage to propose advanced training required, and produce a TNA. The students we spoke to were not confident in this regard, and their concern was echoed in

almost every expert workshop conducted. While students can write a convincing proposal, they are using argumentative skills, derived from their wider reading more so, than their practical research experience. The TNA requires management skills, an understanding of process (the required ordering of courses) and knowledge of the potential learning available to them. From one of the sense-checking workshops:

> ‘I really liked the idea of advanced methods, I really like the idea of having that much more tailored by the students and much more tailored to their projects, much more tailored to their needs. One thing that I feel is missing, given that it’s all about data- driven skills, there’s no element of: where do people go, where do they get the guidance, the maps for this?’
>
> Teaching Lead – Expert Workshops

It is clear that students also feel that they need guidance:

> ‘You should have supervisors who know what you should be looking at, at the start of your research.’
>
> Paul – Workshops with Doctoral Candidates

Early conversations with supervisors about the process of acquiring the data-driven skills, will be needed. Many supervisors use a small range of data-driven skills in their career, using a specialised model. Providing advice on a broader range of skills, as Paul argues, is what is needed. If the aim of a broad-based learning, captured within the 2015 ESRC Guidelines is to offer students the opportunity to shift, where relevant, to new forms of knowledge production which require alternative skills, then it is hoped that their supervisor(s) have the required fluency to provide guidance.

To help students navigate this early TNA, someone additional to the supervision team, may help. A training lead, for example appointed by a DTP, should have the knowledge and awareness of alternative approaches and training available. They potentially can also advise on researcher development training that looks at ‘soft’ or transferable skills that may be essential to a particular doctoral candidate learning a particular data-driven skill (for example, ‘Working with Policy Makers’ training for a student on a collaborative award). The balance must be handled with care, supervisors may be all a student needs; others may need broader guidance.

The next chapter focuses carefully on grounded experiences of learning and teaching data- driven skills to illuminate structural barriers that can limit and deter students. The chapter will incorporate learning established in this chapter, to imagine an alternative structure of supporting data-driven skill learning for UK doctoral candidates in the social sciences.

4. Interventions required to address skills neds

Summary

A legacy of issues impact on the learning experiences for core data-driven skills.

The semester-long presentation of data-driven skills learning did not map well onto the evidence from the literature on pedagogy. Learning around data-driven skills is incremental and time spent repeating lessons from the week before was seen as ultimately wasteful (though necessary if learning is structured across a semester). The disconnect between quantitative and qualitative translated into assignments which did not evaluate a broad-based conceptual understanding of knowledge production, but a focused test on skill for data creation, or data analysis. Other issues include feedback not being available until months after the course ended (another function of large class sizes), and that educators were often early-career colleagues with limited power to improve the teaching or learning conditions.

There were major structural issues facing doctoral candidates in particular, but also impacting on recognised, established and leading scholars. Time was a precious commodity, and doctoral candidates were keen to improve specialised data-driven skills and were frustrated by time-wasting repetition or the absence of any material account being taken of prior learning. They were also reluctant to ask for support from supervisors and colleagues who were equally time-poor.

Specialised and advanced training is not as accessible as it could be. Examples of training that was designed for disabled students was rare, and accommodations were perceived to be difficult to ask for. The cost of training was also prohibitive, even to ESRC-funded students who could 'claim back' costs. Doctoral candidates were challenged by the cost of software and in some cases, the cost of data. A gap was also identified around promoting well-being particularly around students conducting sensitive and emotionally demanding research. Well-being issues were often exacerbated by the increased pressure on doctoral candidates to deliver more with less support and still face a precarious employment postdoc experience. Peer-learning opportunities may work well to mitigate against isolation.

We recommend a number of changes, while recognising that a culture shift is difficult to deliver. Front- loaded investment in a re-imagining of 'core' data-driven skills coinciding with the first half of the DTP2 is recommended. The Guidelines could evolve to trigger a change. Here we recommend an evolution to core training, though space must be available for educating teams to develop within this framework, their version of optimal environments for learning data-driven skills.

4.1 Introduction and purpose of this chapter

This chapter presents working ideas on how to 'close the gaps' identified in earlier chapters and focuses on wider context and structural barriers to learning data driven skills. In exploring these issues, the chapter aims to capture the scale of change required.

This chapter presents evidence from our 5 phases of inquiry to deliver a response to Stage 2 of this research and its component research questions:

- What interventions are needed at a doctoral level?
- What is the scale of change needed in the curriculum?
- More specifically, what is the gap between our current postgraduate research and development guidelines and the core training identified as being needed?
- Do we need to shift how we structure doctoral education to produce confident users of data-driven skills?

The first two elements are approached with a discussion of barriers at institutional-level, across the academic sector and that appear in interpersonal engagements. Legacy issues impact on learning experiences for core data-driven skills and include repetitive material, large class sizes and teaching over semesters with disconnected assignments. The second two elements are approached with a re-imagining of data-driven skills training that proposes a mixed- methods and condensed learning presentation.

4.2 Defining structural barriers to teaching and learning

The chapter begins by examining the way that HEIs have structured and delivered data-driven skill learning. Our analysis draws from literature and landscape reviews, interviews and workshops with experts including doctoral candidates to leading researchers and wider stakeholders as well as the survey to DTP training leads. Barriers identified are thematically presented as operating within institutions, across the academic sector or in inter-personal engagements.

4.2.1 Barriers related to institutional structures

Institutional structures restrict data-driven skill learning. This section will present evidence around repetitive use of the same material, class sizes, semester-long delivery, disconnected assignments that are preventing access to specialised learning. It begins with evidence that the binary imposed by many institutions that separate quantitative and qualitative approaches is a barrier to learning.

Core methods teaching separates quantitative from qualitative approaches to data collection and analysis. Not only are these approaches separated, but there continues to be a disconnect, if not an adversarial nature to this divide. A consequence of this, is that strongly critical narratives are perpetuated that ultimately impact on both approaches as different actors reject the value of the other.

> ‘The first semester course was qualitative, second semester was quantitative and if you did the research masters, you did both and I did not find it helpful that they divided qual and quants. How I see methods, is that the two, they have so much linkage and if you actually discussed them next to each other, or combined, you see the approaches in a very different light. [Students] had already decided that they were going to do qualitative or quantitative methods, so there wasn’t much understanding been the two groups. So creative

thinking and looking at projects from different angles and answering different research questions. Although you might not be using quantitative skills yourself, it is really important to understand what they do and what you can do with those methods.'

Amelie – Workshop with Doctoral Candidates

From here it is clear that learning the data-driven skill isn't sufficient to access this conceptual understanding. Learning approaches in a forum that shows students the connections is required.

Students who feel very committed to a particular methodological approach can feel frustrated by being asked to complete modules that they see as irrelevant. For example:

'I would like to do a masters that was purely quantitative given the work that I'm having to do in my PhD … I mean the list of pretty much everything that I've had to do [in first year], I'd say 90% wasn't covered in my masters.'

Mary – Workshop with Doctoral Candidates

And from a student doing economics:

'I'm just sort of sitting there going well, it's all very interesting but it's absolutely pointless for me to have done any of this.'

Paul – Workshop with Doctoral Candidates

Students are unable to see the general value of knowledge production that falls outwith their specialist field. The separation of quantitative and qualitative approaches is fuelling this perception and this **binary was presented as a form of data-driven skills gap**.

A barrier throughout this project was the paradox, of working with literature and stakeholders that have transcended the qualitative and quantitative divide, while evaluating learning experiences and student cultures that are dominated by this division (Lenihan & Witherspoon, 2020):

'I did struggle with that kind of binary of quants versus qual because I felt like my research, and many others, bridges it.'

Clara – Workshop with Doctoral Candidates

And as one established scholar said:

'I'm a bit fed up with the binary'

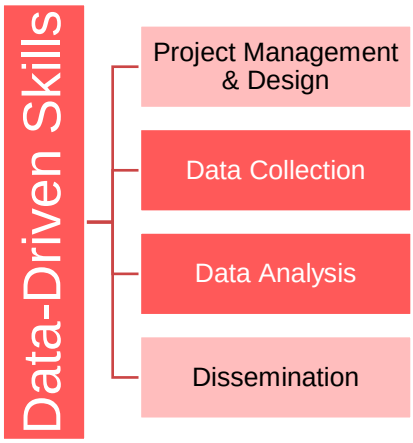
Established or Leading Researcher - Interviews

In this section we explore why the division perpetuates and how it impacts on learning data- driven skills. As discussed in chapter 3, many felt current teaching

focused too much on skills around data collection and analysis.[11] See Figure 10, below, for an illustration. In particular, some respondents felt that the learning in core quantitative courses was too frequentist, and that this was now 'old-fashioned':

> 'Some of the themes are actually quite dated, particularly in quantitative skills that are rooted in frequentist statistics. I believe this is counterproductive and should be reviewed – it makes little sense to require those with expertise and specialising in qualitative approaches, for example, to do 'statistics 101' that has little connection with the cutting edge of quantitative social science. If anything, it adds to epistemological and other divides and misunderstandings.'
>
> DTP Training Lead – Survey

Figure 10 Subset of 'pillars' of training required to produce skilled and confident data-driven researchers



The argument was shared by others who all supported the teaching of numeracy-based epistemologies but felt that students were learning skills that didn't in fact help them develop their conceptual understanding of what data enquiry could achieve (whether due to maths anxiety or other resistance to learning, see section 2.6). Several too talked of the maths-catch up work needed to support frequentist approaches to traditional quantitative learning that further alienated students and divorced the learning from their interest in data-driven research approaches. There are now new approaches to data that could usefully be drawn into learning experiences. The suggestion provided by this scholar and others was to use cutting edge digital skills around social media data or concepts of machine learning to contextualise the learning and reveal the potential.

Hesse-Biber (2015) has championed mixed-methods approaches and educators who are able to see and teach the strengths of both quantitative and qualitative approaches. More could be achieved by a mixed methods presentation using for example, secondary analysis of survey data or digital data:

[11] This represents only two of the four 'pillars' required to produced skilled and confident data-driven researchers

‘The new media stuff is really important - a really good way to move beyond qual[itative] and quant[itative] analysis and into data management side. Our sociologists want to work with social media.’

Established or Leading Researcher – Interviews

More understanding around how methods teachers can take an innovative, and alternative approach to address the limitations of their learning would strengthen our understanding of pedagogy and social research methods. As Purdam says:

‘Teaching social science research methods requires a more coherent pedagogical grounding and robust curriculum design which integrates theory, practice and hands on practical training.’ (Purdam, 2016: 258)

Further evidence of the division came through the workshops with training leads:

‘I also think that qual[itative] scholars, in my experience, [have] been absolutely terrified by quant[itative methods]. Just terrified, and you know struggle to grasp some basic propositions. There needs to be involvement somewhere along the line in order that PhD students understand and recognize the need for conversations across different methodological approaches and what the value in the benefit of that might be in what that might actually deliver for social science research.’

Training Lead – Expert Workshop

Critical to this for Hesse-Biber, an expert participant in Nind & Lewthwaite’s (2018a) work, was making the standpoint explicit, in order to show that knowledge production is anchored, socially, culturally and politically. For students to become knowledge producers, they need the reflexive tools (Hughes, 2016) to appreciate their own standpoint and to have the capacity to evaluate the degree to which they operate objectively, subjectively, or somewhere in- between. This conceptual understanding is inhibited by the division between quantitative and qualitative approaches.

The Guidelines (ESRC, 2015) carefully avoid the reductionism of the qualitative/quantitative binary by using three categories: Learning Outcomes; Principles of Research Design and Data Collection, Analyses and Management. Yet without exception the core training was delivered within master’s programmes with distinct courses for quantitative and qualitative methods often taught by expert scholars entrenched in their own ‘camp’. Further, amongst the participants, there was general rejection of this together with championing opportunities to blur or transcend these boundaries.

‘I would divide it less as quant[itative] and qual[itative] but as methods and contextual expertise. When you understand how to model generally, you can pick up new modelling approaches quickly. Bayesian inference, causality, we have all these methods to work out if something is causal, but without the qualitative side you can’t have a full picture.’

Recognised Researcher – Interviews

Another participant, felt that the binary approach made it harder to see how the researcher has a role to play in guiding the research 'in the field' performing a series of micro-decisions that each need to be understood as impacting on the potential findings:

'Basic skills in modern longitudinal efforts is core in quant[itative] social sciences – feeling in qual[itative] is that it has advanced and I see value in qual[itative] longitudinal interviews for example – but we should see reflections at each step, not one post-hoc justification. I'd like to see this become more key in methodology.'

Recognised Researcher – Interviews

While a mixed method presentation was championed, the binary culture is well established and significant work and care will be required to resolve the issues stated.

There is a strong message that the Guidelines (ESRC, 2015) took the correct approach in avoiding structuring learning according to qualitative/quantitative approaches, yet this has not impacted sufficiently on how training is actually delivered in universities.
The distinction between quantitative and qualitative needs to be surmounted by those working across disciplinary boundaries, particularly where their approach is informed by very different traditions in knowledge production and data-driven skill use. Therefore, for interdisciplinary scholars, learning data-driven skills separated into quantitative and qualitative is a form of data-driven skills gap, as there is no space to consider how different skills can be brought together to develop a rich, nuanced understanding of a research topic perpetuating the binary limits all researchers. A mixed-methods approach was well received by students across the social sciences, as better able to prepare them for choosing and learning specialised data- driven skills. Before presenting evidence around specialised learning, the chapter will discuss repetition of learning and other institutional barriers.

Repetition is considered a data-driven skill gap because it is limiting access to specialised training. Masters level core course are large, with a diverse range of students as will be demonstrated here. To accommodate students without prior methods learning, courses begin at an introductory level.

Students joining master's data-driven skill training varied greatly in prior learning. For many students with prior learning, the methods courses informed by the Guidelines (ESRC, 2015) did not advance their knowledge base despite taking up a third of their master's degree credits. The Guidelines also champion a standard student experience, and this has resulted in the pooling of resources in many universities, offering core research methods learning across a college, rather than within disciplines.

This may be contributing to recent findings from the 'Review of the PhD' report that:

'There is some suggestion, particularly from international commentators that UK training does not provide enough emphasis

on the development of quantitative skills and that the quantitative training that is provided is too basic. DTPs should continue to provide high quality research methods training for all ESRC-funded social science PhD students to ensure they achieve minimum competencies in quantitative and qualitative research and that there are opportunities for students to pursue appropriate advanced level training.'

Tazzyman et al, 2021: 7

Diversity in the classroom doesn't just result in repetition of learning for many, it also puts enormous pressure on the educator (Kilburn et al., 2014b). Thus, the teacher is facing multiple demands on the learning experience they provide. Consolidation of key learning points is critical:

'Repetition [is] useful and actually learning your basics and understanding your basics are really important.'

Training Lead - Expert Workshop

This is a strong argument. Semester-long training exacerbates the need for repetition, yet repetition is required for consolidation to happen. In this section, repetition is understood as repeated learning that is not usefully consolidating key learning.

Some students, who stayed at the university they completed their undergraduate degree at, to study for their PhD reported that the courses they had completed as an undergraduate were repeated: often with the same educators, the same learning materials and sometimes the same assignments:

'I had to do the same courses again. The actual same, with the same lecturers two years in a row. The lectures all over again, do the same pieces of work. Just stop. Just loosen up a bit and look at what the students need rather than what the university regulations are and think about what your PhD is going to need you to do.'

Mary –Workshop with Doctoral Candidates

One doctoral candidate who had the prior data-driven skill learning (though from a different HEI), said:

For me [repeated learning] was 20 credits that I was just sitting around, literally doing absolutely nothing.

Amy –Workshop with Doctoral Candidates

There were many similar stories that emerged from students.

Class sizes notably restrict data-driven skill learning and should be regarded as creating a 'gap'. The discussion of literature in section 2.3 that championed student-led learning strategies such as asking students to influence structure and content, can only work if the class size is small enough to complete the exercise, and the educator has the workload to then respond to what has been agreed.

In practice, social research methods class sizes tend to be large, pan-disciplinary spaces, under-resourced by their host universities. In their Review of the PhD, Tazzyman and colleagues (2021) reported that many senior contributors felt a course that spanned the social sciences disciplines would be too generic, would not recognise different levels of prior learning or disciplinary requirements. This may be particularly true for certain disciplines and economics is often cited in this regard:

> ‘[Students] taking an econometrics module within the economics department is probably going to be better meeting their needs … [students] want the methods that would be the most, bring them to the frontier of what they need to do.’
>
> Training Lead – Expert Workshop

Students coming together from different disciplines are likely to have ‘different goals’ (Nind & Lewthwaite, 2018a: 79), interests, experiences and perspectives. Again, the steer from the Guidelines to standardise the student experience had led many HEIs to deliver core training in pan-disciplinary courses. The term interdisciplinary cannot be used here as the size of classes did not allow for students to take advantage of this diversity and challenge their perceptions around ontology and epistemology. Indeed, most students emerged from core training unable to define these terms or relate them to their research.

Teachers found it difficult to meet learners’ expectations, and equally, some learners felt that their needs were not met:

> ‘I did find the stat[istic]s interesting, but I think the issue that I found with the masters was that everybody wants to do such different thing, everybody was kind of on different levels as well, which didn’t help. And there was just no way that a master’s course, which had those four huge big key modules could encapsulate that for everybody and I think, to be honest, I came out of the masters feeling rubbish and feeling like I couldn’t research, because I got myself so confused with all those things that’d we’d done a little bit of.’
>
> Lucy –Workshop with Doctoral Candidates

Again, many doctoral candidates and researchers echoed these statements. While learning data-driven skills was limited by the pan-disciplinary approach to ‘core’ training it was ultimately the class sizes that presented the most significant barrier.

The management rationale for these class sizes include economy and standardising the student experience. Universities rationalised teaching teams, reluctant to employ a team of educators to teach the same data-driven skill to multiple groups of students (for example teaching chi-square or focus groups multiple times to multiple disciplinary groups). This attitude is aligned to the dominance of skill learning rather than conceptual understanding and fails to acknowledge the pedological demands on data-driven skill learners and the value of bridge- building into disciplinary learning and students own research ambitions.

Large class sizes may too, be due to the expansion in post-graduate taught student numbers over the last five years,[12] though methods class sizes have been large for decades. The students we spoke to struggled with large class sizes, often learning alongside over 100 others (with a wide range of disciplinary homes and prior methods learning; one institution revealed more than 400 students were in one methods class).
Core methods learning is accessed by non-ESRC funded students increasing class sizes:

'Making [core training] core means that the classes are very large so not great teaching or learning environments.'

Established or Leading Researcher - Interviews

This participant continued to reflect that attracting lecturers with strong method skills that could also communicate to these large groups was a challenge. Strong teaching skills are rare, and a lack of investment in educators' pedagogical development and large classes deter talented scholars.

While defendable from a management perspective, the result is that students, often anxious about methods learning, and sometimes reluctant to engage, are challenged further by large class sizes. While leading reports (such as the OECD, 2020) discuss 'scaling up' of training to reach more people, and even cross international boundaries, the UK reality is that methods learning is under enormous pressure and are far larger than the discipline/substantive classes that students complete within their masters.

Smaller class sizes would enhance the learning opportunity and dividing students into 'no prior' and 'prior' learn categories could be a way of improving engagement.

In chapter 3, in discussing the influence of the Guidelines (ESRC, 2015), we presented data that evaluated the broad-based approach to data-driven skills. In order to meet the Guidelines in full, and to engage learners from across the social sciences, and support the learning of those with no prior learning, data-driven skills training tended to be skill focused and offer fairly shallow engagement with 'each' skill. Moving on from the content and context, this section will consider the structural presentation of this teaching and learning.

The imposition of 10-12 week terms with classes taken just once or twice a week may challenge learning and in making data-driven skills harder to learn, constitute a gap:

'I struggle. Jumping in an out of different classes and only doing an hour or two a week.'

Miriam, Workshop with Doctoral Candidates

[12] Total numbers have risen by around 20% from 2014/15 to 2018/19 https://www.hesa.ac.uk/news/16-01-2020/sb255- higher-education-student-statistics/qualifications

One student reflecting on practical activities in Nind et al's (2015) paper on advanced postgraduate data-driven skills training highlighted that pressure to complete exercises within a class's time limits restricted learning, and time outside the class was required to revisit and practice the material to achieve the learning outcome. As argued, students particularly struggle when this practice work is not related to the data-driven skill they wish to use in their doctoral research and where acquiring this skill does not advance their conceptual understanding. Doctoral candidates reported less motivation to invest their 'own' time in practicing skills they viewed as irrelevant, especially as their PhD proposal, inclusive of research design and approach had already been funded. This emphasis on practicing skills, diminishes opportunities for critical reflection where students have time to assimilate learning into their lifeworld and build a metaphorical bridge between the class and their own research interests (Spigner-Littles & Chalon, 1999).

Training was more useful where it involved opportunities to practice and took place over an intense-time frame, such as a week:

> 'Refreshing introductory skills, to practically play around with data is essential in understanding. Hands on training is the most useful, the longer training and smaller groups provide the best opportunities to learn, it also is a chance to build competence.'
>
> Recognised Researcher - Interviews

The participant is reflecting on training around a method that will be used, where a degree of expertise is wanted, rather than a shallower engagement required for data literacy. In this example, 'longer' is used relatively to distinguish between half-day training and events that take place over days. Another participant commented on the value of shorter (<1 day) events:

> 'I usually feel like I get the basics once the training is complete, but never feel like a master. Only if something is used repeatedly do you actually learn something. These trainings show you what is out there but little more. The notion that these are professional development is kind of a misnomer. Nobody is interested in that you went to a workshop, but that you can use [the method] immediately.'
>
> Recognised Researcher – Interviews

And:

> 'I found that having a full day, of especially numerical kind of subjects, was helping me to warm up my brain to then dive into the depths of what you need to know.'
>
> Miriam – Workshop with Doctoral Candidates

These challenges were reported by many participants. Even with 4-8 hours devoted to a method, there is a shortfall. Students have understanding but would not claim to have 'specialist skill'. Still, having these hours clustered, rather than spread out over the semester is a strength of this approach. Specialised training that happens after the masters is presented in something of a vacuum where students' helicopter into an event and leave with no (generally speaking) prior or post training engagement.

Using the training model of condensed week- long courses but structuring it around core learning could provide space to engage with the data-driven concept and practice the skill within the same, longer, timeframe:

> ‘Spending a whole week, and actually being able to say ‘I get this’ rather than spending an hour lecture which is then followed three days later by a sometimes unrelated tutorial. You don’t learn enough to know how to ask questions.’
>
> Teaching Lead – Expert Workshop

From another expert teacher:

> ‘I think [intensive learning] is a good way yes, I would think it’s a lot more preparation.’
>
> Teaching Lead – Expert Workshop

The additional preparation is to enable a flexible learning experience. As the educator interacts with the students and understands their learning needs, they teach best when able to respond with sometimes different examples, or different ways of teaching key concepts. There is time in theory, between weekly lectures to produce new materials. In reality, the large class sizes and high workloads mean that educators can’t exploit the drawn-out presentation of semester-long courses. With condensed learning there is no time, between learning segments so advanced preparation is key. Material not used in class can be provided online to help students practice and this post-course rehearsal will aid learning.

Unless the method is practiced, the learning is soon lost. Some commented that it’s easier to learn a second or third time; but this repeated investment in methods that won’t be used is placing a strain on many doctoral candidates and recognised researchers. Practicing without support (or training) is typical as a post-graduate student takes ownership of their data and thesis, however confidence can disappear if they feel unsupported by their supervisor or that they’ve not applied a methodological framework well:

> ‘If I had had more support, I would have spent more time analysing instead of reading basic material.’
>
> Recognised Researcher – Interviews

Moving away from semester long structures is a pedagogical advancement. The next barrier identified is disconnected assignments that undermine confidence in, and learning of, data- driven skills.

To have **confidence in data-driven skills, core assignments should be connected** and assess not just data collection and/or analysis but at least project management too. Many respondents spoke explicitly about the need for data-driven skills training that permits students to feel confident in planning research. This relates to our own definition of data-driven social research that produces evidence strong enough to withstand scrutiny. The question explored in this section is not what skills/topics do students need, as much as how can we create learning experiences that produce confident researchers?

According to contributors, assignments generally do not work well in this regard. The doctoral candidates and recognised researchers involved in interviews and workshops generally felt the assignments showed their analytical skill, but not a broader understanding of why that skill would be applied over another. Assignments rarely required links to theory or disciplinary learning, meaning that students invested time in producing work that was divorced from their other assignments. Assignment deadlines at the end of an 11-week course meant feedback arrived months later when new learning was happening. Disconnections between quantitative and qualitative courses meant feedback was rarely relevant to new learning. Assignments that worked together, would permit students to share and develop their conceptual understanding. The value of this would be early feedback to help them build confidence in their knowledge (or constructive feedback that led to extended study).

There was some evidence from the landscape review, interviews, and workshops that assignments had progressed from essay-based assessments with increasing use of reports and portfolios. Assignments continued to assess skill, to a greater extent than conceptual knowledge. Secondary data was used and there were examples of students being introduced to repositories such as the UK data archive to support their assignments and this shift, coincided with the Covid-19 pandemic and the greater dependency on online resources. Examples of challenge-led assignments that use data from external partners was rare (see section 3.2.2. for an outline of Wollschleger's (2019) approach). We also found no examples of students being asked to demonstrate project management or dissemination strategies explicitly.

The chapter will now present evidence to support the claim that **space must be made within the Master's for specialised data-driven skill training**. As has been argued, reducing time spent learning skills that would not be used, would allow investment in specialised skills of direct relevance to the doctoral candidate's field of study.
Data from interviews and workshops revealed that students spend a third of their master's year learning data-driven skills at an insufficiently advanced level for application during their PhD.

There is some leeway in the Guidelines for some disciplines to offer specialised or advanced core learning (for example Economics is given as an example). Thus, it is to some extent the disciplines which dictate the advanced level of core training, rather than the individual student though this is undermined if data-driven skills are taught in larger pan-disciplinary classes. Economics is a strong case-example: students are still required to complete 'core' training. The remaining 60-credits of courses (the dissertation is worth 60-credits and the total masters is 180- credits) are not optional disciplinary courses, but advanced methods courses usually focused on econometrics.

While increasing the remit of DTPs to deliver advanced and specialised training is one option, placing greater pressure on doctoral researchers to incorporate data-driven skill learning in their PhD years was resisted by all contributors to this research. Finding ways to eliminate the repetition and move students to advanced training within their masters is another solution. This could also help close the gap

between structured, undergraduate-esque learning and autonomous doctoral research.

Q-Step students help illustrate the scale of barrier caused by core courses that assume no prior learning. Q-Step focused on undergraduate learning (with impact on doctoral education and beyond) was mentioned by six participants in the expert interviews:

> ‘Q-Step has created a culture at my institution that recognises different data starting points for postgrads. Q-Step undergrads coming to PG programmes seem best prepared to handle data and quant[itative methods] (compared with other social science graduates).’
>
> Recognised Researcher - Interviews

The Q-Step programme, launched in 2014 and leading to 17 Universities in the UK hosting a Q- Step Centre aimed to create a step-change in how undergraduate social scientists use and understand social numeric data. The landscape review revealed an extensive impact of the initiative. In late 2020, 81 degree programmes had been launched; 236 new modules developed; 424 different employers had hosted placements. In 2018 alone, this produced 304 industry placements with 992 enrolments on to Q-Step degree programmes. Q-Step programmes offer students between 30-50% of learning with numeric data, the remainder focusing on disciplinary learning. A 2018 audit (Ferrie, 2018) of the Q-Step centres mapped against the 2015 Guidelines revealed that undergraduates had completed all the learning related to quantitative methods and research design, often exceeding the requirements having learned advanced quantitative courses. Ideally these students should be able to spend their master’s year learning at a higher level. This also brings into relief the question of whether all students should be able to present prior undergraduate learning as a way to ‘leap’ over the core training that would be repetitive. Qualitative skill learning has not received investment commensurate with the Q-Step initiative and the landscape review did not reveal undergraduate degree courses with elevated qualitative training (most included a 20 credit module).

It’s clear then that some students are entering their post-graduate courses confident in their skills and could be better served by an advanced or specialised training module. A balance is required, to empower the students in this regard, and to respond if they, from an informed position, identify that further learning is not the best investment of their time. This echoes the findings of our discussions with students:

> ‘It is easier to learn when you're trying to look at [data-driven skill] that you're interested in, as opposed to something that's just been given to you, and so I think, maybe giving more choice to the students, because then you could at least try and tailor it in some way that might be useful to your further research, even if you can’t explicitly use what you've collected.’
>
> Lucy - Workshop with Doctoral Candidates

Evolving data-driven skill training is only one solution and to help students, disciplinary learning experiences could change too. Kilburn et al (2014a, see also

MacInnes, 2018) were right to point out that methods teaching is a different learning experience for most students compared with their disciplinary lecture-based courses. Thus, substantive lectures that foreground data and make knowledge production explicitly visible are helping students prepare for learning data driven skills. This approach has been practiced throughout the Q-Step network through 'embedding' (Buckley et al, 2015). This process sought to increase the visibility of research method into substantive teaching. Here disciplinary lecturers were challenged to present findings not as facts or knowledge, so much as evidence for a theory and the strengths and limitations of that evidence relative to the methodology used to produce it. As one participant said, '… it doesn't feel like you're doing quantitative data for quantitative data's sake' and later in the same piece 'it's so interlinked' (Recognised Researcher – Interviews). Two key points emerge here, first it's not just data-driven skill training that needs to build in relevance, but subject-specific disciplinary learning must do more to acknowledge knowledge production skills. Secondly, the semester-long course model works where a student is engaged with thinking through their data-driven skill learning in all classes they are attending, rather than being helicoptered into a course, once a week. Building towards the dissertation is another key strategy.

Once core training is complete, a student then turns to a dissertation or doctoral supervisor to continue learning data-driven research skills with application in a substantive field, (Boud & Lee, 2005; Fillery-Travis & Robinson, 2018). In the primary research completed for this report, the dissertation was rarely discussed as a site for data-driven skill learning or as a space where confidence could be built. Despite the 3-year PhD project being understood as a learning experience (ESRC studentship awards are also referred to as Training Grants by the ESRC) the dissertation appeared removed from this narrative:

> 'For my dissertation, for my masters I did ethnography. I'd never done that before, and I was struggling to find appropriate materials. Also struggling with finding the time to familiarize myself with what is actually involved in doing offline ethnography. My supervisor was a huge help, which I did appreciate at the time, but I was stressed using a tool which I had never used before, but which I knew absolutely complemented the whole research and added that extra dimension to it. If I'd been aware and started thinking about the topic for my dissertation earlier, I could have perhaps saved myself from the time needed, or been less stressed during the time.'
>
> Miriam – Workshop for Doctoral Candidates

If prepared for adequately, the dissertation is an opportunity to build confidence in data-driven skills. Data-driven skill training can help prepare for this explicitly, and advanced training could be identified in advance to support students who wish to apply new skills.

Confident users of data-driven skills need time to work with and practice the skills they will use in their doctoral research. They also need time to meet those who can advance their understanding of skills. The one-size-fits-all approach used by many universities and mixing all learners together, effectively undermines advanced work and specialised learning. For students to transition to the more autonomous learning

experience of their doctoral research, masters learning needs to evolve to offer more options, more pathways and access to advanced thinkers.

4.2.2 Barriers related to the academic sector

This section turns to barriers which individual institutions cannot resolve. Across the academic sector, and across all career stages, social science scholars are struggling to find the time to access training. This is exacerbated by cost: of training and of some data. The section begins with evidence that scholars do not have knowledge of how to access suitable specialised training.

Beyond DTPs **most contributors were not aware of how to find training to acquire specialised data-driven skills.**

The final gap discussed in this chapter pertains to accessing specialised or advanced training. The five phases of enquiry uncovered a paradox at the heart of advanced data-driven skill learning. Substantial and comprehensive specialised training is widely available to UK social scientists, yet a minority are aware of opportunities, and most are unable to assess whether training advertised will meet their needs. To some degree, this reflects a highly active and 'busy' training and development market. There is much to be admired in the range of opportunities available, yet it is not always clear what they comprise and how well they will be delivered. If supervisors, established and leading researchers are aware of training, they are unsure whether they are eligible to attend. There appears to be a need to consider how, where and to some degree when doctoral candidates and recognised researchers obtain information about training and development opportunities.
As part of their training grant, ESRC students are provided a nominal Research and Development Support Grant (RTSG) that is equivalent to around £750 for each funded year and can be used to support data-driven skills training or transferable skills such as attending conferences. The grant is available for each year of funding including the 1+ master's year, though not all of the DTP training leads who contributed to the Workshops were aware of this.

While influential, the gold standard is not accessible to all, and this inevitably constitutes a data driven skills challenge amongst the wider community of social science doctoral students who do not have access to core training. In the UK, there are still many master's programmes financed by other funders, leading to PhD study that do not require core methods training to be complete. From the interviews we learned that self-funding students avoid research training- intense masters programmes often because they don't think they'll 'use' the learning or feel the risk of low marks could devalue their investment in post-graduate education.

Access to advanced training is particularly key for qualitative research which in core training can be taught as much as 18 months before a student begins data collection or analysis. Several DTPs reported that advanced qualitative training continues to be popular, particularly ethnography, analysis (notably software training including NVivo) and situated ethics. In the 2021 analysis of training needs data at Scotland's DTP (the Scottish Graduate School of Social Sciences), the most asked for training (47% of the 326 who had completed TNAs) was around 'interviewing', something covered by core training. Another example, provided above in relation to

dissemination is writing, not just as a means of producing a visible output, but as a form of analysis:

‘Working on research proposals and papers together - the latter has really helped my understanding of qualitative approaches’

Recognised Researcher - Interviews

There is then a place for training that takes place over a longer timeframe and allows people to develop skills with their own data and practice their data management and analysis skills during the doctoral years.

Currently most doctoral candidates obtain advanced and specialised data-driven skills training after their masters, and during their doctoral years. Whether students need a little or a lot of advanced training, they must fit this alongside their research. Advanced training is available from each student’s university, from their DTP, from national initiatives such as the National Centre for Research Methods (NCRM) and a range of other units, institutes, and commercial organisations. From the landscape review there are significant resources available in HEI’s for research development training that focuses on academic practice and strong networks between HEIs built by Research Development Practitioners. There is much less evidence of a comprehensive advanced methods programme offered by individual HEIs:

‘The level of your methods training is highly dependent on your institution and how much your institution is interested in methods and in what methods people use in your department.’

Amelie – Workshop with Doctoral Candidates

Half of the doctoral candidates who took part in the workshops had heard of NCRM. Those that had, had used their training and were highly positive in their evaluation. As the participants of the student workshops were all ESRC funded, perhaps they should have been aware of the NCRM.

‘It is amazing that, it doesn’t matter how much advertising and promotion, still people out here have never heard of [ESRC investment].’

Training Lead – Expert Workshop

There has been significant investment in how NCRM are communicating to students directly and through DTPs, and the findings here show that this work should continue. A contributor working with another ESRC investment that provides advanced training added:

‘I think that 50% is already a generous estimation, I think we are below that threshold.’

Training Lead – Expert Workshop

In the UK Careers in Research Online Survey (Vitae, 2019a), 12.5% of researchers at postdoctoral level and beyond had not accessed any training or CPD activities in the previous 12 months, and the majority (63%) had accessed fewer than five days (n=6392). Similarly, in the Principal Investigators and Researcher Leaders Survey

(Vitae, 2019b), 11% had not undertaken any training or CPD activities, and 77.2% had undertaken fewer than five days (n=2327).[13] The years when students are engaging with master's and doctoral research are relatively learning- intensive compared with the years beyond the PhD. It may also be that if students are unable to get the training (in advanced methods for example) during their master's or PhD, that they will be unable to find the time to get the training later on. If this is the case, then providing easier access to advanced training for doctoral students will improve the skill set of social scientists emerging from UK universities.

In general, the early-career participants in this study were able to generate more examples of training providers than mid-late career participants. While this may be because they are more likely to attend training (or have done so recently), it would appear that supervisors, the established and leading researchers, ought to have had detailed knowledge of the training and development offers that are available. This is mostly about fulfilling their role in developing new researcher talent but should also be a way to make sure that they are looking at opportunities which they themselves want to take up. Several established and leading researchers also hadn't heard of NCRM. Thus, doctoral candidates need help finding training, and their supervisors may in turn, benefit from learning about the UK training landscape. As a minimum ESRC funded supervisors should be expected to understand this landscape (especially the ESRC's investments) and how to maximise opportunities for their students.

Some report being overloaded with information and unable to discern what training would best meet their needs: 'It is always harder to find training, than to be trained' (Established and Leading Researcher – Interview). As a national provider, NCRM may be uniquely situated to provide this clear 'signposting and co-ordination' (Wider Stakeholder - Interviews) and to ensure that their online repository of training resources is improved and extended to produce 'a hub of resources' (Wider Stakeholder - Interview) including the new open access materials HEIs and DTPs have produced during the pandemic. This would substantially improve access for people who were currently struggling to attend timetabled events (in person or online) and this could help deliver outcomes for the equality, diversity, and inclusion strategy.

> 'There is so much data available from so many different sources. There is nothing that says where and what you need to focus on. All the time exploring, and vetting is a waste of time, for the basic skills being told where to go would be more helpful. It would be fantastic if ESRC could vet material on statistics and centralize the best materials.'
>
> Recognised Researcher - Interview

While NCRM have an extensive and valuable online presence, students and colleagues need to know this repository exists and it should be easy to navigate. Thus, a useful map of provision is needed, and this could be accessed through an

[13] The Researcher Development Concordat, to which UKRI is a signatory, suggests that researchers should have a minimum of ten days professional development each year: https://www.vitae.ac.uk/policy/concordat/full

app, to better fit how early career scholars communicate, network and source information. This data would not just signify what training was available but guide users to the best training for them including information on practical examples, duration, complexity, disciplinary approach (if any), learning styles, opportunities to ask questions and so on. A map that evidenced where there is overlap between university, DTP and NCRM training would help clarify the role and function of each unit in the training landscape. This single site for information would help ESRC investments such as the UK Data Archive and NCRM who produce significant quantities of training of a significant quality, promote their opportunities:

In designing advanced methods short courses, teachers are faced with the issue of creating sufficiently challenging content while ensuring that the training accommodates a potentially wide range of skills and experience (Kilburn et al., 2014b). For teachers and students, challenges also arose in clearly communicating the level of the training and what 'advanced' meant for different training topics. To address these issues, Kilburn et al. (2014b) suggest that teachers should provide detailed descriptions of a course's aims, content, and required knowledge or experience prior to registration. A typology introduced by FAIR4S (2019) could be usefully applied: 'basic', is analogous to having awareness towards comprehension; 'intermediate' comes with the ability to apply and 'expert' means the 'ability to change practice in this skill' (OECD, 2020: 22). The NCRM currently use a similar typology of 'entry' indicating no or almost no prior knowledge, 'intermediate' indicates some prior knowledge or 'advanced' symbolises specialised prior knowledge. If all ESRC training investments including NCRM and DTPs utilised a single framework, and this was also embedded in university's training needs analysis, then doctoral students would find it easier to identify suitable training. In theory, setting up such a typology across the UKRI could then help research councils measure and compare progress towards upskilling (percentage increase in social scientists having intermediate digital skills by the early career stage; proportion of post-doctoral researchers with expert digital skills leaving the academy to join industry).

From the data around skills gaps, it may be that investing in digital skills could be of notable benefit to NCRM in raising their profile and to social researchers looking for this training. Most DTPs mentioned a new Training Network set up by NCRM with the DTPs that has the potential to co-ordinate training and facilitate partnership working and this was highly valued by the training leads. Another benefit of developing a training typology is to flag who can attend. Established and leading researchers who can find time, are unaware that they can attend training provided by ESRC investments:

> 'I would tend to read around methods. I've also been on courses, online via zoom. Been on loads recently trying to catch up. Often in house - STATA/R. Not used NCRM but have been meaning to. I feel embarrassed doing it. As staff it becomes a barrier in thinking you shouldn't be using resources - being exposed. Maybe something marketed to staff specifically? Suspect there would be a lot of take-up/demand. I know lots of people that would be interested.'
>
> Established or Leading Researcher – Interviews

This section has focused heavily on UK advanced training and the landscape review revealed some trends which could influence a doctoral intervention. A profile of three international summer schools provided insight into overseas demand for, and scale of advanced training available, see Figure 11, below. Each site was chosen because they had been identified from the interviews or DTP survey as offering a significant contribution to the training landscape, and are summarised below.

Figure 11 Landscape Review – Leading overseas summer schools

Overseas Summer Schools		
US Summer school. This event takes place over two months and has a strong global reputation. Almost all of their provision is quantitative, (2 of around 80 qualitative courses available). There is not much around research design, or information to help students choose which course would help them meet their research aims, so a degree of expertise is required to navigate. Critically, this summer school offers four- week courses and this is significantly longer than anything provided by DTPs of the NCRM. This summer school did not directly offer training in transferable skills, but their methods training did mention engagement, impact, and also communication and networking skills.	**Northern European School.** Based in Northern Europe, this summer school has a strong reputation for methods training, with a thematic focus around health and environment. Students attend for 2 weeks. Again, principles of research design were not included, and there were no examples of qualitative methods. There were references to big data and social media, which would presuppose some text-based data-driven skills even if the analytical approach used a statistical model. Again, communication and networking skills were explicitly mentioned within methods training as were ethical and legal issues.	**Central European School.** Located in Central/Eastern Europe, this summer school has a stronger qualitative programme and strong disciplinary threads throughout the events offered. There was less emphasis on traditional approaches, such as interviewing and more on emerging methodologies (though with a rich history of) arts-based approaches, including photo elicitation. This school also embedded transferable skills in their methods training and delivered an extensive provision including teaching and work experience; language skills; ethical and legal issues; engagement and impact; research exploitation and communication and networking skills (offered in relation to working digitally).

The cost of these summer schools will be prohibitive to many. Based on 2021 fees and using conversion rates correct in January 2022, they range from £450 for a

week in central Europe, up to £3, 850 for 4 weeks in the US. This cost covers fees only. The University of Essex Social Science Data Analysis summer school is the closest UK based host, offering the level of provision demonstrated by the first and second example. A week long course at Essex for academic delegates costs £1100.[14] Most training in the UK takes days rather than weeks. This may be adequate if a student needs a range of training support. For example, a student studying gendered violence on the university campus could require training in sensitive interviewing, critical discourse analysis, writing for impact and emotional safeguarding and find this knowledge from 4 distinct training events. In contrast a student who is taking a deep-dive into a particular data-driven skill, and particularly common in anthropology and economics, may well benefit from access to an extended event. Currently ESRC students can use their Research Training and Support Grant to access overseas training and in addition to the learning, can build international networks. ESRC-funded students can apply for an Overseas International Visit grant which could sustain the additional living costs associated with an extended stay away from home. The landscape review did not reveal equivalent financial support available to non ESRC-funded students.

One limitation of the landscape review is that it affords little scope for a critical evaluation of overseas provision. From the review of the summer schools, the two that focused predominantly on training around numeric data, tended not to have any disciplinary anchoring to the training, or acknowledge the disciplinary context of data production and analysis. In contrast the third example of a summer school, dominated by a broad range of qualitative approaches, did emphasise disciplinary context. For the former to work, according to the literature reviewed, the learning would need to allow students to build the conceptual bridge between the training, and the student's own research aims and ambitions. The events promoted did not specifically mention that time would be devoted to this, though this does not equate to evidence that it is not. The drive towards stronger numeric data-driven skills may in truth, diminish the critical thinking skills, predicated on disciplinary engagement, that is a key strength of the social sciences. These findings confirm a rich landscape of specialised training opportunities in the UK. In discerning the 'best' course, having someone to talk to would be beneficial:

'Having someone who can provide information like a site librarian, who can direct me to the correct courses, I need that guidance.'

Recognised Researcher - Interview

This idea should map onto the new Professional Development role proposed by Tazzyman et al (2021) in the Review of the PhD. As stated, most training is done by doctoral candidates and one major reason for this is time pressures on academic staff. **A lack of time limits investment in data-driven skills**. From DTPs there was evidence that students were less likely to engage with their Training Need Analysis process the closer they got to submission, citing time pressures (to participate and to engage in training even if required). Early career scholars struggled too:

'One of the challenges is time. It takes a lot of time to learn and do [a new method]. Confidence is another area that can be hard to

[14] See: https://essexsummerschool.com/summer-school-facts/fee-structure/

overcome. It requires continuous learning and re-engaging. This takes time.'

Recognised Researcher - Interviews

One of our participants, established in their career, also reflected that early career scholars are often working in a state of uncertainty, with short-term contracts and pressure to optimise employability. Investing their time in training can feel particularly risky where that learning may not be valued or relevant to the next research leader willing to hire them. There is little structural support to encourage training:

'It is having the actual time to do it. There isn't time built in to increase your skill set Universities expect staff to upskill themselves, but don't give them time/money to do so.'

Recognised Researcher - Interviews

Many wanted to devote more time to research skills. Many early career scholars reflected that they didn't leave their PhDs feeling confident or expert in the method(s) they used. Paradoxically, and in turn, they and many of the established and leading researchers reported that they were hired because they were perceived as being an expert in the methods they had used (often leading senior academics/principal investigators to feel that they didn't have to re-train or refresh skills because their 'team' would take care of the data). There is evidence there that even though they have very little time, PhD students have the most time and resources to engage in advanced training and expertise can lead to jobs. If the advanced training came earlier, staring in their master's year, they would have more time to complete lengthy training, practice their skills and have time to write and disseminate their work before moving into post-doc positions. This should build confidence in their data-driven skills by the time they are paid a salary to use them.

Supervisors, it has to be acknowledged, are generally time-poor and operate with a set of expectations which they may not have been prepared for. While structural support could be a solution including training grants or embedding training requirements into annual reviews, recognition is needed that training takes time: and something must be removed from academic workloads for supervisors to manage training to improve their data-driven data skills. One participant reflected in the interviews about potential solutions argued that 'adding in' more activities, was not an option:

'Also have to contemplate the perception of a professional development requirement on staff already stretched and feeling over-managed.'

Recognised Researcher – Interviews

Further, the competitive culture can leave academics unwilling to admit training would be useful.

'There's a bigger point, which is the need to address the needs of supervisors and the anxieties of supervisors. Perhaps building in an element of training or networking for supervisors through the NCRM or updating them and bringing them with you letting them know that

you know, okay, you did your PhD many years ago, you may not be up to speed with data driven research, but we can help you.'

Training Lead – Expert Workshop

It seems academics can be resistant to being trained. One contributor reported that their institution had only recently mandated supervisor training, 'it's still very light touch' (Training Lead – Expert Workshop). The opportunity for supervisors was undermined by this being centralised training and so knowledge specific to the social sciences, for example, exposure to the value of NCRM for them and their students, could not be included as it was not relevant to colleagues from other research council domains.

For some established and leading researchers, the best training for learning new, or consolidating old skills, was teaching others:

'They say if you want to learn something, organise a course to teach! I use STATA and R has become more popular: to learn this I taught undergraduates.'

Established or Leading Researcher - Interviews

This view was echoed by other established scholars, and also those at the recognised researcher stage who felt teaching work enhanced their understanding of foundational concepts. There is a clear barrier here to continued training but perhaps no straightforward solution given the structural nature of workloads are managed at the University level.

To manage the tension, participants reported 'overloading' themselves with learning, and lacking sufficient time to 'play' with their new skill. Thus, advanced data-driven skill learning was difficult to acquire:

'Time is a huge factor. Information overload is a massive barrier. Supervisors/advisors are overwhelmed themselves and do not have a lot of time to support students properly. I think this is a large barrier to students learning and developing new skills. The students don't have any power, and just have to go figure it out themselves.'

Recognised Researcher – Interviews

This issue was felt by contributors at all stages of their career:

'I mean, as a researcher, I'm always faced with exactly this conundrum you know, should I learn this new technique or this new method or this new programming coding skill, or should I wait until it becomes necessary for me. But there's never time, right? There's never enough time.'

Training Lead – Expert Workshop

Funding bids to Research Councils rarely ask for training budgets in order to cost competitively and if they do, these are restricted to the immediate team, so with little impact on the wider culture within the university. Further, the participants reported that usually only one person went to the training, and these tended to be junior members

of the team. While the data-driven skills might be learned using this approach, without a network to discuss and consider the application of learning, the team might not benefit as much as they could, impacting on confidence and ultimately research quality.

The 2015 Guidelines have impacted well because they used the DTP commissioning, and potential of significant funding, to optimise engagement. It could be that using this strategy again may help. Many social scientists, however, operate outside of a DTP framework and so other models of incentivisation could be utilised. The Research Excellence Framework also has a component around Research Environment where universities evidence the learning opportunities available to post-graduate students. If this explicitly asked for examples of sharing knowledge then the ESRC could have access to the carrot, as well as the stick, for sector wide change. Nevertheless, sector-wide recognition that all career stages are time poor and at their limit, is required.

Researchers also reported **frequent barriers to accessing training, including cost, travel distance, and time** (Brown, 2021; Nind et al., 2020). While the increased availability of online training addresses some of these barriers, it can also reduce the quality of the learning experience for some students, particularly those who benefit most from opportunities to interact and network informally with research peers (Brown, 2021; Zauner, 2021).

Access has been a big issue for DTPs during the pandemic:

> 'All our training has shifted to online but we are mindful of lack of access to Wi-Fi for students and study space as well as those with care responsibilities and we aim to provide a wider range of blended learning.'
>
> DTP Training Lead - Survey

All students can struggle to attend training. There are additional barriers for people with caring responsibilities and who work part-time. Around 4% of ESRC funded students are completing their doctoral research part-time (which compares to HESA 2018/19 statistics for all research domains of around one quarter of post-graduate research students being part-time).[15] The knowledge gained by DTPs from delivering online learning should provide more access options for students going forward and careful monitoring should reveal that inclusion improves.

The landscape review revealed that for one large DTP, of the 1467 students who had attended a training event in 2020, 19 students requested an adjustment. Of these 19, 6 required an adjustment due to caring responsibilities whilst 13 related to a self-declared disability (most requiring live captioning or access to transcripts). In 2018/19, there were more than 11,000 postgraduate disabled students researching at UK HEIs (HESA, 2020), equating to around 10% of all students and far lower than

[15] See: https://www.hesa.ac.uk/news/16-01-2020/sb255-higher-education-student-statistics/numbers

the proportion of disabled people in the UK (19% in 2018/19[16]). These figures are likely to be an underestimation as disclosure is still perceived to be 'risky' (Grimes et al, 2019) particularly around impairments viewed as 'learning difficulties' including autistic spectrum disorders and dyslexia. While students can record their disability with their HEI, this information is not shared with DTPs or other training providers such as NCRM and so students must ask for accommodations each time they engage with training, potentially reducing further the volume of requests. A training network that invests in inclusive design will better support all learners (Persson et al, 2015). For example, providing captions as a standard accommodation will also benefit learners who do not have a hearing impairment; and explaining to students the key learning points from visual material will also help learners who do not have a visual impairment. Advances in how we use digital technologies for learning and teaching give unprecedented opportunities for inclusion (Hale & Allam, 2020).

One early career scholar discussed training on R that was designed specifically for hearing impaired learners and included British Sign Language support, described as a 'uniquely excellent experience' (Recognised Researcher - Interviews). This may be another way of imagining the role of a national provider like NCRM of not just providing extended learning opportunities or unusual methods but investing in accessible learning spaces for significant and valued members of our academic community that have traditionally not been served well. Currently NCRM do ensure accessibility but this example shows there is demand for 'higher' access standards to some training events.

Training costs money and this limits data-driven skill training delivery and learning. Though DTPs have the expertise within their networks, they do not have the resources (time, funding, and incentives) to deliver extended training. They rely on good will from colleagues to deliver training for free and often this limits training contributions to less than one day:

> 'I think NCRM with their funding package could focus more on 3 day-3 week training that is needed but which [DTP] can't resource. I have had folk on tenured contracts at [partner] universities turn me down [to deliver training] because NCRM will pay them for the session. That makes our job really tough.'
>
> DTP Training Lead – Survey

NCRM receives a funding grant and charges students a small fee as a sustainability model. ESRC funded doctoral candidates can use their Research Support Training Grant to attend this data-driven skill training meaning that the ESRC is investing twice: through NCRM core funding and through student access costs for training. It is clear though, that this is the kind of training that can transform competency and confidence:

[16] See: https://www.gov.uk/government/statistics/family-resources-survey-financial-year-201819

‘Four-day training was better, because there were four days you could go back and ask questions. The longer period was much more helpful in internalising.’

Recognised Researcher – Interviews

Training often incurs a cost. One DTP routinely charged non-ESRC funded students to attend their training. Another two DTPs charged students who were not studying at one of the partner universities. In contrast half of the DTPs made their training free and available across their region showing a gap in how access has been interpreted. If access is seen as part of the equality, diversity, and inclusion agenda then more could be done to open training up beyond those with grants and access to additional training funds. For impact, this could extend to all ESRC/UKRI investments including NCRM and the UK Data Archive for example. The current structure is effectively producing a data driven skill gap for many non-ESRC funded students.

There is need for approaches to methods training at an advanced level that fits better to support researchers as they are practicing their research; and that allows them to progress their skills. There is no single timeframe that best fits here, students will need different types of training at different stages as their projects develop, as will later career stage researchers.

Of course, face to face learning could also be a barrier and one DTP who provided travel expenses lamented that these could only be paid post-event. The payment helped but prevented those who couldn’t afford to wait from attending. The experience of increasingly online-based training and development, boosted by the demands of the pandemic, should offer a good way forward on this matter. There certainly appears to be an appetite for online and asynchronous learning from early career researchers (Interview data), that may well suit established and leading researchers too.

4.2.2.1 Cost of data and software

The cost of data and software limits data-driven skills. Most software needs to be purchased and this cost can deter students. ‘Core’ courses that use frequentists data and are skill dominant require software purchases that may not be used again. There are alternatives particularly for using numerical data. In 2015, Q-Step centres were pioneering the use of R software, with Glasgow the first to launch R as an alternative to SPSS. R has the advantage of being free, being open to new programming and code development and students have to learn ‘coding’ as a skill (as opposed to the point and click format of SPSS). The success of Q- Step centres has influenced undergraduate and postgraduate social science learning across the UK. Coding makes the mechanics of research design and analysis transparent and is a gateway skill to advanced quantitative analysis and digital skills. Students are keen to be skilled in coding. Coding has value beyond analysing numeric data. It can also be used for example, to manage a large literature review. By drawing on project management skills, students may be better able to see the relevance of a broad-based approach to data-driven skills training.

In 2021 one DTP reported sufficient training needs analysis requests to deliver beginner and advanced training courses in SPSS, Excel, R, Stata, and Python. There is equivalent growth in other areas, for qualitative methods (software such as NVivo, using mobile phones in data collection, digital ethnographies); for teaching and dissemination (software such as Zotero, Trello and Mentimeter, raising social media profiles, mining social media to demonstrate 'reach') and project management (digital copyright, bibliography software and equality and accessibility legislation). Importantly, as identified in Zauner's (2021) report on digital skills, students' competency in using digital tools or software may be restricted by the licences purchased by their university. Here, the use of free and accessible software, such as R makes learning inclusive and becomes a fundamental condition for widening participation. A future iteration of the Guidelines (ESRC, 2015) is likely to include multiple digital platforms and endorsing free or low-cost platforms will impact positively on the student experience.

Another area of cost (but which also dovetails with time) is access to data, and particularly administrative data. ESRC have invested in this area and support access to large datasets. Additionally, some DTPs are able to host 'steer' studentships: the dataset steer; and the advanced quantitative methods steer. This investment is a strategy to develop advanced data-drive skills in two areas considered to require targeted investment. Even administrative data held by UKRI research councils (such as the Medical Research Councils) come with costs and frequently, time-delays. Doctoral candidates who had negotiated the cost of access and a timeframe with the owners of the administrative data during the proposal stage, very often had to wait more than a year of additional time (to what was agreed) and increasing costs (in one case a tenfold increase):

'Admin[istrative] data has a complex process to get access to the data. Many people then just don't have enough time… A PhD student will have challenging time constraints and might need to launch plan B.'

Established or Leading Researcher - Interviews

Such concerns are valid, and students with delays in access will be facing more time-pressure to complete their projects and less capacity to practice their data-driven skills. They may also face a significant gap between data-driven skill training and acquiring data despite planning well and could be required to repeat training. This is one particular case where some scholars are disproportionately disadvantaged:

'The UK has so much good data, and the access issues (and skills to analyse it) are big barriers to its exploitation.'

Wider stakeholder – Interviews

If a solution to transcending the quantitative/qualitative divide is to increase mixed methods training with survey data, then such data ought to be more accessible to PhD candidates.

4.2.3 Barriers related to inter-personal cultures

This section will examine how interactions with others, notably supervisors and peers, impacts on capacity to learn data-driven skills. **A model, where PhD supervisors 'train their students' in how to do research may not be experienced by all and this constituted a data-driven skill gap for some.** In this section we move from training being understood as an organised event, and focus on the individualised sharing of practice between a supervisor and their doctoral student. Presenting one students' view, retrieved through diary circle methods, Nind et al, (2020: 805) reported they felt 'fortunate' that their supervisors were explicitly spending one afternoon working through an approach to analysis. This was echoed in the workshops:

> 'I acknowledge the fact that we cannot have supervisors which are experts in everything. I do feel that somehow, we need to be able to have access to people who can help us, because sometimes I hear from other PhD students, who don't. I can spend the whole afternoon with my supervisor looking at my syntax.'
>
> Amy – Workshop with Doctoral Candidates

This contributor did have some data-driven skill support from their supervisor(s), though felt there was a general lack of support felt by their peers:

> 'It is always a trade-off between your supervisor being an expert on the method and your supervisor being an expert on the topic.'
>
> Amy – Workshop with Doctoral Candidates

One influence of the Guidelines (2015) was to shift culture in some universities, to ensure all ESRC-funded supervisors completed ethics training within the last 5 years. There could be similar scrutiny on supervision workloads and what supervisors 'do'. There are huge differences in what academics 'get' in principle for supervision and greater differences in what they practice.

Doctoral research is an example of experiential learning with the supervisors there to act as assurers and reassurers:

> 'There is no manner of mastering the fundamental principles of a practice – the practice of scientific research is no exception here – than by practising it alongside a kind of guide or coach who provides assurance and reassurance, who sets an example and who corrects you by putting forth, in situation, precepts applied directly to the particular case at hand' (Bourdieu & Wacquant, 1992: 221, original emphasis)

Contributions from Kilburn et al (2014a) and Nind & Lewthwaite (2018a) demonstrate that students benefit from this approach in their data-driven skills learning at earlier stages of their education. Some contributors to this research felt data-driven skill training was not within the domain of their supervisor's role. According to this apprenticeship model, (Carter, Smith & Harrison, 2021), it is within the PhD that the student can learn by doing.

The apprenticeship model, however, relies heavily on the knowledge base and pedagogical talents of the small supervision teams supporting each student. If the supervisors have sound understanding of data-driven evidence building they should be able to co-produce with the student, a useful learning experience. It is the supervisors that teach their students (Hager, 2005) about how to engage and disseminate their work, with an emphasis on thesis production (Hodkinson & Hodkinson, 2004) and critical academic writing (Zukas & Andersen, 2012):

> 'If you have supervisors who have both been in academia for almost a whole career, and it's almost, a disconnect, between what I'm facing as an early career [scholar]. What the situation was like when they started off – there's so much more pressure now, the amount of stuff you need to have on a CV. I've been told, 'you don't need to go to conferences or write articles until you finish with your PhD' and 'you do that afterwards' but I'm not really sure they understand that I need some sort of financial job after.'
>
> Amelie – Workshop for Doctoral Candidates

The culture gap between doing a PhD 'then' (when supervisors graduated) and now, was mentioned by many doctoral candidates and recognised researchers as an issue. Students are dependent on their supervisors and discussed needing 'permission' (Workshop for Doctoral Candidates) to publish or prepare publications, to attend conferences and for placements. A lack of encouragement in such endeavours resulted in lost opportunities. The chapter will next consider the value of peer networks, and this may be one strategy towards closing the culture gap.

Isolation can be considered an issue limiting data-driven skill learning. In relation to the broader learning experience, research students have reported peer support and networking as important avenues for accessing help with methods, digital skills, and general emotional support and reassurance when faced with research challenges (Brown, 2021; Nind et al., 2020; Zauner, 2021). For example, learners may not require a full course or refresher of a research method or digital tool but can receive informal tutorials or troubleshooting advice from peers (Zauner, 2021). This idea was highly valued by the doctoral candidates who participated in the workshops.

In Kilburn et al.'s (2014b) report on short courses in advanced social research methods, methods teachers reported 'co-learning' as a way to address potential disparities in learner skills and experience on a course. Co-learning in this sense involves more experienced learners using their skills to support those with less experience:

> 'Having direct contact to someone who as a little bit, you know, a couple of steps up from you? Or someone who you can speak freely to, because I like learning from other people and I like to save my time and save myself from failure.'
>
> Miriam – Workshop with Doctoral Candidates

As argued, the pressure on data-driven skills teachers is significant and large class sizes, prior/no prior learning and pan-disciplinarity impact hugely. A potential strategy

to accommodate variation in skills, could be to invite those with prior learning to contribute their perspectives and expertise via a co-learning approach. As doctoral candidates employ data-driven skills that others do not possess (particularly around digital skills), there is an opportunity here:

> ‘In terms of contribution to the department, if you learn new methods that no one else in the department knows about, why not have, once every semester, let’s have a round of what the PhD students have learned through the advanced training.’
>
> Amelie – Workshops with Doctoral Candidates

It’s possible that this could be harnessed as a strategy, with students attending expensive and extended training courses being provided with a fee waiver if they develop training for peers. If recognised (for example, status as an ESRC training fellow) doctoral candidates may be attracted to the role of peer mentor. It should be worth their time investment as they increase their confidence in the skill and consolidate their learning. It is worth sounding a note of caution though: supervisors and PIs are paid and may be expected to perform this role. Asking colleagues or students at the start of their career to do this work for free, or less remuneration is exploitation. Senior academics could influence a culture shift to adequately recognise the mentoring role as vital teaching leadership, and this should be a precursor to the formation of effective peer networks.

Although not a specific focus or frequently found element of the literature, forms of team- based learning were frequently alluded to as a pedagogical tool. In Nind et al.’s (2020) study on students’ perspectives on learning social research methods, learners valued the dialogic aspect of learning, which could occur both as part of formal teaching and in more incidental learning encounters. Having the opportunity to engage and discuss with students or researchers from different disciplines allowed learners to process their research experiences and learn from each other.

This is supported by focus group findings from Brown (2021) and Zauner (2021) that social science postgraduate students find significant benefit from the incidental encounters and networking opportunities at training events. In our workshop with doctoral candidates, access to a peer network was viewed as an exciting evolution where peers could bounce problems off of each other towards solutions (particularly if this revolved around a data-driven skill or methodology such as coding, or discourse analysis). There was strong support in the student workshops for a peer-learning network that revolved around data-driven skill; topic; or particular barrier to learning (working on an interdisciplinary project, or a collaborative one, or being part-time). One doctoral candidate wanted to meet others who used R coding, to help them problem solve through errors in the script:

> ‘If a page [online platform] existed and someone could go in there and collaborate. These people exist, we just need the way to be able to find each other, … if I had someone who knew how to read my syntax and help, I would have been moving so much faster.’
>
> Amy – Workshop with Doctoral Candidates

Students have also reported valuing the opportunity to network and collaborate with more experienced researchers (Kilburn et al., 2014b). This approach can be seen as a more peer- based and informal version of supervision and mentoring. While small group work was viewed as useful and valued by learners, teachers found it challenging to incorporate with time constraints, particularly with larger course groups (Kilburn et al., 2014b). Providing some space or opportunities for ESRC funded students to network early on in their studies, particularly if structured around a topic or data-driven skill should enhance advanced, informal learning of data driven skills.

Team-based learning is the antithesis to the apprenticeship of doctoral research (Park, 2005; Carter, Smith & Harrison, 2021) that requires lone-working leading often to isolation. Peer- support networks can help students feel connected, which in turn can improve well-being, and then help submission rates. Submission rates can be understood as a proxy indicator that students have an appropriately sized project (it can realistically be completed during funding) and that they have the motivation and energy to persist through the intellectual (and admittedly structural) challenges they will face.

Peers can form networks that are maintained throughout a person's career and could be one pathway to impact as some scholars migrate outside the academy upon graduation. Such networks are valuable for PhD students (Sweitzer, 2009; Jairam & Kahl, 2012; Hernández- Hernández & Sancho-Gil, 2015), not just for skill development. Peer networks also work to navigate through the wider student experience and support engagement and motivation with the research (Janta, Lugosi & Brown, 2014). They are also useful for challenging imposter syndrome (Nind et al, 2020). This approach is not teamwork in a controlled sense, with students orchestrated to deliver an activity. Rather it is a natural phenomenon of friendships that benefit those able to form them. The ESRC guidelines (2015) have steered doctoral training partnerships towards 'cohorts' to create fertile spaces where informal networks can germinate. Learning data-driven skills can be easier if working in a group. It can be reinforcing to see others struggle and recognise that some learning is difficult. It can also help to teach others the things we are getting to grips with.

The link between data-driven skills training, relationship building, and leadership, is a tentative one. In imagining new ways of learning data-driven skills, it is valuable to consider gaps around team-working generated through all phases of data collection. Here we argue that training in working with others, together with practical experience working with others, may help with learning data-driven skills. Related to gaining experience developing communication skills and networking: DTPs had seen an increase in requests for training around leadership and particularly around the leadership skills required when working with teams of researchers. There was some evidence that post-graduate students were making these opportunities for themselves, forming research clusters, and creating student-led activities. Such initiatives may not be recognised by early career scholars as examples of leadership. Recognition of events and networks as a signifier of leadership could help students understand and develop their profile. Some DTPs were aiming to provide this:

‘If a student facilitates training for us we write to them making it clear how their achievements are valued by the academy, carefully using language from promotion criteria for example around leadership and influencing others. We also have an Expenses Policy that ensures any PhD or ECR facilitator can be paid for their contribution.’

DTP training lead - Survey

Helping students explicitly see their value is a form of enabling practice, providing a fertile context for co-production and confidence building. Peer-networking has evolved, and the Guidelines could move from a vague expression of cohort building, to guide DTPs towards communities and networks themed around data-driven skills, around disciplines and around well-being. Further, such an initiative, could run optimally across the UK.

4.3 Identifying interventions at the doctoral level

The evidence presented has demonstrated the indivisible impact of structural barriers and training gaps. As discussed throughout chapter 3 and consolidated here as context before presenting a re-imagination of training, the Guidelines should evolve to include more explicit reference to project management and digital skills. Specifically, the Guidelines should require a broad-based conceptual learning and include the following aspects:

- Champion the use of free software and open data as much as possible
- Deliberately address building capacity around digital skill training, specifically:
  - Digital research design
  - Data capture, curation, and storage
  - Intellectual property, safeguarding data and legal expertise
  - Research integrity training that includes all forms of data and helps students demonstrate rigour in their work
  - New pathways to dissemination and publication
- Highlight training on building dissemination strategies inclusive of publishing open data; using social media and peer-review articles. The Guidelines could include critical writing and literature reviews into ‘core’ training
- Conceptual understanding of open access and skill development in cleaning data files; fixing data inconsistencies, generating data in multiple file formats for preservation and dissemination and generating code lists should be added to core training
- Conceptual understanding of AI and machine learning. Skills in these areas are not required in core learning
- Research integrity training including reproducibility and incorporating standards relative to text or visual data, for example transparency and credibility
- Learning on ethics, reflexivity and standpoint relating to all forms of data and inquiry
- Championing spaces for teamworking, dissemination strategies and peer-learning to enhance learning experiences and help students prepare for research or academic ‘practice’

- Promotion of assignments that focus on conceptual understanding, transcend the quantitative/qualitative binary, and are challenge-led
- The integration of data-driven skills and disciplinary learning by explicitly identifying the dissertation as a site of learning and championing embedding into disciplinary classes.

If the Guidelines are too prescriptive, the new delivery will be static and rigid. Further, a decision is needed about whether this change is delivered by DTPs or by HEIs (straightforward for those with one, or a small cluster of HEIs located close by each other; will place additional pressure and demands on educators and students who belong to a larger DTP and/or where the HEIs are geographically dispersed). Unless the change impacts on HEIs there is a significant risk of perpetuating a two-tier system where ESRC funded students have access to resources that others do not. This will not deliver the step-change required to close data-driven skill gaps for most social scientists. HEIs could make changes sustainable by investing income earned through methods learning, into methods learning. Courses with larger student numbers raise more money for university departments than smaller course. The large research methods classes that use small teaching teams, are producing profit, that is invested outwith the methods-learning spaces. It is a different approach to disciplinary learning where smaller classes are used, often with number-caps to better control teaching workloads and the learner experience. University managers could make a change to ensure finances raised by methods learners are invested in methods learning. Re-modelling investment could support larger teaching teams with workloads that accommodate course refresh and greater engagement with pedagogical literature than is currently permitted. In turn larger teams could then accommodate smaller class sizes. HEIs will need encouragement and the front-loaded ESRC investment together with stronger reporting pathways (for example to the REF scrutiny of the Research Environment) will help.

Data-driven skills educators need time to engage with these recommendations and the pedagogical literature to re-imagine a new delivery of training. This re-imagination should also be built on debate with colleagues and over a period of time that allows experimentation, capacity-building and engaging with student feedback. Delivering training over intensive- courses rather than semester-long will improve learning experiences.

The gaps identified in this chapter are more ambiguous and nebulous than those featuring in chapter 3 yet there are ways the ESRC's Guidelines could evolve to a) recognise these challenges as gaps and b) present a framework of doctoral study that will support advanced learning. Here, 'advanced' describes the data-driven skills required by a student as they advance in their doctoral studies and may mean a re-visiting of earlier learning.

The Guidelines should evolve to include post-core expectations for advanced learning that:

- Move to a 'skills' focus, reinforced by conceptual learning and champion data-driven learning experiences where students can use their own data, or have space to reflect on their research-context during training to help them become a specialist

- Encourage peer-networking spaces that span the UK
- Supports researcher safety and well-being strategies that advance with doctoral researchers
- Make Supervisor training and development mandatory and includes awareness of ESRC investments and the training landscape, well-being support and guidance on creating and maintaining supportive supervisions. Championing use of the UKCGE provision may help.

The ESRC should adopt a single framework to aid promotion of training opportunities across their investments. Encouraging HEIs and the other Research Councils that form the UKRI will maximise the efficacy of this model and significantly improve data-gathering potential. The framework should capture how challenging the learning will be, whether prior learning is needed and what engagement-level is required by students. The framework should make it explicit if recognised, established and leading researchers are welcome.

A single site, regularly updated, curated, and broadcast to help researchers at all levels learn about training delivered through ESRC investments should be created. Developing this into an app, or having a website that can be used on mobile and handheld devices was seen as particularly useful by the doctoral candidates who participated in the workshops.

Finally, how we disseminate data and findings is changing and working with REF frameworks and strategising across the UKRI landscape to lead to greater recognition for outputs beyond peer-review journal articles is recommended. Recognising open data, open learning materials and blogs for example. A REF framework that awarded 4* to these kinds of outputs when produced by doctoral candidates and recognised researchers would allow them to demonstrate academy-readiness and industry-readiness. This will particularly support capacity building around digital skills.

4.4 Changes (and the scale of change) needed in the curriculum

The recommendations begin by re-imagining core learning. Data-driven skills will be better delivered through a foundation course, a new ‘core’ course with mixed-methods presentation, specialised training incorporated into the master’s year, and more explicit use of the dissertation as a site for data-driven skills learning.

4.4.1 Foundation courses

A major barrier is learning in large classes with diversity in prior learning. To avoid repetition in data-driven skills learning in the master’s year, a foundation course could be completed by students with no prior learning, or by students wishing to rehearse and consolidate prior learning. This will be taught intensively, with a suggested duration of two weeks, based on content covered and providing time for networking and skill practice.

The content proposed is informed by a landscape review of undergraduate methods training. This course is not credit-bearing but should include formative assessments to help students gauge their learning.

Content should focus on skills, though conceptual learning will be useful too:

- Introductions to ethics
- Ontology of research methods and epistemological pathways
- Levels of measurement and sampling
- Univariate and bivariate analysis with numeric data
- Data production such as survey design, interviews, participant observation and focus groups
- Using documents as data
- Thematic analysis with discursive and narrative data
- The fundamentals of data literacy.

Undergraduate data-driven skills learning is advancing, and new QAA subject requirements may further improve provision at this level. This raises the importance of systematically conducting and revisiting TNAs undertaken with students, that can also capture prior or work- based learning, at the point a studentship award is made. The TNA should also record the anticipated specialised training required for the PhD (see section 4.4.3 for the rationale). We support the Review of the PhD's (Tazzyman et al, 2021) recommendation that DTPs appoint professional development leads. Working with an academic training lead at DTPs, these new appointments could support this work and build relationships with students at the start of their PhD journey, optimising their capacity to work as a guide.

As many of these students may have undertaken their initial training outwith the UK, or have an interdisciplinary focus, or face other barriers to learning, this course could be offered centrally by the ESRC giving students the chance to network with others who face similar challenges. To manage numbers, students could choose between face-to-face and distance-learning presentations and thus accommodate those who work part-time or who face barriers to learning away from home.

There are feasibility issues with this proposition, and these will be discussed in section 4.4.2 (below) and in chapter 5.

4.4.2 'Core' training

Removing the foundation material will allow students to move to advanced learning in a new 40-credit module. Advanced content then could focus on project management and include a broader range of data collection and analyses approaches that build confidence and range in the application of data-driven skills. Intellectually, this course should be as demanding and challenging as master's disciplinary courses.

Emphasis on conceptual learning should ensure that:

- Students have greater understanding of how epistemological choices will impact on findings produced

- Students recognise the importance of data/project management and dissemination strategies including ethics, research integrity and demonstrating rigour. This learning should relate to traditional and digital approaches.
- Awareness of and understanding in AI and machine learning as a research tool is established
- Using existing data such as surveys, existing repositories of visual and text data; social media; big data is promoted

In terms of data driven skills, it is anticipated that students will gain understanding in:

- The appreciation of theoretical frameworks that underpin data and analysis
- Curation and analysis of numeric, text and visual data with appreciation of the potential of using archives and secondary data
- Application of data ethics including digital data; reflexivity; and standpoint
- Fundamentals of coding – if applied to project management tasks such as creating a map of literature reviewed, this will help engage students not wishing to pursue a career analysing numeric data
- Analysis approaches such as regression, and appreciation of the potential of using specialised analysis such as spatial, network, econometrics and/or structural equation modelling)
- Use of specialised analysis approaches such as discourse analysis, phenomenology, and critical realism
- Maintenance of emotional well-being and the importance of, and strategies towards safeguarding-self
- Development of skills in digital data capture, curation, and storage, safeguarding digital data including awareness of rules around intellectual property
- Open data access issues and skill development in cleaning data files, fixing data inconsistencies, generating data in multiple file formats for preservation and dissemination
- The building of dissemination and impact pathways. If this learning draws on text/visual analysis as a means of 'controlling the narrative' of dissemination strategies, it is likely to engage students not wishing to pursue a career analysing text/visual data.
- The use of a wide range of analytical software (R, Nvivo for example), championing the use of free software as much as possible.

To engage with the wealth of talent and expertise, it is recommended that existing research development teams working in HEIs (often centralised teams working in professional services rather than academic/disciplinary units) form teaching teams with existing data-driven skills experts to re-imagine how this course will work for each HEI or DTP.

Learning will still be broad-based and aim to deliver fluency, though it should focus on conceptual learning. The assignment(s) should be challenge-led and if possible, use real-world data. A portfolio that captures reflexive learning around project management and dissemination skills is key, though students could personalise their work by drawing on understandings of data collection and analysis. This will be strengthened further if students can use this to bridge towards their master's

dissertation and planned doctoral research though this will require greater investment in teaching teams, discussed further in section 5.3. This module should be mixed methods and use numeric, narrative, and visual data. Where possible/meaningful, a new course should resist using terms quantitative and qualitative.

Like the Foundation Course, new core training should be delivered in a condensed presentation potentially, over three-four weeks, rather than over a semester. Pedagogically this would be a good strategy. The doctoral candidates we spoke to championed this approach and there was support from many recognised, established, and leading researchers. Digital advances have improved access to training for many, yet face-to-face is considered to be the most valuable form of teaching and learning.

The class size should be kept manageable and broadly be in line with disciplinary class sizes. Further, class size should enable spaces for learner-teacher interaction and student-led pedagogy (usually ruled out by larger groups or mass lecture approaches). Space should be provided for small teamwork too. If demand exceeds 50 students, then multiple classes should operate separating students into groups of allied-disciplines. Multiple classes should not be taught by the same people or workloads will be unmanageable. It is not repeating the lecture that takes the time, but creating learning pathways to individual students and responding to their ambitions and concerns.

If this re-imagined course is offered by HEIs, then it should be accessible to most (a residential should not be needed) but there will be more pressure on class sizes. The scale of change required, and the number of ESRC studentships available, may mean that it should be offered regionally, either by a DTP or by a small cluster of DTPs. This would require some/all students to attend a residential presentation. It is vital this course does not exclude or marginalise students or educators with impairments, with care responsibilities, who are studying part-time or students/educators who face other barriers to attending a course over this length of time. As summer schools and conferences are routinely run on this type of presentation, it is imagined that barriers can be removed. Evidence they can be removed is needed before action is taken. One solution might be that this course is offered by one DTP/university as a distance- learning option. The feasibility issues are discussed further in section 5.3.

Whether the presentation is live, residential or distance-learning, the intensity of this model requires a pedagogical shift and time investment is required to produce high-quality learning materials. Specifically, we recommend production of new high-quality online and in person learning resources, with emphasis placed on blended learning that includes interactions, masterclasses, mentoring and post-course support (Durrant et al, 2015). Ideally the teaching team should be available post-course and could run monthly ‘surgeries’ for students to seek advice as they develop their dissertation plans.

4.4.3 Specialised training module

As discussed, core training is currently delivered over 60-credits. The re-imagined core training is a 40-credit course. This leaves 20-credits which we recommend is

used by students to generate a specialised training portfolio. This strategy will facilitate advanced training and development driven by personal needs and underpinned by previous learning.

This course is fundamental to the re-imagining of data-driven skills training and its ambition and is geared towards closing data-driven skills gaps, as well as providing a foundation for doctoral learning. This course should:

- Champion the use of free software and open data as much as possible
- Require each student to complete four courses that map directly to their dissertation and PhD research plans
- Ideally use courses produced by ESRC investments such as DTPs, NCRM, CDTs and UK Data Archive for example. They could use courses offered by HEIs

The four courses completed should include at least one data-driven skill and one broader transferable or research development skill. The PhD supervisor (and dissertation supervisor if different) should help the student make the decision and the DTP training lead and/or the DTPs professional development lead can offer advice and sign-off for ESRC funded students. For equity, courses should combine to offer between 16-32 hours instructor-led training.

Figure 12 An imagining of the Personalised Advanced Training in practice

For example:

A student preparing for emotionally-challenging overseas fieldwork may select:

- Interviewing & Sensitive Research
- Ethics, Well-being & Researching Overseas
- Object Elicitation
- and Translating Transcripts
-

A student preparing to work with administrative data could identify:

- Data Management and Cleaning
- Linking Administrative Datasets
- Structural Equation Modelling introduction
- SEM Intermediate
-

A collaborative student working with a civil society organisation may choose:

- Participatory Action Research
- Inclusive Research Practices
- Reflexivity
- Using Social Media for Research Dissemination

Examples are provided in Figure 12 (above). Note in the second example, a student could join a single course that covers introductory and intermediate and if equivalent to 8-16 hours of training, could be considered as delivering half of the requirements. This kind of flexibility is vital to accommodate students and will also benefit some

disciplinary fields where the depth of data-driven skill learning is critical (this may include but is not restricted to economics and anthropology).

In order to link this training to the masters, students could participate in this training and then produce a single assignment to be assessed by their dissertation supervisor. HEIs will need to create this course so that students can earn credits with the submitted assignment (it is the assignment that earns the credits rather than individual courses assessing the students' attainment). This could be a reflective dissertation proposal that critically evaluates their planned approach to data production and analysis; ethics and research integrity; and potential dissemination pathways (sharing data and findings).

There are feasibility issues and larger DTPs will be unable to accommodate this innovation for non-ESRC-funded students. As discussed, universities could host this innovation with greater delivery of internally provided specialised data-driven skills training.

The TNA conducted at the point that studentship awards are decided should include an indication of what specialised training would usefully be completed before the PhD. For 1+3 students this is conducted before the master's year begins. This provides the time required to allow DTP training leads and professional development leads to start planning their training programmes. It is critical that students are encouraged to revisit their choices at the end of the core training. Therefore, the specialised courses should be completed post-winter break assuming an autumn start to the masters to give space for students to change their mind, and space for training providers to meet demand. For this to work optimally, all DTPs should use an agreed TNA template. This module is part of a strategy to increase awareness of the training available from ESRC investments that can support doctoral research.

4.4.4 The dissertation

Remarkably, through the data collection conducted, the master's dissertation was not identified as a site of learning data-driven skills. This may be due to the supervision model of a single academic guide, and a strong disciplinary emphasis to the dissertation research. More could be done to flag to students the opportunity provided by the dissertation in terms of developing: project management skills; project design and data-driven skills. Students should be encouraged to consider using the dissertation as a pilot to help them build confidence ahead of their PhD research and to think strategically then about what challenges the dissertation should provide.

One strategy to help students identify their dissertation as a space for practicing data-driven skills is to provide an assignment to the specialised training module that explicitly includes dissertation planning as suggested in section 4.4.3.

4.4.5 Re-imagining 'advanced' training

The re-imagining of training delivered through the master's year should yield significant results. Students beginning their doctoral research should have greater conceptual fluency of knowledge production options; and so be more able to defend

their selected data-driven skills, will have secured specialised learning relevant to their PhD and practiced autonomous and strategic decision making in delivering their dissertation. Their training is not complete. A framework is required to underpin continuing development around data-driven skills. This section outlines some principles to inform this structure.

Awareness of training available in the UK is low, despite training from ESRC investments being substantial and of high quality. The specialised training module will improve awareness. A key strategy will also be supervisors whose awareness could improve and whose influence is significant. Mandatory supervisor training that includes detail on the training landscape is needed.

Developing an app, or website that can be used with handheld devices, that provides access to training provided by ESRC investments will deliver a co-ordinated display that allows social scientists to easily search. Information about training should cover what level the training is pitched at, who is eligible, how the training will be delivered as well as duration and location. It is recommended that time be taken to determine how to make training accessible to disabled students, and to employ strategies regardless of whether a student has declared an impairment. Training should be promoted declaring accessibility and declaring where it is not accessible. Using a typology to manage this that is used by all ESRC investments (ideally all UKRI investments) and HEIs will make it easier for users to evaluate the best training for them. It will also help ESRC produce data on added value to the academic community (for example, percentage increase in training on machine learning across all ESRC investments).

There will continue to be demand for training in digital skills, AI, and machine learning. To build capacity, a strategy that mobilises recognised researchers and doctoral candidates is recommended. There are wider structural issues that can undermine career trajectories for those who teach data-driven skills and mindful of this, opportunities to lead training in this field should carefully consider how investment will lead to recognition and reward. All students should be strongly encouraged to improve their digital skill working.

A structured framework to doctoral training should be included in the ESRC Guidelines that encourages engagement with training around dissemination strategies, building on learning delivered in core training. Couched as being industry-ready and academy-ready, this training should include advancing skills around open access data; social media strategies and producing academy-ready outputs such as book chapters and peer reviewed journal articles. Spaces where students can learn from each other could be invaluable here. Moving away from the concept of training being led by an advanced educator, here we use peer-learning as a way to form networks that advise and support.

Peer-learning is also a strong strategy to improve well-being but it’s vital to also structure learning in this area to communicate clearly that this is not an individual issue but linked to structural issues around precarity.

5. Recommendations for implementation and conclusion

Summary

Digital skills, dissemination strategies and project management are the three significant areas where an intervention is required. Alongside this, we identified structural barriers that, if challenged, will also impact positively on data-skill training. Removing these barriers is tantamount to a culture shift and this is required to deliver the step-change required. This chapter recaps on the main conclusions from our findings and presents data informing 5 factors that could impact on the implementation of our recommendations.

Front-loaded investment in a re-imagining of 'core' data-driven skills coinciding with the first half of the DTP2 is recommended. We recommend that the Guidelines evolve to trigger a change and an evolution of 'core' training. For this to be feasible, space must be available for teaching teams to develop their version of optimal environments for learning data-driven skills. If the Guidelines are too prescriptive, the new delivery will be static and rigid. Further, a decision is needed about whether this change is delivered by DTPs or by HEIs (this is straightforward for those with one, or a small cluster of HEIs located close by each other, but will place additional pressure and demands on educators and students who belong to a larger DTP and/or where the HEIs are geographically dispersed). Unless the change impacts on HEIs there is a significant risk of perpetuating a two-tier system where ESRC-funded students have access to resources that others do not. A two-tier system will not deliver the step-change required to close data-driven skill gaps for most social scientists.

HEIs could make changes sustainable by investing income earned through methods learning, back into methods learning. This could support larger teaching teams with workloads that accommodate course refreshes and greater engagement with pedagogical literature than is currently permitted. In turn, larger teams could accommodate smaller class sizes. HEIs will need encouragement and the front-loaded investment together with stronger reporting pathways. For example, the Research Excellence Framework (REF) could help by firstly accepting a broader range of outputs including teaching materials and open data. The REF could also evolve its Research Engagement element to measure progress towards stronger teaching and learning experiences. This could include reporting on class sizes, embedding and advancing digital skill learning.

Further work would usefully consider how best to standardise the Training Needs Analysis. Related to this, work to standardise the ways in which training is promoted (skill level but also presentation type and suitability) and defined will improve measurement tools (for example, permitting measurement of the uptake of digital skill training) and student engagement.

While this research captures some of the structural barriers facing recognised, established, and leading researchers (time, workload, access issues), we champion a more focused enquiry and future work in this area. This would help the ESRC and UKRI focus on how to deliver the promise of a research career outlined within Vitae's

Concordat[17] (2019), and Research and Development People and Culture Strategy[18] (2021).

5.1 Introduction and purpose of this chapter

This chapter moves from considering what data-driven skill gaps exist, and how the Guidelines could evolve to capture this learning, to consider how HEIs and DTPs could implement the recommendations. Ultimately, the aim is to reform data-driven skills training so that it produces confident social scientists, who can lead on the production of robust and rigorous research. The chapter closes with a view of future work. Investment in post-doctoral career support for (re)developing data-driven skills is recommended.

5.2 Finding: Culture shift required

In conclusion, the aim of this research was to evidence the data-driven skill gaps with particular emphasis on doctoral education in the UK's social sciences. There were only a few gaps identified ultimately: with digital skills, dissemination strategies and project management being the three significant areas where an intervention is required. Sections 3.4 and 4.3 recommend changes to the Guidelines to meet these data-driven skill gaps.

During the data collection and analysis, two more areas of intervention appeared. First the delivery of data-driven skills training during the masters is not producing students with specialised skills directly relevant to their planned PhD research, or with confidence in their skills. Issues with repetition, mixed ability classes and learning skills considered irrelevant, were key. A mixed methods presentation will enable students to build connections between methods. The use of a Foundation Course will stream-line teaching and students with no prior learning can focus on skill building. Then core training can focus more on conceptual learning and a challenge-led assignment will bring students together to practice not just data-driven skills but practice 'being a researcher'. The innovation of a specialised training module will deliver confidence-boosting learning in time to be practiced with a dissertation and engage students with the wide range of training available to support them through their doctoral research, from ESRC investments.

Second, a culture shift is required across the sector. Doctoral candidates, recognised, established and leading researchers reflected on inadequacies and barriers, often individualising these and assuming that others had access to 'better'. Across the data, we can see that the problems are not individual but structural. No one has time, or resources to invest in data-driven training. Educators don't have time to engage with advances in the literature around methodological teaching. Disinvestment in courses, leading to larger class sizes and smaller teaching teams signify that a culture shift is needed. Data-driven skills training is 'core' and should be invested in, to at least equivalent levels to disciplinary courses. Because of the years of disinvestment, a front-loaded model with an injection of ESRC investment, that is

[17] The Concordat to Support the Career Development of Researchers — Vitae Website

[18] R&D People and Culture Strategy (publishing.service.gov.uk)

then made sustainable by HEIs adequately supporting teaching and learning, is recommended. How this might work in practice, is the focus of the next section.

5.3 Recommendations for implementation: Re-imagining data skills teaching and learning

The culture shift required by these recommendations is significant. This section will consider the practical issues and will outline perceived risks and is grounded particularly in data produced by the sense-checking workshops. The section will then turn to feasibility and the decisions that are triggered by the move to condensed learning and support of larger teaching-teams.

Risks were identified during the sense-checking workshops and flagged by the student and training experts involved in the workshops:

> 'It could easily be made worse rather than better if we're not careful.'
>
> Training Lead – Expert Workshop

Five areas of our recommendations were viewed as potential sources of risk and requiring careful management. Each of these is taken in turn to specify the potential risks and how these can be mitigated:

- Investment in training teams
- Avoid creating a two-tier system
- Moving on from the qualitative/quantitative binary
- Managing specialised training module and
- Advancing pedagogical skill to deliver condensed learning

Firstly, investment in data-driven skills educators is vital and urgent. The re-imagined training model is a major departure from mainstream and traditional delivery for many educators and time to engage with the pedagogical literature and resources to produce new learning materials is required. Further, recognition that core data-driven skills teaching is more demanding than disciplinary teaching should be reflected in continued investment and changes to workloads. This requires a culture shift away from large class sizes and needs adequate workload modelling for teaching teams. Without this, the significant changes proposed are unlikely to succeed.

Second, while it may be useful to ensure new learning experiences prioritise ESRC-funded students, this should be considered as a pilot with commitment to roll-out delivery to include all students and avoid, a 'two-tier system' (Training Lead – Expert Workshop) and effectively widening the data-driven skills gap between students with ESRC funding and those without. Reflecting on their masters' experience, one expert recalled how important it was that they could choose to fund a masters and learn alongside ESRC-funded students:

> 'What would have happened to me today, if I'd been doing the, you know, the 'loser' course.'
>
> Training Lead – Expert Workshop

Separating students by destination (PhD or exit) at the master's level will control class sizes but will also miss the opportunity to deliver a step-change valuable to those moving to employment. Many of the data-driven skills identified in this research as vital, are equally relevant to industry and academic sectors.

In considering a phased approach, the new model proposed, could work with a centralised offering, a single, or regional course that brings together ESRC students. This could work in a pilot phase as a means to generate new learning materials that can be shared widely. There was concern that this would trigger the two-tier gap identified above; would impact negatively on accessibility and:

> 'There's also a bit of a risk of, as an institution, losing control of the training. You have to trust in the fact. You're not quite sure if what [was taught] and who taught it and how they taught it, as an instructor, was what you want them to learn.'
>
> Training Lead – Expert Workshop

If a centralised offering is to be resourced, then built within the planning, should be a move to scale it up, to be as inclusive of non-ESRC funded students as possible and engaging the expertise of many data-driven skills educators as is possible.

Third, while existing teachers will be a strong fit with this model, care must be taken to deliver and value mixed methods teaching. It will undermine the approach for existing qualitative teachers for example, to only teach parts of the new course that directly map with current teaching. If this happens, then the qualitative/quantitative binary will reappear. One option is to consider appointing new educators, freeing up existing team members to produce specialised training according to their expertise. In fact, this could add value. If HEIs produced master's courses where students could choose to invest their 'free' credits in disciplinary or data-driven skills, then something like the specialised training module could be delivered to non-ESRC students, as training would be available 'in house'. Here the skill and expertise of traditional educators could prove invaluable as HEIs produce new advanced methods courses that build on core training in a particular methodology. For example, a 20-credit course on ethnography, social network analysis, participatory action research or archival research will strengthen the options available to students selecting their masters' curriculum.

Fourth, this model achieves a great deal with the same credit-load as is currently used. Communication is key to reinforcing this. It will be easy to interpret this model as providing additional pressure on students for two reasons: there is an extension of data-driven skills requirements notably digital skills and because this new approach makes explicit the need for training around project management and dissemination. If communicated well, the approach should encourage students: this was our experience of the student workshops. The use of the foundation course, the use of the 20-credit assignment to support specialised training and the recognition that learning needs to continue through the PhD and beyond are critical. Without clear communication there is a risk that students feel more pressure:

‘There is a lot of things that we want out students to do, and not enough space, and there seems to be kind of ways, to try and shoot more stuff into them.’

Training Lead – Expert Workshop

Related to this is the need to present the core masters learning as an anchor to support advanced learning as the students need it:

‘Students, when we give training very early on in their doctoral careers, they don’t yet see the need or the benefit from it, and it’s really only once they’ve hit the problems that they come back and are eager to go for the training. So, I think front-loading everything is not the way forward.’

Training Lead – Expert Workshop

Finally, condensed-learning periods need careful investment of pedagogy, and care is needed to check accessibility. Intensive learning weeks may be less accessible for some learners including part-time learners, those with caring responsibilities or impairments. As discussed, providing students with the option of a distance-learned delivery will be critical, and this brings further pedagogical challenges:

‘Based on my own experience, teaching people for eight hours in front of a screen for a whole week, you’re going to be, by Friday, bending their mind.’

Training Lead – Expert Workshop

A part-time contributor responded to the challenges of intensive learning:

‘For me, I would find it really challenging to get one week off from work.’

Miriam – Workshop with Doctoral Candidates

And then reflected on the challenges of semester-long learning:

‘The teaching model is covering two or three different subjects in one day, for me that was difficult in terms of trying to find that time to prepare for three different subjects. So my head was jumping from one thing to another. I think there is a lot of psychology around how we absorb information and how we learn. I don’t think our brain is capable of switching to one task to another in such a small period of time.’

Miriam – Workshop with Doctoral Candidates

Here the challenge is exacerbated by ‘silo-ing’ quantitative and qualitative learning into distinct courses. Providing condensed learning opportunities over week(s) rather than months is pedagogically strong, though care will be needed to construct either distance learning courses that can work with other priorities; or sufficient notice of condensed learning to allow planning and removal of learning barriers.

The new configuration does not remove or reduce methods training. It still champions the kind of repetition that consolidates key learning. It does remove the repetition that is impacting negatively on confidence around data-driven skills. There are 60-credits in the masters that are not core methods or dissertation. From the evidence collected, we recommend that students have both disciplinary and advanced methods options to choose from. This will require many HEIs to increase their provision of advanced courses. As one contributor reflecting on their choices said:

> 'I would have loved to have been able to do a larger proportion of the credits in research methods courses. Because I already had an undergraduate degree in [discipline] and I didn't need to do another [social inequalities] course. What I needed at that point was to spend more time doing that intensive training to build my confidence.'
>
> Training Lead – Expert Workshops

Retaining disciplinary options enables interdisciplinary learning and may be particularly beneficial to overseas students or even those moving between UK universities where courses on offer are distinct thematically from their undergraduate learning. Improving access to advanced training was requested by the expert students during their workshops:

> 'I asked as well and the [University] used to have an advanced quantitative course but not anymore.'
>
> Rehana – Workshop with Doctoral Candidates

And from another:

> 'I asked about something called the advanced quantitative methods add on to the 1+3 when I went for my interview but it wasn't available in the majority of pathways'
>
> Mary – Workshop with Doctoral Candidates

Advanced courses may not attract sufficient students to operate, unless HEIs within a DTP or region work together to allow access to learning across the partnership. This needs monitoring, as the new core training impacts and confidence in data-driven skills builds there may be more demand for these kinds of learning opportunities.

Though it didn't appear as a concern in the workshops, the new timing of the foundation course and the new core learning will need to be considered carefully. The move to intensive learning has implications for other learning which is likely to remain using semester-long structures. This may require some 'squeezing' of learning before the winter-break into 8 weeks, but there'll be less demand on the students' time through these 8 weeks and so the disciplinary learning quality could improve. As each HEI structures their learning differently it is not possible to impose a one size fits all onto how the disciplinary courses will respond to the intensive learning model proposed for methods.

These two courses should be less expensive (two-week (estimated) Foundation course plus four- week (estimated) core course equates to a 6-week stipend and fee investment) compared to the current model of a fractional award (6-months stipend and fee investment) or a 1+ masters (12-months stipend and fee investment). For students with a master's qualification that includes a dissertation but insufficient methods, a TNA that accounts for prior learning may be able to avoid the core training and move students immediately to the specialised training module. However, currently the ESRC pays fees at the same level during the 1+ year as the +3 years; and to help encourage HEIs to match investment (in larger teaching teams. workloads that include elevated pedagogic engagement, teamwork, student engagement, refreshing learning materials; working with a local industry on real data for challenge-led assignments) money saved, could be invested in a greater fee to HEIs for 1+ training relative to the +3 years.

Further the ESRC may optimise HEI engagement by negotiating with REF to increase the range of outputs recognised either as publications or as significant contributions to the Research Environment. Including learning materials and open access data would mean teaching-team investment in excellent learning experiences will also generate recognition and income if submitted to the REF exercise. A culture shift here, could trigger changes within HEIs who could also explicitly recognise these outputs for promotion. This then enables data-driven skills teachers to advance in their careers because of their excellence in learning and teaching. In this way, excellent materials and learning becomes sustainable as educators are clearly rewarded for their contribution.

5.4 Towards future work (stage 3)

While attempts were made to consider a whole-career approach to data-driven skills gaps where possible, the focus of this work is firmly around doctoral education and intervention. Data produced here could be used as a springboard for new investigations that focus explicitly on enablers and barriers experienced by recognised, established and leading researchers around learning new, and maintaining existing specialised data-driven skills. It is recommended that this work consider afresh, the culture change required to facilitate engagement with training for colleagues in secure posts. The colleagues involved in this research all have roles in teaching, or leading training around data-driven skills and new research would usefully engage with academics whose teaching is disciplinary in focus. Engagement of disciplinary colleagues would ensure proposals are inclusive of the full academic body. Another recommendation is that research development colleagues are involved as learning about their experience around transferable skills broadly, and training staff members specifically would be useful. Colleagues outside the social sciences may well relate to the academic sector barriers highlighted in section 4.2.2, and future work should consider involvement of non-social researchers.

A re-examination of the Training Needs Analysis process is a key recommendation of this research. As well as using it earlier to inform the choices available for the specialised training module, it has been recommended that a standard TNA be used with all ESRC-funded students. What elements are standardised, and how the new TNA is operationalised (through HEIs, more centrally through DTPs, or managed by the national provider, NCRM?) should be the focus of further work.

Appendix A References

Adeney, K. and Carey, S. (2009) Contextualising the teaching of statistics in political science, *Politics*, 29 (3), 193–200.

Aguado, A. (2009) Teaching research methods: learning by doing, *Journal of Public Affairs Education,* 15 (2), 251–260.

Aitchison, C. & Pare, A. (2012) Writing as craft and practice in the doctoral curriculum. In A. Lee & S. Danby (Eds.) *Reshaping Doctoral Education: International Approaches and Pedagogies*. Florence: Taylor and Frances pp.12-25

Atkinson, M. P., Czaja, R.F. & Brewster, Z.B. (2006) Integrating sociological research into large introductory courses: Learning content and increasing quantitative literacy. *Teaching Sociology* 34 (1) 54-64

Balan, P., Clark, M. & Restall, G. (2015) Preparing students for flipped or team-based learning methods. *Education & Training* 57 96) 639-57

Barnacle, R. (2005) Research education ontologies: exploring doctoral becoming. *Higher Education Research & Development* 24(2) 179-188. https://doi.org/10.1080/07294360500062995

Barone, L., Williams, J. U Micklos, D. (2017) Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. PLOS *Computational Biology* 13 (11) e1005858 https://doi.org/10.1371/journal.pcbi.1005755

Bennett, L. and Folley, S. (2014) A tale of two doctoral students: social media tools and hybridised identities. *Research in Learning Technology* 22 (23791)

Biesta, G., J. Allan, and R. Edwards. (2011). "The Theory Question in Research Capacity Building in Education: Towards an Agenda for Research and Practice." *British Journal of Educational Studies* 59 (3): 225–239.

Bone, J., Emele, C. D., Abdul, A., Coghill, G., & Pang, W. (2016). The social sciences and the web: From "Lurking" to interdisciplinary "Big Data" research. *Methodological Innovations Online*, *9*(1), 1.

Bourdieu, P. & Wacquant, L. (1992) *An Invitation to Reflexive Sociology*. Cambridge: Polity Press

Borrell-Damian, L., Morais, R. & Smith, J.H. (2015) *Collaborative Doctoral Education in Europe: Research Partnerships and Employability for Researchers. Report on DOC-CAREERS II Project.* Brussels: EUA European University Association.

Boud, D., and A. Lee. (2005). 'Peer Learning' as Pedagogic Discourse for Research Education. *Studies in Higher Education* 30 (5): 501–516.

Boud, D. and Lee, A. (2008) Framing doctoral education as practice. In D. Boud and A. Lee (Eds) *Changing Practices of Doctoral Education* (Abingdon, Routledge), 10–25.

Brown, S.E. (2021). '*The balance has got to be struck': Training development internship final report.* SGSSS Internship Report. Scottish Graduate School of Social Sciences.

Buckley, J., Brown, M., Thomson, S., Olsen, W. & Carter, J. (2015) Embedding quantitative skills in the social science curriculum: case studies from Manchester. *International Journal of Social Research Methodology*. 18(5) 495-510. Doi: 10.1080/13645579.2015.1062624

Campaign for Social Sciences (2019) *The importance of the Social Sciences for the Industrial Strategy*. https://campaignforsocialscience.org.uk/wp-content/uploads/2019/06/The- Importance-of-the-Social-Sciences-for-the-Industrial-Strategy.pdf Accessed 30th August 2021

Campaign for Social Sciences (2020) *Vital Business: The Essential Role of the Social Sciences in the UK Business Sector.* Sage Publications

Carter, J. (2021) *Work Placements, Internships & Applied Social Research London*, Sage

Carter, S., Smith, K. and Harrison, N. (2021) Working in the borderlands: critical perspectives on doctoral education *Teaching in Higher Education* 26(3) 283-292 DOI: 10.1080/13562517.2021.1911098

Charmaz, K. (2008) Constructionism and the Grounded Theory Methods. Handbook of Constructionist Research. New York: The Guildford Press 397-412

Charmaz, K, & Bryant A. (2011) Grounded theory and credibility in D. Silverman (Ed.). Qualitative Research: issues of Theory, Method and Practice. 3rd Edn. London: Sage 291-309

Chung, E-K., Jung-Ae, R. Young-Hong, B., & Oh-Sun, A. (2009) The effects of team-based learning in medical ethics education. *Medical Teacher* 31 (11) 1013-17

Clarke, G. & Lunt, I. (2014) International comparisons in postgraduate education: quality, access and employment outcomes. Bristol; HEFCE International comparisons in postgraduate education: quality, access and employment outcomes: Report to HEFCE by Gillian Clarke and Ingrid Lunt, University of Oxford (ioe.ac.uk) Accessed 24th October 2021

CMI (2018) 21st Century Leaders: Building employability through higher education Accessed 16th July 2021 https://www.managers.org.uk/knowledge-and-insights/research/building- employability-through-higher-education/ Accessed 24th October 2021

Corbin, J. M. & Strauss, A. L. (2008) Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Los Angeles: Sage publications

Cox, R. D. (2012) Teaching qualitative research to practitioner–researchers, *Theory Into Practice,* 51 (2), 129–136.

D'Andrea, V. (2006) Exploring methodological issues related to pedagogical inquiry in higher education. *New Directions for Teaching and Learning*. 107(3) 89-98

DCMS (2020) *National Data Strategy* https://www.gov.uk/government/publications/uk- national-data-strategy/national-data-strategy Accessed 30th August 2021

De Marrais, K., L. Moret, and E. Pope. (2018). "'I Found a Fit': Doctoral Student Narratives ofComing to a Theoretical Home in a Qualitative Research Class." *International Research in Higher Education* 3 (2): 83–98.

DeLyser, D., Potter, A. E., Chaney, J., Crider, S., Debnam, I., Hanks, G., Hotard, C. D., Modlin, E. A., Pfeiffer, M. and Seemann, J. (2012) Teaching qualitative research: experiential learning in group-based interviews and coding assignments, *Journal of Geography*, 112 (1), 18–28.

Denton, A. W. (2018) The use of a reflective learning journal in an introductory statistics course *Psychology Learning & Teaching* 17 (1) 84-93

DORA (2012) San Francisco Declaration on Research Assessment. Annual Meeting of the American Society for Cell Biology 16.12.12. https://sfdora.org/read/. Accessed 19th September 2021

Dowling, R. and Wilson, M. (2017) Digital doctorates? An exploratory study of PhD candidates' use of online tools. *Innovations in Education and Teaching International*. 54(1) 76-86

Durrant, G., Luff, R., Wiles, R. & Crow, G. (2015) *Consultation on Training Needs in Advanced Social Science Research Methods*. http://eprints.ncrm.ac.uk/3765/1/TNA%20report_final_3.pdf Accessed 30th August 2021

Earley, M. (2009) Developing reflective researchers. In M. Garner, C. Wagner, & B. East, L., Stokes, R. & Walker, M. (2014) Universities, the public good and professional education in the UK. *Studies in Higher Education* 39(9) 1617-1633.

East, L., Stokes, R. & Walker, M. (2014) Universities, the public good and professional education in the UK. *Studies in Higher Education* 39(9) 1617-1633. https://doi.org/10.1080/03075079.2013.801421

Eisenhart, M. and Jurow, S. (2011) Teaching qualitative research. In N. Denzin and Y. Lincoln (Eds) *The SAGE Handbook of Qualitative Research* (4th edn) (Thousand Oaks, CA, SAGE), 699– 714.

ESRC (2015) Postgraduate Training and Development Guidelines. 2nd Edition. Swindon: ESRC https://esrc.ukri.org/files/skills-and-careers/doctoral-training/postgraduate-training-and- development-guidelines-2015/ Accessed 19th September 2021

European Commission (2016) MORE3 study: Support Data Collection and Analysis Concerning Mobility Patterns and Career Paths of Researchers. European Commission – MORE3 Second Interim Report. Accessed 17th October 2021

European Commission (2018a) *Results of the 3rd Monitoring Exercise*. https://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=612086 Accessed 17th October 2021

European Commission (2018b) *Turning FAIR into reality: Final report and action plan from the European Commission expert group on. FAIR data*, Publications Office of the European Union, http://dx.doi.org/doi:10.2777/1524

European Commission (2020) *Digital Skills and Jobs* (issue October) https://ec.europa.eu/digital-single-market/en/policies/digital-skills Accessed 17th October 2021

FAIR4S (2019) *EOSC FAIR4S, Skills and capability framework*, European Open Science Cloud https://eosc-fair4s.github.io/ Accessed 17th October 2021

Ferrie, J. (2018) *Training Mapping Exercise: Q-Step Centres to the 2015 ESRC PG Guidelines*. Submitted to the Nuffield Foundation and ESRC 1st June, 2018.

Ferrie, J. & Spreckelsen, T. (forthcoming 2023) "Teaching Methods: Pedagogical Challenges in Moving Beyond Traditionally Separate Quantitative and Qualitative Classrooms" *Open Scholarship of Teaching and Learning* Expected 2(2)

Ferrie, J. & Scott, A. (2021) The PhD and Missing Skills: Capacity and Capability Building for Engagement with Industry. *Journal of Perspectives in Applied Academic Practice* Vol. 9 (3) pp. 22-31

Fielding, G., Smith, N., Tyson, J., Adams, J., Johns, G., Webb, A., Gray, L., Tinson, P. & Maclean, F. (2019) *UCISA Digital Capabilities 2019. Survey Report* https://www.ucisa.ac.uk/-/media/Files/UCISA/Publication-files/REPORTS/2019/ucisa_2019_DCS_Report_20190531- v2.pdf?la=en Accessed 17th October 2021

Fillery-Travis, A., & Robinson, L. (2018). "Making the Familiar Strange – A Research Pedagogy for Practice." *Studies in Higher Education* 43 (5): 841–853.

Flinders, M. (2020) *Fit for the Future: Research Leadership Matters* Swindon: ESRC https://esrc.ukri.org/files/research/fit-for-the-future-research-leadership-matters/ Accessed 19th September 2021

Garner, M., Wagner, C. & Kawulick, B. (2009) (Eds) *Teaching Research Methods in the Social Sciences* (pp. 104-110) London: Ashgate

Glaser, B. G. (1978) *Theoretical Sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press

Glaser, B. & Strauss, A. (1965) Awareness of Dying, Chicago: Aldine

Glaser, B. & Strauss, A. (1967) The Discovery of Grounded Theory: Strategies of Qualitative Research. Chicago, IL: Aldine

Great Britain. Department for Business, Energy & Industrial Strategy. (2017). Industrial Strategy: building a Britain fit for the future. https://www.gov.uk/government/publications/industrial- strategy-building-a-britain-fit-for-the-future Accessed 17th October 2021

Grimes, S., Southgate, E., Scevak, J. & Buchanan, R. (2019) University student perspectives on institutional non-disclosure of disability and learning challenges: reasons for staying invisible. *International Journal of Inclusive Education* 23 (6) 639-655 https://doi.org/10.1080/13603116.2018.1442507

Hager, P. (2005) Current theories of workplace learning: a critical assessment. In. N. Ascia, A. Cumming, A. Dunnow, K. Leithwood and D. Livingstone (Eds) *International Handbook of Education Policy*. London: Kluwer.

Hale, C. & Allam, A. (2020) Turning the remote access revolution into reasonable adjustments: Guidance on including disabled people in face-to-face meetings using videoconferencing technology. Chronic Illness Inclusion Project. https://chronicillnessinclusion.org.uk/our- work/access-inclusion-with-elci/ Accessed 11th October 2021

Halford, S. & Savage, M. (2017) Speaking sociologically with big data: symphonic social science and the future for the big data research *Sociology* 51 (6) 1132-1148 https://doi.org/10.1177/0038038517698639

Henderson, M., Finger, G. and Selwyn, N. (2016) What's used and what's useful? Exploring digital technology use(s) among taught postgraduate students. *Active Learning in Higher Education* 17(3) 235-247

HESA (2020) Higher Education Student Statistics: UK, 2018/2019 – Student Numbers and Characteristics. Accessed 11th October, 2021

Hernández-Hernández, F., & Sancho-Gil, J.M. (2015). "A Learning Process within an Education Research Group: An Approach to Learning Qualitative Research Methods." *International Journal of Social Research Methodology* 18 (6): 651–667.

Hesse-Biber, S. (2015) The problems and prospects in the teaching of mixed methods research. *International Journal of Social Research Methodology* 18(5) 463-477. Doi: 10.1080/13645579.2015.1062622

Heyborne, W. H. & Perrett, J. J. (2016) To flip or not to flip? Analysis of a flipped classroom pedagogy in a general biology course. *Journal of College Science Teaching* 45 (4) 31-37

Hillman, J. (2014) Mathematics after 16: The state of play, challenges and ways ahead London: Nuffield Foundation

Hodgen, J., Pepper, D., & London, C. (2010) *Is the UK an outlier?* London: Nuffield Foundation.

Hodkinson, P. & Hodkinson, H. (2004) Rethinking the concept of community of practice in relation to schoolteachers' workplace learning. *International Journal of Training and Development*, 8(1): 21-31. https://doi.org/10.1111/j.1360-3736.2004.00193.x

Howard, C. & Brady, M. (2015) Teaching social research methods after the critical turn: Challenges and benefits of a constructivist pedagogy. *International Journal of Social Research Methodology*. Vol. 18 511-525

Huggins, C. M. & Stamatel, J. P. (2015) An exploratory study comparing the effectiveness of lecturing versus team-based learning *Teaching Sociology* 43 (3) 227-35

Hughes, A. (2016) Exploring normative whiteness: ensuring inclusive pedagogic practice in undergraduate fieldwork teaching and learning. *Journal of Geography in Higher Education* 40 (3) 460-477 https://doi.org/10.1080/03098265.2016.1155206

Humphreys, M. (2006) Teaching qualitative research methods: I'm beginning to see the light.
Qualitative Research in Organizations and Management. 1(3) 173-188

Jairam, D., and D. H. Kahl Jr. (2012) "Navigating the Doctoral Experience: The Role of Social Support in Successful Degree Completion." *International Journal of Doctoral Studies* 7: 311–329.

Janta, H., P. Lugosi, and L. Brown. (2014) "Coping with Loneliness: A Netnographic Study of Doctoral Students." *Journal of Further and Higher Education* 38 (4): 553–571.

Jisc. (2018) *Researcher Profile. Six Elements of Digital Capabilities Model*. https://repository.jisc.ac.uk/6622/1/JiscProfile_researcher.pdf Accessed 17th October 2021

Jisc. (2020) *Researcher digital experience insights survey 2020: UK Higher Education pilot survey findings.* https://digitalinsights.jisc.ac.uk/documents/268/DEI_Researchers_Insights_1920_v1.2- FINAL.pdf Accessed 17th October 2021

Kilburn, D., Nind, M., & Wiles, R. (2014a). Learning as Researchers and Teachers: The Development of a Pedagogical Culture for Social Science Research Methods? *British Journal of Educational Studies*, *62*(2), 191–207.

Kilburn, D., Nind, M., & Wiles, R. (2014b). *Short courses in advanced research methods: Challenges and opportunities for teaching and learning*.

Kottman, A. (2011) Reform of doctoral training in Europe. The emergence of a new form of educational governance. In J. Enders, H de Boer & D. F. Westerheijden (Eds) *Reform of Higher Education in Europe* Sense Publishers. Pp39-54

Kumar, V., & Aitchison, C. (2017) Peer facilitated writing groups: A programmatic approach to doctoral student writing. Teaching in Higher Education 23 360-373 https://www.doi.org/10.1080/13562517.2017.1391200

Kurata, L., Matsubayashi, M. & Takeda, M. (2017) Research data management in Japanese universities and research institutions: Status report based on questionnaire survey. *Journal of Information Processing and Management* 60(2) 119-127

Lenihan, A. & Witherspoon, S. (2020) *Vital Business. The Essential Role of the Social Sciences in the UK Private Sector* Campaign for Social Science https://acss.org.uk/publications/vital- business-the-essential-role-of-the-social-sciences-in-the-uk-private-sector/ 30th August 2021

Lewthwaite, S. & Jamieson, L. (2019) *Big Qual – Why We Should be Thinking Big About Qualitative Data for Research, Teaching and Policy* – Blog LSE https://blogs.lse.ac.uk/impactofsocialsciences/2019/03/04/big-qual-why-we-should-be- thinking-big-about-qualitative-data-for-research-teaching-and-policy/ 30th August 2021

Lewthwaite, S., & Nind, M. (2016). Teaching Research Methods in the Social Sciences: Expert Perspectives on Pedagogy and Practice. *British Journal of Educational Studies*, *64*(4), 413–430.

Lindner, A. M. (2012) Teaching quantitative literacy through a regression analysis of exam performance. *Teaching Sociology* 40 (1) 50-59

Linneman, T. J. (2021) From measures of association to multilevel models: Sociology journals and the quantitative literacy gap. Teaching Sociology 49 (1) 45-57 https://doi.org/10.1177/0092055X20980427

Littlejohn, A., Beetham, H., and McGill, L. (2012) Learning at the digital frontier: a review of digital literacies in theory and practice. Journal of Computer Assisted Learning 28 547-556

Lonie, A. & Francis, R. (2017) *An Australian Bioscience Data Capability Project Report – Phase 2* see: https://www.melbournebioinformatics.org.au/project/aust-bioinformatics-commons/ Accessed 17th October 2021

MacInnes, J. (2010) *Proposals to Support and Improve the Teaching of Quantitative Research Methods at Undergraduate Level in the UK* [online] Swindon: ESRC. Retrieved November 21, 2019: https://esrc.ukri.org/files/research/qmi/final-report-strategic-advisor-for-quantitative- methods-proposals-to-support-and-improve-the-teaching-of-quantitative-research-methods- at-undergraduate-level-in-the-uk/ Accessed 17th October 2021

MacInnes, J. (2018) *Effective Teaching Practice in Q-Step Centres*. London: Nuffield Foundation

Manathunga, C. (2019) 'Timescapes' in doctoral Education: the politics of temporal equity in higher education. *Higher Education Research & Development* 38(6) 1227-1239 DOI:10.1080/07294360.2019.1629880

Mazzucato, M. (2013) *The Entrepreneurial State: Debunking Private Versus Public Sector Myths*. London: Anthem Press

Mateos-Garcia, J., Bakshi, H. & Windsor, G. (2015) *Skills of the Datavores Talent and the Data Revolution* https://media.nesta.org.uk/documents/skills_of_the_datavores.pdf 30th August 2021

McVie, S., Coxon, A., Hawkins, P., Palmer, J., & Rice, R. (2008) *ESRC/SFC Scoping Study into Quantitative Methods Capacity Building in Scotland*. Final Report [online] Retrieved November 21, 2019: https://esrc.ukri.org/files/research/esrc-sfc-scoping-study-into-quantitative-methods- capacity-building-in-scotland/ Accessed 17th October 2021

Michaelsen, L. K., Knight, A. B. & Fink, L. D. (2004) *Team Based Learning: A Transformative Use of Small Groups in College Teaching*. Sterling, VA: Stylus

Moltzou, A. (2019) *Towards Social Data Science: With Great Quantity Comes Great Responsibility* Towards Data Science https://towardsdatascience.com/towards-social-data- science-f90c5c020855 Accessed 30 August 2021

Mons, B. (2020) Invest 5% of research funds in ensuring data are reusable. *Nature*. Vol. 578/7796 pp.491-491 http://dx.doi.org/10.1038/d41586-020-00505-7

Morgan, D.L. (2007) Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1 48-76

Morris, P., Ida, A. K., Migliaccio, T., Tsukada, Y. & Baker, D. (2020) Collaborative learning in sociology research methods courses: does race matter? *Teaching Sociology* 48(4) 300-312

Nairn, K. (2020) Learning through doing: the potential of a collective editing pedagogy Teaching in higher Education 25 (7) 888-901 https://www.doi.org/10.1080/13562517.2019.1632827

National Audit Office (2019) *Challenges in Using Data Across Government* https://www.nao.org.uk/report/challenges-in-using-data-across-government/ 30th August 2021

National Data Strategy (2020) *Policy Paper* https://www.gov.uk/government/publications/uk- national-data-strategy/national-data-strategy Accessed 30th August 2021

National Science Foundation (2011) *Rebuilding the Mosaic: Fostering Research in the Social, Behavioural, and Economic Sciences* (Arlington, VA, National Science Foundation Directorate for Social, Behavioral and Economic Sciences).

Nind, M., Curtin, A. & Hall, K. (2016) *Research Methods for Pedagogy*. London: Bloomsbury

Nind, M., Holmes, M., Insenga, M., Lewthwaite, S., & Sutton, C. (2020). Student Perspectives on Learning Research Methods in the Social Sciences. *Teaching in Higher Education*, *25*(7), 797– 811.

Nind, M., Kilburn, D., and Luff, R. (2015). The teaching and learning of social research methods: developments in pedagogical knowledge. International Journal of Social Research Methodology, 18 (5), 1-9. https://doi.org/10.1080/13645579.2015.1062631

Nind, M., Kilburn, D. & Wiles, R. (2015) Using video and dialogue to generate pedagogic knowledge: teachers, learners and researchers reflecting together on the pedagogy of social research methods. I*nternational Journal of Social Research Methodology* 18 (5) 561-576 https://doi.org/10.1080/13645579.2015.1062628

Nind, M., & Lewthwaite, S. (2018a). Hard to teach: inclusive pedagogy in social science research methods education. *International Journal of Inclusive Education*, *22*(1), 74–88.

Nind, M., & Lewthwaite, S. (2018b) Methods that teach: developing pedagogic research methods, developing pedagogy International Journal of Research & Method in Education 41 (4) 398-410 https://www.doi.org/10.1080/1743727X.2018.1427057

OECD (2016) Building Skills for All: A Review of England https://www.oecd.org/unitedkingdom/building-skills-for-all-review-of-england.pdf Accessed 30 August 2021

OECD (2020) *Building Digital Workforce Capacity and Skills for Data-Intensive Science* OECD Science, Technology and Industry Policy Papers, No. 90, OECD Publishing, Paris, https://doi.org/10.1787/e08aa3bb-en.

Park, C. (2005) New variant PhD: The changing nature of the doctorate in the *UK Journal of Higher Education Policy and Management* 27(2)189-207 DOI: 10.1080/13562517/2021.1900814.

Payne, G. (2014). Surveys, Statisticians and Sociology: A History of (a Lack of) Quantitative Methods. *Enhancing Learning in the Social Sciences*, *6*(2), 74.

Persson, H., Ahman, H., Yingling, A. A. & Gulliksen, J. (2015) Universal design, inclusive design, accessible design, design for all: different concepts – one goal? On the concept of accessibility
– historical, methodological and philosophical aspects. *Universal Access in the Information Society* 14 (4) 505-526 https://doi.org/10.1007/s10209-014-0358-z

Piatetsky, G. with Gandhi, P. (2018) How many data scientists are there and is there enough? https://www.kdnuggets.com/2018/09/how-many-data-scientists-are-there.html. Accessed 19th September, 2021

Probst, B., D. Harris, J. Pehm, R. Lindquist, O. Mora, V. Hallas, and S. Sandoval. (2016) "In Our Voices: A Collaborative Reflection on Teaching and Being Taught." *Qualitative Social Work: Research and Practice* 15 (3): 331–345.

Punch, S. (2012) "Hidden Struggles of Fieldwork: Exploring the Role and Use of Field Diaries." Emotion, *Space and Society* 5: 86–93.

Purdam, K. (2016) Task-based learning approaches for supporting the development of social science researchers' critical data skills. *International Journal of Social Research Methodology* 19 (2) 257-267

Rainford, J. (2016) Becoming a doctoral researcher in a digital world: Reflections on the role of Twitter for reflexivity and the internal conversation. *E-Learning and Digital Media* 13(1-2) 99-105

Ralston, K. (2020). 'Sociologists Shouldn't Have to Study Statistics': Epistemology and Anxiety of Statistics in Sociology Students. *Sociological Research Online*, *25*(2), 219–235.

Roberts, S., Hine, C., Morey, Y., Snee, H., & Watson, H. (2013). *"Digital methods as mainstream methodology": Building capacity in the research community to address the challenges and opportunities presented by digitally inspired methods*.
Royal Society (2019) *Dynamics of data science skills* https://royalsociety.org/-/media/policy/projects/dynamics-of-data-science/dynamics-of-data-science-skills-report.pdf?la=en-GB&hash=212DAE7D599B0A48687B372C90DC3FEA 30th August 2021

Ryan, M., Saunders, C., Rainsford, E., & Thompson, E. (2014) Improving research methods teaching and learning in politics and international relations: A 'reality show' approach. *Politics* 34 85-97

Saldaña, J. (2009) Popular film as an instructional strategy in qualitative research methods courses, *Qualitative Inquiry*, 15 (1), 247–261.

Savage, M. and Burrows, R. (2007) The coming crisis of empirical sociology, *Sociology*, 41 (5), 885–899.

Scott Jones, J. & Goldring, J. E. (2014) *Skills in Mathematics and Statistics in Sociology and Tackling Transition. Higher Education Academy STEM Project: Skills in Mathematics and Statistics in the Disciplines and Tackling Transition*. York: HEA

Scott Jones, J. & Goldring, J.E. (2015) 'I'm not a quants person'; Key strategies in building competence and confidence in staff who teach quantitative research methods. *International Journal of Social Research Methodology* 18(5) 479-494 Doi: http://dx.doi.org/10.1080/13645579.2015.1062623

Shaw, K., Holbrook, A. and Bourke, S. (2011) Student experience of final-year under- graduate research projects: an exploration of 'research preparedness', *Studies in Higher Education*, 38 (5), 711–727.

Skemp, R. R. (1976) Relational understanding and instrumental understanding. *Mathematics Teaching.* 77, 20-26

Social Media Research Group (2016) *Using Social Media for Social Research: An Introduction*
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_d ata/file/524750/GSR_Social_Media_Research_Guidance_-_Using_social_media_for_social_research.pdf 30th August 2021

Sowell, R., Zhang, T. & Redd, K. (2008) Ph D Completion and Attrition: Analysis of Baseline Program Data from the Completion Project. Washington, DC: Council of graduate Schools, p.17

Spigner-Littles, D. A. & Chalon, E. (1999) Constructivism: A paradigm for older learners. *Educational Gerontology* 25 203-209

Starke-Meyerring, D. (2011) The paradox of writing in doctoral education: Student experiences. In L. McAlpine and C. Amundsen (Eds.) *Doctoral Education: Research-Based Strategies for Doctoral Students, Supervisors and Administrators*. Dordrecht: Spring ppp.75-95

Strangeld, J. A. (2013) Promoting active learning: student led data gathering in undergraduate statistics. *Teaching Sociology* 41 (2) 199-206

Sweet, M. & Michaelsen, L. K. (Eds.) (2012) *Team-Based Learning in the Social Sciences and Humanities: Group Work That Works to Generate Critical Thinking and Engagement*. Sterling, VA: Stylus

Sweitzer, V. (2009) "Towards a Theory of Doctoral Student Professional Identity Development: A Developmental Networks Approach." The Journal of Higher Education 80: 1–33.

Tazzyman, S., Moreton, R., Bowes, L., Wakeling, P. & Stutz (2021) Review of the PhD in the Social Sciences. Swindon: Economic and Social Research Council
https://esrc.ukri.org/files/skills-and- careers/review-of-the-phd-in-the-social-sciences/
Accessed 9th October 2021

The British Academy (2013) *Fuelling prosperity: research and innovation as drivers of UK growth and competitiveness.* Joint statement from the Academy of Medical Sciences, the British Academy, the Royal Academy of Engineering and the Royal Society. Available at: http://www.britac.ac.uk/templates/asset-relay.cfm?frmAssetFileID=12529 Accessed 1 February 2014.

The British Academy (2016) *Crossing Paths: Interdisciplinary Institutions, Careers, Education and Applications*. https://www.thebritishacademy.ac.uk/publications/crossing-paths/. Accessed 19th September 2021

Thien, D. (2009) Encouraging gender analysis in research practice, *Journal of Geography in Higher Education,* 33 (3), 351–367.

Thrift, N. (2008) *Research Careers in the UK: A Review* https://www.readkong.com/page/research-careers-in-the-uk-a-review-5164567 Accessed 30th August 2021

Tinati, R., Halford, S., Carr, L. & Pope, C. (2014). Big data: Methodological challenges and approaches to sociological analysis. Sociology 48 (4) 663-681 https://doi.org/10.1177/0038038513511561

Universities UK (2009) *Promoting the UK Doctorate: Opportunities and Challenges* https://dera.ioe.ac.uk/26279/1/PromotingUKDoctorate.pdf Accessed 30th August 2021

Van den Eynden, V., & Corti, L. (2017). Advancing research data publishing practices for the social sciences: from archive activity to empowering researchers. *International Journal on Digital Libraries*, *18*(2), 113–121.

Vitae. (2011) *Vitae Researcher Development Framework.* Available at: https://www.vitae.ac.uk/vitae-publications/rdf-related/researcher-development-framework- rdf-vitae.pdf/view. Accessed 20th September 2021

Vitae. (2019a). *Careers in Research Online Survey (CROS) 2019 UK aggregate results*.

Vitae. (2019b). *Principal Investigators and Research Leaders Survey (PIRLS) 2019 UK aggregate results*.

Vitae. (2020) What do doctoral graduates do? Available at: https://www.vitae.ac.uk/doing- research/are-you-thinking-of-doing-a-phd/what-do-doctoral-graduates-do. Accessed 24th October 2021

Wagner, C., Garner, M. & Kawulich, B. (2011) The state of the art of teaching research methods in the social sciences: Towards a pedagogical culture. *Studies in Higher Education*. 36 75-88

Wathan, J., Brown, M. & Williamson, L. (2011) Understanding and overcoming the barriers increasing secondary analysis in undergraduate dissertations. In J. Payne & M. Williams (Eds.), *Teaching Quantitative Methods: Getting the Basics Right* (pp. 121-141). London: Sage.

Williams, M., Collett, C., & Rice, R. (2004) *Baseline Study of Quantitative Methods in British Sociology*. Birmingham/Durham: C-SAP/BSA

Williams, M., Payne, G., Hodgkinson, L., & Poade, D. (2008) Does British sociology count? Sociology students' attitudes toward quantitative methods. *Sociology* 42 1003-1021

Williams, M., Sloan, L., Cheung, S. Y., Sutton, C., Steven, S. & Runham, L. (2016) Can't count or won't count? Embedding quantitative methods in substantive sociology curricula: A Quasi experiment. *Sociology* 50 (3) 435-452

Willis, J.W. (2007) *Foundations of Qualitative Research: Interpretive and Critical Approaches*
London: Sage

Wilson, S. G. (2013) The flipped class: A method to address the challenges of an undergraduate statistics course. *Teaching of Psychology* 40 (3) 193-99

Windsor, G. & Mateos-Garcia, J. (2015) *Analytic Britain: Securing the Right Skills for the Data- Driven Economy* https://media.nesta.org.uk/documents/analytic_britain.pdf Accessed 30th August, 2021

Wollschleger, J. (2019) Making it count: Using real-world projects for course assignments. Teaching *Sociology* 47(4) 314-324
https://doi.org/10.1177/0092055X19864422

Zauner, J. (2021). *Getting ready for the fourth revolution: Digital skills and learning. SGSSS Internship Report.* Scottish Graduate School of Social Sciences.

Zukas, M. & Andersen, L. L. (2012) Taking a Break: Doctoral Summer Schools as transformative pedagogies. In A. Lee & S. Danby (Eds) *Reshaping Doctoral Education International Approaches and Pedagogies*. Oxon: Routledge.

Appendix B Methodology Detail and Use of Grounded Theory

This appendix reveals more detail about the methodology in order to provide background and establish the rigour of this work. After an overview of grounded theory, the discussion will turn to each phase of research inquiry.

Adoption of grounded theory

Our methodological framework was informed by Charmaz's (2008) approach to Grounded Theory (GT). The research team do not claim to have used any one version of GT specifically, but rather borrowed from its ideology and analytical approach. Grounded theory worked well for this project design. The constant comparative method permitted us to check emergent findings and recommendations (understood as developing theory about data-driven skills gaps) against each of the initial phases of data collection. Thus, we grounded the findings and recommendations in the reviews and the primary data collection. This commitment to ensure recommendations are useful and viable, yet also able to deliver change where it is required, led to the use of 'sense-checking' expert workshops. Influenced greatly by the Delphi approach, this strategy allowed us to ground recommendations not just in the data produced but check them against expert views of what will work well in practice.

Using a grounded theory approach, each research phase informed the others. For example, the literature review was performed in two stages: a broad field-capture as the project launched, and a targeted, purposive capture after the initial interviews and survey were complete. The landscape review was one of the first activities we engaged in and has also been revisited during and after other research phases have completed as we've developed a 'theory' to be tested. For example, we returned to the landscape review to determine the core masters learning of UK universities not part of a doctoral training partnership to explore the influence of the ESRC's 2015 Postgraduate Training and Development Guidelines.

Our approach produces a new 'theory' of data-driven skills training that is grounded in the data and validated by those with expertise in practice.

In 1965, Glaser and Strauss used Grounded Theory (see also Glaser and Strauss, 1967 for the early outline of the method) as an approach to qualitatively understand the sociology of dying in six San Franciscan hospitals. Glaser and Strauss looked at the interplay between hospital management, staff and patients and thus used a form of triangulation to draw out similarities and differences in the sub-fields they explored. This approach became known as the constant comparative method. In this research we are exploring the interplay between skills learners, skills leaders, and skills providers across numerous 'sites' to fully understand where skills-gaps exist, and how gaps may be filled. We also set up the research phases of the study so that early findings could inform later data collection, inspired by Glaser and Strauss's 1967 work in which they outlined a process of collecting data, initial analysis, informing direction of next data collection phase would continue until theoretical saturation was achieved.

A strength of the GT approach for this project, is that the skills gap is essentially about finding out an unknown: what skills do social scientists need that are not currently provided for? In a typical GT research study, the literature available can be used to orient the study towards more in-depth field work which focuses more closely on the knowledge gap: here relating to data- driven research skills. Glaser (1978) also called on researchers to approach the field as a 'blank slate'. This was difficult to achieve given the legacy many of the research team have in developing data-driven research training programmes. Our approach is supported by Corbin and Strauss (2008) and their acknowledgement that data production is co-constructed meaning and the researchers cannot stand beyond this relationship. It is supported further by Charmaz and Bryant's (2011) constructivist GT that emphasises reflexivity and rejects assumptions that researchers should and could set aside their prior knowledge to develop new theories. Here constructivist relates to meaning being located within an individual (rather than constructionist: meaning located within relations/between people) and this supports our decision to speak with individuals through our qualitative survey and expert interviews to understand more about their reality in relation to data-driven research skills.

GT is a strong choice to be influenced by because it helped develop a theory of what skills gaps exist, and we could test the emergent theory with experts in the sense-checking workshops. The first analysis phase involved data from the landscape and literature reviews; and primary data from the survey and interviews: the data was 'fractured' or 'opened' to reveal substantive codes in the data: a form of descriptive code which were then developed as we found relationships between these categories as we drew our different forms of data together. As we understood more about these relationships, we performed axial coding (linking across the codes), converting substantive into theoretical codes, the building blocks of our theoretical arguments. Following the expert workshops we arrived at core codes, a higher order conceptualisation which signifies our contribution to the field. These form the substantive base of chapters 3 and 4.

Outline of research phase 1: Landscape review

The landscape review was a significant piece in our puzzle and aimed to understand what training was being made available through key ESRC investments in the UK, and equivalent providers overseas. This desk-based activity relied on publicly available information. Some websites did not have public access to training available. If we could not find it, doctoral candidates and other users would also face barriers.

All 14 Doctoral Training Partnerships (DTPs) were included in the desk-based review, and six had training sufficiently visible to complete a 'mapping' against the ESRC's (2015) Postgraduate and Training Guidelines. The Guidelines were discussed in detail in section 1.2, though it is useful here to make some observations about the way they construct data-driven skill learning. The Guidelines introduce a framework that promoted a breadth of understanding for social science researchers, and carefully avoided a quantitative/qualitative dichotomy. In so doing, the Guidelines avoid any suggestion of a hierarchy (where one approach is valued over the other, or is considered more robust).

The Guidelines were chosen as an anchor for this mapping exercise because they have underpinned the commissioning of DTPs to shape the 'core' 1+ provision (master's level) of PhD funding awards. In reviewing 'advanced' training provided by DTPs against the Guidelines we could understand i) how training 'added value' to the core training completed in the 1+, and ii) detect training or skills development that had not been captured in the 2015 Guidelines. Finally, the 2015 Guidelines include 'transferable skills' and 'researcher development' training which allows us to consider not just methods when thinking through data-driven skills, but also skills broadly associate with 'doing' research and 'being' a researcher, including project management, communication, and dissemination of findings.

Other ESRC investments were added into the desk-based review. The 17 Q-Step centres have driven forward a step-change in how some social scientists use and understand quantitative methods with programmes focusing heavily on the undergraduate student. Since Transitional Funding in 2019 (to September 2021), Q-Step centres had a mandate to articulate with post- graduate training. Again, most Q-Step centres had websites where the training available was not publicly visible. Of those that did, there was a clear and established contribution to quantitative methods training that could be accessed by post-graduate and established career academics. Practice though was varied, and so the data has been synthesised to allow comparison with other ESRC investments such as Centres for Doctoral Training and DTPs.

This work was extended to include global leaders in data-driven social research skills. Summer school programmes that badged themselves as offering advanced data-driven skills training were reviewed, featuring schools based in the U.S., Northern Europe, and Central Europe. Further, doctoral programmes with a greater taught element (as compared with UK ESRC- funded doctorates) were reviewed to determine implications for data-driven skills. This included a review of programmes in the U.S, Canada, and Germany. The inclusion of global actors allowed a comparative approach, to consider where UK training can be considered world-leading and where gaps are emerging.

Outline of research phase 2: Literature review

The literature review sits in the context of significant work in the UK and globally to make data available at unprecedented levels. The grey literature on UK policy heralds a 'data opportunity' to improve research capabilities and decision-making processes in the UK through the incorporation and mobilisation of high-quality data (National Audit Office, 2019; National Data Strategy, 2020). There have also been significant increases in literature examining the teaching of data-driven skills and research methods (Nind, Kilburn and Luff, 2015) suggesting a growing culture around pedagogy in this field. The literature review used grey literature and peer-reviewed academic articles and books.

Finding 'gaps' created a challenge in forming a literature review, as we were essentially trying to find what is missing. What do we search for? Early attempts at a shallow, but systematic capturing of the field is outlined in figure A1, below. The approach did not adequately produce convincing evidence of where the gaps are, or how they may be overcome. It was a useful foundation for a critically engaged

approach that used key contributions and their references to approximate something akin to a snowballing approach. That is, we started reading, critically engaging with the material, and mining it for gaps and other barriers to learning about data, evidence building and data-driven skills. We used reference sections and citation searches to reach a wider literature base. The literature is drawn from UK scholars and other predominantly English-speaking writing, and this is one limitation of this review.

There are three main themes that emerged through the review of the literature around data driven skills gaps, and these have informed the data chapters 2, 3, and 4.

- An evaluation of the ESRC's 2015 Guidelines as a foundational document
- Engaging with pedagogical approaches to social research methods teaching and thinking critically about how this could help us further understand data-driven skills gaps and how we might close them
- Considering the gap that exists around digital skills

The initial literature review process was as follows:

- Identification and primary mapping to the search strategy set out in Figure A1 below
- Selection of studies to include, based on a secondary mapping structure as follows:
  - Does the publication include a definition of data skills?
  - Does the publication set out data skills requirements?
  - Does the publication include specific references to or information about the life-course stages (researcher or outside academia)?
- Assessment of whether the publication provide an understanding of skills development / acquisition / learning outcome

Figure A1 Initial literature review search strategy

A		B		C		D		E
Discipline or organisation		**Career stage**[19]		**Skills area**		**Mechanism**		**Location**
Social Science		'Early-career'		Data skills		'Investment in…'		UK [ON/OFF]
'Other disciplines'		Post-graduate		Data handling		'Provision of …'		'Other locations'
Arts		Doctoral		Data analysis		Training		England
Humanities[20]		Post-doc		Data management		Capacity dev.		Scotland
Biological sciences		Researcher		Data ethics		CPD/Ct. Prof. Dev.		North. Ireland
Engineering		'Mid-career'		Data-driven research		Methods learning		Wales
Physical sciences[21]	**AND**	University scientist	**AND**	Data-intensive research	**AND**	Formal learning	**AND**	
Economics		Lecturer		Qualitative research method(s)[22]		Informal learning		
Social research[23]		Senior Lecturer		Quantitative research method(s)		Class-based learning		
Medicine/medical		PI/Principal investigator		Digital research method(s)		Work-based learning		
Environmental sciences		'Late career'		Computing science		Curriculum		
Computing science		Supervisor		'New forms of data'		Pedagogical approach		

[19] Developed in line with models such as the Flinders (2020) typology: Phase 1 Doctoral; Phase 2 Post-doc; Phase 3 University scientist [Lecturer/Senior Lecturer]; Phase 4 Professor (but broadened to cover potentially relevant additional terms)
[20] Inclusive of: ancient and modern languages, literature, philosophy, history, archaeology, anthropology, human geography, law, politics, religion
[21] Inclusive of: Mathematics, physics, chemistry, artificial intelligence, computer science, plus space science and astronomy
[22] Inclusive of: Narrative methods, ethnography, diary methods, grounded theory, qualitative comparative analysis
[23] Inclusive of: Anthropology, communication studies, education, geography, history, law, linguistics, political science, psychology, sociology, behavioural science

Cultural studies	Professor	Social media data	Skills acquisition	
	Research leader	Research leadership/culture[24]	Drivers of …	
	Unit/centre director	Multi-/interdisciplinary/MIDRI	Barriers to …	

[24] Supporting academic development, creating conducive work cultures

Outline of research phase 3: Expert interviews

The expert interviews were conducted with users of data-driven skills and strategic stakeholders. These interviews were treated as elite interviews, with the aim of hearing the voices of active researchers (in the UK and internationally) who have a strong reputation for data-driven research and often for creating data-driven skills learning experiences.

We aimed to be varied in terms of hearing voices representing different kinds of researcher, using different research methodologies and at different stages of their career. Thirty-eight interviews were completed in this phase (see tables A2, A3, A4, and Appendix D for an overview of the sample). As these were in-depth, rich, qualitative interviews, we did not aim for representativeness per se, but used purposive sampling to secure diversity.

Table A2 Interview sample by lifecourse stage

Lifecourse stage	Number of invitations sent	Number of interviews completed
Ph 1 – Doctoral	10 (16%)	7 (18%)
Ph 2 – Post-doc	9 (14%)	7 (18%)
Ph 3 – Lecturer / Senior Lecturer	13 (21%)	8 (21%)
Ph 4 – Professor	31 (49%)	16 (42%)
Total	63	38

Table A3 Interview sample by qualitative/quantitative tradition

Research tradition	Number of invitations sent	Number of interviews completed
Qualitative	28 (44%)	17 (45%)
Quantitative	40 (63%)	24 (63%)
Total	63*	38*

NB: * This does not represent a 100% breakdown of the sample, as individuals were able to identify as using both qualitative and quantitative methods, which several did

Table A4 Interview sample by gender

Gender of interviewee	Number of invitations sent	Number of interviews completed
Male	34 (54%)	22 (58%)
Female	29 (46%)	16 (42%)
Non-binary	0 (0%)	0 (%)
Total	63	38

As well as reflecting on their own research skills, and skills gaps, the participants shared their views on training provision and skills gaps more broadly. The majority had sufficiently navigated training to produce research robust enough to establish research careers. Their expertise in a particular method/methodology/skill set/approach was listened to, and not contested. That is, expert participants were not asked to defend their choice of approach to research, but rather to think more broadly about the enablers and barriers to developing rigorous research skills. This data has been analysed borrowing from the grounded theory approach described above.

Outline of research phase 4: qualitative survey to doctoral training partnerships (DTPs)

The qualitative survey to DTP training leads featured 10 questions that we hoped would feel manageable for DTPs to complete during the busy studentship competition period. Sub- questions were used to prompt extended answers. The survey was sent to the training lead if identified on the DTP website (where they existed, they were often a Deputy Director) and to the Director if there was no identifiable training lead. Eight of the 14 DTPs responded, and those that did provided strong and extensive answers. The DTP leads have also been invited to attend the sense-checking expert workshops.

The survey again used the device of foregrounding the ESRC's Guidelines, thus anchoring responses in the sector-wide standard of broad social science training. Five themes were covered: core training; training provided by the DTP; training needs analysis; working in a network and international comparison. As such, the survey aimed to take respondents through the work they do in their DTPs to assess what training will be useful for their students and how this informs the training they provide.

As with the reviews and expert interviews, the analysis features in chapters 3 and 4. This particular research phase has led to some fascinating theories: one being that qualitative digital skills have been sought by post-graduate students and will become 'must have' for all career stages in the same way that quantitative digital skills (particularly around machine learning and AI) have become 'must-have'.

Outline of research phase 5: Sense-checking expert workshops

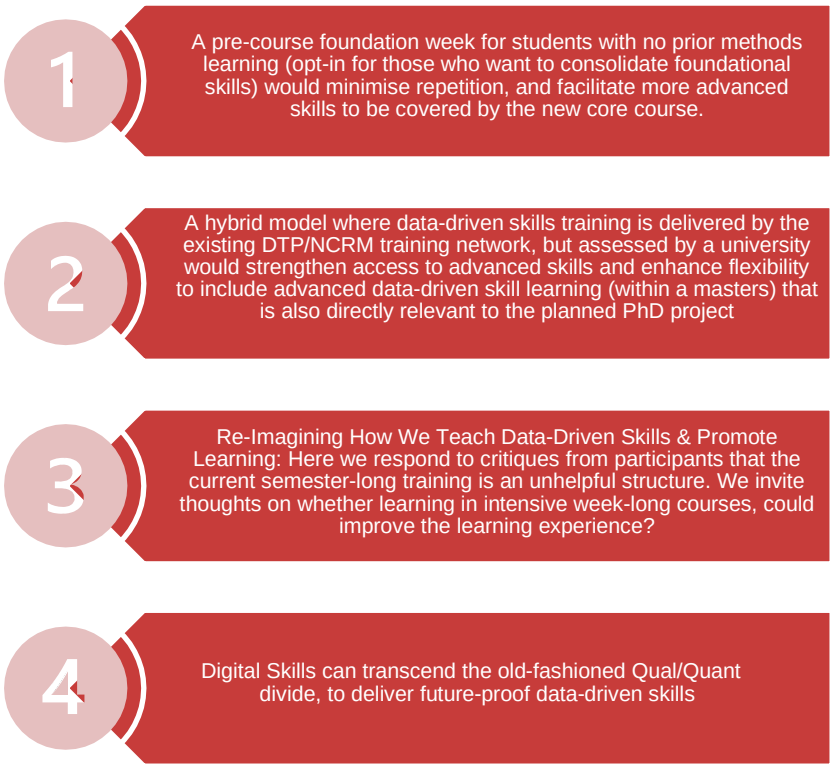
The sense-checking expert workshops are a major strength of this project and elevated the methodology to one of increased rigour, as theories generated from earlier analysis were explored with experts and tested against their practice. As is typical of using qualitative approaches broadly and being influenced by grounded theory in particular, the form of this consultation and the experts invited evolved as we absorbed learning from the earlier research phases. For example, it was decided to produce a workshop just for doctoral candidates so that 'propositions' could be tested with those currently 'learning' skills. To accommodate their availability, eleven students were involved in two workshops. All but one of the doctoral candidates who participated chose to be named (see Appendix E) and recognised for their contribution. Pseudonyms are used in the report against quotes to provide some distance between what was said, and who said it. The doctoral candidates all agreed to the use of pseudonyms. Four workshops were conducted with data-driven skill

trainers: including those managing HEI courses, to those leading within ESRC investments such as DTPs. In total fourteen colleagues were involved in the workshops.

As emerging recommendations suggested a big shift in how data-driven skills training should be delivered, and with support from the project Steering Group, we chose to populate the remaining workshops with people directly involved in delivering or planning the data-driven skills training to a broad range of social science doctoral candidates. We also decided to host four workshops to allow smaller groups to convene. This worked well to create spaces where participants could negotiate the strengths and limitations of our findings and propositions. Further, they could offer insight into the feasibility of recommendations. This practical insight was invaluable in shaping this report generally, and chapter 4 and 5 in particular.

Ahead of each workshop, participants were sent a 21-slide overview of the project to date and four propositions. The propositions aimed to capture the 'problems' and 'barriers' we had found through our research and propose solutions. The slides, converted into a word format, are available in Appendix D and are summarised in figure A1, below. Informed by the data analysis, the propositions were designed to trigger debate. They did not necessarily work together. This too was by design, to encourage participants to evaluate the propositions, rejecting elements that in their view would not enhance data-driven skills amongst doctoral candidates. To interrogate responses, a series of prompts were developed, and these too were shared in advanced (also found in D). The prompts were designed to encourage experts to move from a position of support (or not) for a proposition: to think through the feasibility of each solution. Their expertise in working through the implications in terms of structure, time and other resources required has been invaluable in forming the recommendations that feature in the Executive Summary and in concluding chapter 5.

Figure A1 Four propositions provided to experts ahead of sense-checking workshops



Source: Authors' own elaboration. Full details available in Appendix D

Appendix C Conceptual Framework

Approach to 'data'

In line with the terms of reference for this study, the study team worked with a broad definition of 'data' and 'data skills' throughout the work undertaken. This ensured an inclusive and wide- ranging examination of the skills required to exploit a range of data types. The primary data collected for this project championed a removal of the division between quantitative and qualitative learning (for example, within master's courses) yet the 'binary' persisted in how participants spoke of data-driven skills and training. We have perpetuated this binary by repeating this terminology to be faithful to the data where necessary, but aimed to transcend the binary wherever possible. The study also considered new and emerging forms of data, and the skills required to work with these data. Significant here is the rise in demand for digital data- driven skills, and our working definition for this is presented in section 3.3.1.

The term 'data-driven research' is understood inclusively: all forms of data production and/or of data analysis that informs and/or is used by social science researchers from archives to zines. There is more than a question of what is included, but also of purpose and standard. Data- driven research produces a robust and rigorous evidence base:

- In which the researcher is confident
- That is open to scrutiny
- That ensures the quality of the work can be evidenced regardless of the approach taken and is transparent, credible and/or reliable, valid
- That is sufficiently strong to inform decisions within academia and beyond
- Data-driven skills includes data production and management. It is not limited to the skills required to analyse existing data

Approach to 'skill gaps'

For the purposes of this study, 'skills gaps' are understood in three ways:

- The lack of access to training in a particular skill, method, or field where the presence of this opportunity would strengthen the researcher and the research they produce
- The lack of knowledge to evaluate what skill, method, field, or provider would best fill the gap identified. This second point relates to career stage as well. For example, the senior academic who has identified a skill gap but isn't aware of training opportunities for established and leading researchers
- The structural barriers limiting high quality teaching, or learning, of data-driven skills in the UK

Approach to doctoral education

The study focuses on doctoral education. As with the ESRC funding framework, this is inclusive of all i) social sciences, ii) masters provision, and iii) different learners (for example, part-time). The social sciences are a broad church. ESRC investments in interdisciplinary practice, including 'steer' funding for interdisciplinary ESRC funded

studentships, meant that our attention drifted toward academic disciplines in the arts, humanities, sciences and professional degrees (medicine, engineering, veterinary science). Where literature or primary data referred to non-social science disciplines, we have made it clear which field or discipline the knowledge is sourced from.

To deliver the aims of the project outlined in the introduction, we developed a conceptual framework. In doing so, we anticipated some themes would emerge from better understanding of the landscape, gaps and how to close gaps.[25]This relates to the prior knowledge of the research team: half of us have been or are currently working to provide data-driven skills training to UK social scientists.

Figure A2, below, captures our conceptual framework mid-way through the analysis and shows, through colour coded cells, how emergent themes build on research aims. The supporting evidence has been presented in chapters 3 and 4.

[25] The conceptual framework presented here has evolved over the course of the study, informed by all five phases of the research

Figure A2 Working conceptual framework and emergent theoretical codes for understanding data- driven skills gaps

Source: Authors' own elaboration. Key: Red – the aim of the project and anticipated codes; Yellow – extension of the project aims based on evidence; Green – findings; blue –findings/recommendations

Methodologically, we aimed to capture the 'field of practice'. That is, we wanted to examine not just what we (as educators or students) know about teaching and learning data-driven skills, but also what the performance of teaching and of learning data-driven skills entails. To explore this, we invited participants to appear in multiple phases of inquiry. For example, we interviewed doctoral candidates and held a

workshop for doctoral candidates. DTP training leads were invited to participate in the survey and were also invited to attend a workshop. We read literature produced by researchers who also took part in one of the other phases of inquiry. By approaching researchers at different stages in their career, and offering multiple ways of engaging, we were able to capture a complex narrative that illuminated data-driven skill practices.

Appendix D Information sent to participants ahead of expert interviews

Capture of 21 Slides presented in advance of expert workshops

1. **Data Driven Skills Gaps in UK Doctoral Education**

Technopolis with Simon Gallacher, John MacInnes & Jo Ferrie

2. **Aims & Objectives**

Scope the capacity-building needs in the social sciences to support data-driven research and identify where additional investment is needed, and how the ESRC could add value

- Landscape review of UK and overseas provision
- Literature review of data-driven skills gaps
- Short answer survey to DTPs
- Interviews with doctoral candidates and established scholars

Workshops with doctoral candidates (complete) and doctoral trainers (forthcoming) to sense- check findings

- How is current training working? How could it be improved?
- Imagined alternatives present (final slides)

Work alongside CFE Research/University of York project: Review of the PhD; sharing evidence, informing emergent recommendations

3. **Share our Working Definitions**

Data-driven research produces a robust and rigorous evidence base:

- That the researcher is confident in
- Open to scrutiny
- The quality of the work can be evidenced regardless of the approach taken: Transparent, credible and/or reliable, valid.
- Sufficiently strong to inform decisions within academia and without: in industry and policy fields for example.

Digital skills

- Relate to any data collected from a digital source (such as social media); analysis of data from a digital source even if it uses a 'traditional approach' (such as discourse analysis of tweets) and dissemination of research findings digitally (on social media for example)

Skills Gaps are understood in two ways:

- The lack of access to training in a particular skill, method, or field where the presence of this opportunity would strengthen the researcher and the research they produce
- The lack of knowledge to evaluate what skill, method, field, or provider would best fill the gap identified.

4. **Findings so far: Core Training:**

- ESRC 2015 Postgraduate Training & Development guidelines are regarded as the gold standard across the UK
- Are generally delivered across 60 credits of methods courses split: quantitative, qualitative & research design from a foundational level
- Class sizes are large, interdisciplinary spaces and those with prior learning find it repetitive
- Training Needs Assessments are used, and only 'count' prior post-graduate learning. Recognition of undergraduate and/or work-based learning would help move students from 'core' to 'advanced' learning
- The current approach to learning and assignments perpetuates a divide between qualitative and quantitative

5. **Findings so far: Advanced Training**

- Only some students are taught advanced data-driven skills by supervisors
- Students struggle to find advanced data-driven training by themselves, and are unable to evaluate whether a course will meet their needs
- More people at all stages of their career are seeking data-driven training in quantitative digital skills, particularly around AI and machine learning
- There is also an increase in doctoral students seeking data-driven training in qualitative digital skills, but this is not (yet) evident amongst mid or established career scholars, which could be causing a gap in available supervisors for data-driven digital qualitative research projects.

6. **Findings so far: Relating to Pedagogy**

- Access to messy data, and 'mission orientated' assignments build confidence and engagement
- They also help build-in data/project management and dissemination skills
- Learning with others provides a supportive environment, and develops collaborative skills
- Teaching others is the best way of consolidating knowledge on methods and data-driven skills
- Internships and placements build confidence in applying data-driven skills and a great way to develop career-defining networks

7. **This Workshop**

- Will be recorded. Video and audio recordings will be deleted as soon as the transcript has been checked for accuracy
- The anonymised transcript will be retained by the research and team and shared with the ESRC and stored on password protected secure Cloud, fully meeting GDRP requirements
- Identifying elements will be redacted that relate to you, (university, a particular doctoral training partnership) and mentions of discipline, or broad approach (quantitative/qualitative) may remain
- It is important to note that the propositions are here to be tested and are not agreed policy positions of UKRI-ESRC

8. **Imagining Better**

- Four propositions follow, which may 'close the gap'
- In assessing the value of each, please consider your own teaching, and what works to build students' confidence
- We value your expertise in 'sense-checking' the propositions so please be critical or imagine alternatives
- Propositions can be agreed or disagreed with, and we welcome alternatives, expansion, and elevation of these ideas
- We particularly welcome reflections on academic practice: how would these propositions actually work in practice?
- ***They may be mutually exclusive and are designed to trigger debate***
- In the workshop, contributors will decide the order in which we discuss the propositions

9. **Proposition 1**

Students need less repetition and greater access to more advanced data-driven skills training as part of their core training before beginning their doctorate

10. **Proposition 1 – part 1**

A pre-course foundation week for students with no prior methods learning (opt-in for those who want to consolidate foundational skills) would facilitate more advanced development to be covered by the new core course.

Our landscape review of undergraduate training provision showed that to minimise repetition, foundational learning would include:

- Introductions to ethics
- Ontology of research methods and epistemological pathways
- Levels of measurement and sampling
- Univariate and bivariate analysis with numeric data
- Data collection, interviews, participant observation and focus groups
- Using documents as data

- Thematic analysis with discursive and narrative data
- The fundamentals of data literacy

Prompts to determine feasibility:

- Would you agree with the content above, and do you consider it 'foundational'?
- Training needs analysis should recognise undergraduate and work-based learning to exempt students from the foundation week where appropriate*
- Would this produce stronger researchers and improve capacity to discuss research with people outside of the academy?
- Could this advanced core course be delivered within a 40-credit model? Assuming this would 'save' 20 credits (based on the current practice), is this best dedicated to student- specific advanced training? This leads to next proposition

11. Proposition 1 – Part 2

- Advanced content then could include a broader range of data collection and analyses approaches that build confidence and range in the application of data-driven skills
- Students will have greater understanding of how epistemological choices will impact on findings produced
- Relative to the foundational course, this new core training would place greater emphasis on data/project management and project dissemination including ethics, research integrity and demonstrating rigour.
- In terms of data driven skills, it is anticipated that students will gain understanding in:
  - Software (R, Nvivo for example)
  - Fundamentals of coding
  - Confident use of analysis approaches such as regression, and appreciation of the potential of using advanced analysis such as spatial, network, econometrics and/or structural equation modelling)
  - The curation and analysis of discourse and visual data with appreciation of the potential of using archives and secondary data
  - Appreciation of theoretical frameworks that underpin data
  - Using advanced analysis approaches such as discourse analysis, phenomenology, and critical realism
  - Data ethics
  - Building impact pathways
- Data used in learning should be student-generated or derived from industry/civil society and be 'mission focused' to help students see the value of a rich and rigorous evidence base.

12. Proposition 2

A hybrid model where data-driven skills training is delivered by the existing DTP/NCRM training network, but assessed by a university would strengthen access

to advanced skills and enhance flexibility to include advanced data-driven skill learning (within a masters) that is also directly relevant to the planned PhD project

13. **Proposition 2 – part 1**

For example:

- Attendance at four advanced courses offered across our existing NCRM/DTP training network and drawing on methods and transferable skills to meet their doctoral ambitions. Students already have access to attend this training, but this anchors these opportunities clearly in the master's year, improving awareness of their availability - and has an assessed component
- To earn their 20 credits, students submit a single reflexive 2,000 word essay that 'defends' their intended dissertation research: submitted to their supervisor for assessment so early supervision should be conducted to ensure strong fit between supervisors' expertise and students' choices. For example:
  - A student preparing for emotionally-challenging overseas fieldwork may select Interviewing & Sensitive Research; Ethics, Well-being & Researching Overseas; Object Elicitation and Translating Transcripts
  - A student preparing to work with administrative data could identify Data Management and Cleaning; Linking Administrative Datasets; Structural Equation Modelling introduction; SEM Intermediate
  - A collaborative student working with a civil society organisation may choose Participatory Action Research; Inclusive Research Practices; Reflexivity; Using Social Media for Research Dissemination

14. **Proposition 2 – part 2**

Prompts to determine feasibility:

- What are the implications for this '20 credit' student-specific advanced training being offered through NRCM/DTP advanced training provision?
- Would learning at a national level improve network building? Help students find their peers using similar advanced approaches?

15. **Proposition 3**

Re-Imagining How We Teach Data-Driven Skills & Promote Learning

16. **Proposition 3 – part 1**

Here we respond to critiques from participants that the current semester-long training is an unhelpful structure. We invite thoughts on whether learning in intensive week-long courses, could improve the learning experience? What we have learned:

- Anxiety about data-driven skills training is still high, and students continue to rate courses less highly than their substantive learning
- Teaching across semesters is difficult, an intense week(s) long courses would work better, particularly if extended support by remote and/or peer mentoring

were available, to permit practice and rehearsal of skills. This approach would increase confidence and networks.
- Is it better to locate this core learning at the beginning of a masters? Or closer to the dissertation?

17. Proposition 3 – part 2

Prompts to determine feasibility:

- Are intensive methods training weeks seen as a positive evolution?
- If so, should these be taught within the university? By the DTP to promote cohort building? Or should we, could we, 'scale-up' and draw all ESRC funded students together across the UK?
- What are the practical challenges of moving to an intensive teaching/learning model?
  - What timeframe or resources are required to re-design current courses into intensive week-long units for example?
  - How would you utilise the significant growth in pedagogical literature to enhance methods-learning environments?
- Do you think these shorter, intense periods of learning would be viewed positively by students?
  - For example, they may feel happier to devote a shorter period of time to learning approaches they do not wish to use in their own work (to paint with a very broad brush, economists using discursive data; anthropologists using numeric data).

18. Proposition 4

Digital Skills can transcend the old-fashioned Qual/Quant divide, to deliver future-proof data- driven skills

19. Proposition 4 – part 1

- A learning environment using web-scraping to produce numeric, visual, and narrative data will allow students to more clearly see the ontological roots of a range of approaches.
- Exposing students to ethics and data collection/curation of real and messy data (such as social media/digital data) rather than teaching data sets, will fill skill gaps.
- Students will benefit from greater access to 'practice' material (they can repeat exercises with new hashtags related to their research) and will enhance their data cleaning, data management, data collection and analysis skills.
- Could web-scraping support learning from across the social sciences better than traditional/current learning environments? For example, permit exposure to discipline- relevant data (student generated hashtags for example).
- Assignments could also evolve to transcend the quant/qual divide by supporting students to produce a portfolio, reflexive writing and/or critical evaluations of different approaches to knowledge production

- Assignments would build confidence as are focused less on skill, and more on a broader understanding of knowledge production.
- A shift to assignments that help students capture the epistemological choices available in their research would help students defend their research design, and 'see' the application of their learning into future projects and different fields.

20. Proposition 4 – part 2

Prompts to determine feasibility

- Do you support this proposition?
- Do you have capacity to teach this? What resources/timeline would be required?
- What do we 'lose' from this approach?
- What other skills gaps exist in your experience?
- What would make a positive change for your students?

21. Thank you for your expertise and your time

Technopolis, with Jo Ferrie, Simon Gallacher & John MacInnes

Appendix E Contributions

The research team members for this study were as listed below:

- Dr Jo Ferrie, University of Glasgow
- Dr Martin Wain, Technopolis
- Dr Simon Gallacher
- Dr Eleanor Brown
- Rebecca Allinson, Technopolis
- Dr. Peter Kolarz, Technopolis
- Prof. John MacInnes, University of Edinburgh
- Laura Sutinen, Technopolis
- Rita Cimatti, Technopolis

Thanks to colleagues at the ESRC for their support throughout the research process. Thanks to the Steering Group for their contributions.

Thank you to all the participants who contributed during pressurised working conditions exacerbated by the Covid-19 pandemic. Particular thanks to early career lecturers who are named, with permission in recognition for their contribution while managing a complex and demanding workload: Dr Sharon Greenwood, Dr Kristina Saunders, Dr Thees Spreckelsen.

Thank you similarly to the doctoral candidates, also named with their permission for their unpaid and expert contribution: Mats-Philip Ahrenshop, Ana Alonso Curbelo, Viktoria Eriksson, Gosia Kolomanska-Lawes, Stephen Manders, Erin McNulty, Elena Mylona, John Pierce, Jen Pink, Nafsika Zarkou.

technopolis group

www.technopolis-group.com