Training the next generation of doctoral researchers in data science: The impact on publications and beyond

Papagiannidis, S., Meadows, M. & Panagiotopoulos, P

Original citation & hyperlink:
Papagiannidis, S, Meadows, M & Panagiotopoulos, P 2023, 'Training the next generation of doctoral researchers in data science: The impact on publications and beyond', IEEE Transactions on Technology and Society, vol. (In-Press), pp. (In-Press). https://doi.org/10.1109/TTS.2023.3243395

DOI 10.1109/TTS.2023.3243395
ESSN 2637-6415

Publisher: IEEE

Training the next generation of doctoral researchers in data science: The impact on publications and beyond

Savvas Papagiannidis, *Newcastle University Business School,*

Maureen Meadows, *Centre for Business in Society, Coventry University*

Panos Panagiotopoulos, *Queen Mary University of London*

***Abstract* — Developments in data science methods have changed how we design, review and publish social science research. The impact on academic development has been multifaceted: new research opportunities have come with additional demands on training researchers to develop advanced data skills and apply them to research outputs. For early career researchers, meeting existing work demands and investing time in data skills can be a difficult proposition. In this paper, we consider the challenges that doctoral and early career researchers face when it comes to short- and long-term career goals and discuss how to collectively overcome them. Recommendations are organised around the key areas identified by the *Learning, Leading, Linking* framework. We emphasise that doctoral researchers in social sciences should be supported to develop their skills and pursue meaningful collaborations with other disciplines and external stakeholders as domain specialists.**

***Index Terms* — computational data science, doctoral training, publishing, academic careers**

I. INTRODUCTION

Data science refers to any field of research that involves the processing of large amounts of data in order to provide insights into real-world problems [1]. Data science methods have reshaped social science research over the past few years. In a short period of time, data science techniques have become essential in most fields of study and practice. Rapid growth has been spurred on by the proliferation of complex and rich data in science, industry and government, which are potentially available for research. These methodological trends have set new expectations about research questions, data sources and analysis methods [2], [3]. As a result, researchers at all levels need to understand the potential outcomes of novel data science methods and how to realise them. Such engagement can vary in depth and breadth from one researcher to another. For example, researchers traditionally familiar with quantitative methods might opt to focus on different areas of development from researchers interested in mixed methods or qualitative projects. It has been argued that such distinctions do not apply to data science methods, which can more naturally integrate quantitative and qualitative thinking [4].

The impact is also evident in the publication process. Before the widespread use of data science methods, the boundaries of existing research methods were typically defined within disciplinary requirements and improvements would be limited to software for statistical analysis. Computational data science methods have changed the landscape radically, by redefining and extending the range of sources, and uses of data in social science projects. When engaging with new methods, those interacting with complex data science projects often find themselves in a position of not being able to assess the validity of the process and outcomes. For example, researchers in information systems that attempt to apply machine learning are often guided by pragmatic and cultural considerations in terms of what academic outlets are willing to consider [5]. Over time, such considerations can affect someone's career progression as the projects tackled and outputs generated can significantly shape their research and career profile.

Considering the above, there is a pressing need to consider how we train the next generations of researchers, not just when it comes to data science skills, but also the wider competencies that can help them make the most of such skills. In this paper, we focus our attention on doctoral research programmes and how they are responding to the challenge of integrating new methods at the intersection of the computational and social sciences. In PhD research, advanced data skills can extend beyond the immediate needs of the project and publication potential, as is often the case with a variety of typical skills covered in research methods courses. Doctoral researchers have to manage time constraints, while implementing a research project and mastering the necessary toolset to become domain experts. We elaborate on these balances in the next section and focus on the key reasons for the necessity for, and significance of, data science competencies in doctoral programmes. Our subsequent recommendations highlight the need to establish meaningful collaborations with other disciplines, while building advanced data skills that will contribute to career development. We approach the topic from a publications and career development perspective. Although the insights come from our experience in Business and Management research, we believe they are widely applicable in the social sciences and doctoral training programmes in other disciplines.

II. Data Science Competencies in Research Training

Growth in data science education has been primarily driven by the introduction of taught programmes that focus on business analytics and machine learning methods, following the increasing needs for trained employment in the area. As data science methods make it possible for businesses to examine large data sets in their respective industries of operation, it is expected that graduates should have sufficient understanding of analytics and the skills to collaborate with data scientists and other experts. However, such growth has not been observed at the doctoral level and limited research has considered the training needs of doctoral programmes as the attention is on courses at the undergraduate or postgraduate level [6], [7].

Traditionally, doctoral and early career researchers are required to become familiar with baseline research skills conventionally applied within their disciplines. Increasing competition in maturing academic fields, coupled with the expectation of interdisciplinary collaborations, have led researchers to explore novel ways to undertake their research. Following these developments, data science techniques have created new research opportunities by adding to our methodological arsenal and novel ways of tackling research topics.. *"The single biggest stimulus of new tools and theories of data science is the analysis of data to solve problems posed in terms of the subject matter under investigation. Creative researchers, faced with problems posed by data, will respond with a wealth of new ideas that often apply much more widely than the particular data sets that gave rise to the ideas."* [8, p. 22] However, at the same time new data science techniques created new training requirements. Apart from a small number of specialised programmes, few doctoral courses can confidently claim that they have caught up with the methodological advances that harness the potential of data science. Introducing data science methods at the doctoral level could face issues similar to those faced at lower degree levels, like lack of motivation to engage with quantitative methods or limited background in statistics and programming. Still, the fundamental difference is in the range of applications and motivations for learning: taught courses focus on employability and practice, while doctoral courses focus on research and academic development [9].

Starting from the curriculum side as a broad guide, doctoral training programmes could potentially cover topics related to a) algorithms and applications, b) data management and c) analytic methods [10]. For example, algorithmic applications need to consider how to tackle domain-specific research questions and data management actions needs to consider competencies related to data visualisation, validation, classification, integration, wrangling and data infrastructure. Analytics can include competencies related to statistical inference, study design, diagnostics and machine learning. These competencies will typically be highly interrelated, requiring sufficient development of a set of prerequisite competencies before making the most of another area. For instance, even if someone is well versed in statistics, it will not be possible to engage with data analytics without having sufficiently developed coding skills. Similarly, it may be impossible to effectively present findings without the required visualisation skills that are appropriate to the project undertaken. In complex projects, the exact skillset needed can be difficult to outline in advance. In addition to the above, the skillsets developed may not be fully transferrable from one project to another even within the scope of the same PhD research, hence requiring additional investment in new learning and practise. These learning challenges can discourage early career researchers from engaging with complex data projects when the perceived benefits from learning investment – often measured as potential to publish peer-reviewed outputs – can be risky. Instead, in the pursuit of productivity and while operating under time constraints, early career researchers may pursue more conventional projects that result in outputs that are perceived as safer, because they only require traditional, well-established research methods. Such perceptions may be based not just on the risky proposition of having to learn and execute a new methodological approach, but also on how these are perceived by other stakeholders, such as journal editors and reviewers.

Still, if social science studies do not include data science competencies beyond the ones typically covered in quantitative methods, they risk limiting the scope of research objectives that are topical and relevant to theory and practice. For example, a shortage of relevant skills and competencies limits the potential of projects to utilise mixed methods approaches, where qualitative insight can be combined with sources like social media data or crowdsourcing [11], [12]. This particularly applies to interdisciplinary projects, defined as "*research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or field of research practice*" [13, p. 26]. Such projects require innovative techniques of methodological synthesis to enhance contributions to knowledge and thus the chances of converting research into a strong peer-reviewed output. This raises an interesting dilemma: should researchers opt for more outputs based on existing methods that are likely to make incremental contributions to knowledge, or invest in applying and developing new methods that could have greater impact, but require more time to produce tangible outputs? Reflections such as the ones by Meng [4] strongly support positive thinking about the integration of tasks traditionally viewed as qualitative (e.g. interpretation, sense-making) to enhance all types of projects where data science methods are applied.

It is worth noting that reflections like the above do not imply that complex data science methods will *a priori* result in better PhD projects or increased chances of publication in highly esteemed outlets. On the contrary, research should be driven by "big questions" and not by "[big] data" [11] or, as captured succinctly by Boyd and Crawford [14, p. 663]: *"[big data involves] mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy".* Therefore, more attention should be paid to creating the necessary research capacity to deal with data in creative ways that are still appropriate and relevant to the research objectives. Neither does it follow that introducing data science in doctoral programmes is to say that such training would result in cross-functional, cross-discipline, omni-knowledgeable researchers who can tackle any research objective that can deliver valuable analytical insights from day one. Research is a journey that often lasts many decades. As such, there is a need for a long-term orientation when it comes to defining the set of competencies that future researchers will need. Such competencies will not only help doctoral students tackle the immediate objectives of their thesis, but also prepare them for future projects. The latter could be relatively speaking more important as a thesis requires specific knowledge and skills to make an original and substantive contribution, but graduates need a broader repertoire of analytic skills [15].

Developing such skills within a supportive environment and following a structured pedagogic approach can lay the foundations for students to independently extend their data science skills later. Starting a new job and the demands for immediate outputs can often place a great deal of pressure on early career researchers, leaving little time to learn and apply new methods. Quantitative methods are expanding all the time, in terms of both the development of existing and the creation of new methods, making it a challenging proposition for someone to keep up with such developments [16]. As a result, students may choose to overlook methods that are better alternatives to ones they already know – or even to decline to tackle problems that are beyond their immediate methodological capabilities. Hence, doctoral students need to recognise the value of "*sustained lifelong commitment to learning, retooling, knowledge refreshing, and dynamic skill building. This is neither easy, quick, nor necessarily intuitive; however, it is absolutely essential for a perpetually successful career*" [9]. We draw on these observations as a springboard for presenting further recommendations in the next section.

III. DEVELOPING RESEARCH SKILLS AND COMPETENCIES

To consider potential ways to address the challenges faced in social science doctoral training, we organise our recommendations using the "*Learning, Leading, Linking"* framework [17] as a guiding approach. The framework was developed with consideration to the organisational challenges and tensions of big data applications. It covers areas that are pertinent to the context of this paper, namely both the trainee perspective – the doctoral researcher – and the organisational perspective, such as designing social science graduate programmes. The framework identifies three areas of tension related to big data applications: organisational learning (*Learning*), organisational leadership (*Leading*) and societal tensions (*Linking*).

Organisational learning or *Learning* highlights aspects that sit principally within the boundary of the organisation in the development of hard skills, soft skills and learning processes. These are at the core of doctoral training programmes and our main area of attention. Organisational leadership or *Leading* refers to connections between the internal and external environments in the form of strategic vision and dynamic responses to data-driven applications. Societal tensions or *Linking* refers to challenges mainly driven from outside the organisation – this provides the perspective of embedding training into data ecosystems and in response to challenges around new forms of data and data science methods [18].

Using the framework as a guiding lens, we put forward a set of recommended actions – as summarised in Table 1 – on how data science applications can be better embedded into doctoral training. We consider how such recommendations can impact doctoral training that aims to develop future research leaders, irrespective of whether they are in an academic or commercial organisation. Building on insights developed in the previous section, we believe that such a perspective is necessary not only for immediate employability prospects or realising publication strategies (shorter-term career goals), but also for preparing researchers for changes in how data-driven environments affect longer-term considerations of theory and practice. This consideration is necessary in response to fundamental changes in expectations about how social science researchers can access and deploy data science methods.

TABLE I
LEARNING, LEADING, LINKING FRAMEWORK RECOMMENDATIONS

Challenges and areas of attention		Key recommendations
Learning	*Hard skills*	– Provide regular training and upskilling in response to changes in data science practice and publication expectations. – Involve external partners to fill gaps in advanced skills and support doctoral supervision and training. These could include interdisciplinary collaborations within universities or engagement with national data centres, institutes and government statistical units.
	Soft skills	– Provide training that builds awareness of the individual characteristics (such as personality traits and competencies, motivation, social skills, etc.) that support effective use of data skills. – Introduce training on soft skills that are important for complex data projects, which are often interdisciplinary in nature.
	Learning processes	– Create awareness, not just of how to operate in a diverse environment in relation to data, but also how to contribute to such an environment, for example, how to actively address data management and data quality issues in complex projects. – Ensure that researchers have experience of engaging with data science experts and other domain specialists in at least one interdisciplinary project. This could be organised in the form of a data study group, consulting project or placement.
Leading	*Strategic vision*	– Establish academic development that incorporates data skills as a key dimension for all research activities. For example, create awareness of how academic projects can develop comprehensive strategies that reflect data management challenges. – Create mentoring schemes that encourage not just sharing of experiences but also create opportunities for long-term collaborations.
	Academic impact	– Explain how researchers can use data-driven approaches to create value for stakeholders beyond academic publications, for example, in providing training related to the sustainable data management of outputs from research projects.
	Data knowledge practice	– Facilitate knowledge sharing and participation in mentorship networks and communities with expertise in data science practice.
Linking	*Data ecosystems*	– Introduce training that creates awareness of diverse data ecosystems and their associated characteristics, opportunities and challenges (e.g. smart cities, health data, data utilities, marketplace platforms). – Develop an appreciation of the differences between academic-to-academic, academic-to-business and academic-to-government collaborations. – Provide training that integrates the perceived legitimacy and trustworthiness of data in the eyes of key stakeholders involved in projects.
	Responsible data science practice	– Incorporate guiding principles, ethical considerations and practical examples of responsible use of data science methods.

Learning starts with the availability of high-quality skills training when it comes to managing, analysing and presenting complex datasets. Such training needs to incorporate both the methodological perspectives as well as guidance on what different disciplines, interdisciplinary approaches and publication outlets expect in terms of the application of new data methods (e.g. [2], [4], [9], [19]). Collaborations with specialist providers like national data centres, institutes and government statistical units can significantly enhance both baseline skills for all researchers as well as specialist help for those who need more complex training. For instance, the Alan Turing Institute in the UK provides additional training and enrichment placement courses for all researchers working on projects related to data science and artificial intelligence [20]. Enabling all doctoral projects to tackle topical issues with suitable data skills can entail significant productivity gains and positive spill-over effects compared to projects that follow more traditional methods. For example, projects that tackle data that are not big in size can also benefit from skills in data collection and management. Another area of significant advancement in social sciences comes from the interaction between qualitative and quantitative methods [4]. Methods like sentiment analysis and natural language processing help blend the boundaries between descriptions of large datasets and researchers' subjective interpretations. Subsequently, these methods have gradually paved the way for publications that deploy mixed methods on new datasets and topics (e.g. [21], [22], [23]).

Further to the knowledge of methods and tools, soft skills are significant enablers to support the effective use of data in academic and commercial projects [24]. Soft data science skills include aspects like communication and social skills, and awareness of key personality traits and competencies. For instance, researchers need to develop their ability to communicate with colleagues from different disciplines and be able to relate to their disciplinary norms and contemporary issues of data practice. Doctoral training programmes tend to focus more on hard skills, with much less attention given to soft skills. This may be due to time constraints, or even because there is an implicit assumption that such skills have been developed sufficiently in previous taught programmes.

Learning processes are equally relevant to the direct provision of skills. Being able to adapt to different contexts and offer creative solutions to complex problems that have both academic rigour and practical relevance is a non-trivial task even for senior researchers. This is why doctoral students and early career researchers need to develop such skills and continue practising them within a supportive environment. A deep appreciation of data science practices cannot be achieved by working on graded assignments as part of a training programme or from working on one's own thesis. Creating opportunities to engage in interdisciplinary projects or to engage with external partners and actively support the engagement process can demonstrate in action the benefits of being involved in such projects [25]. It can be difficult for interdisciplinarity to be attained and sustained if it is only added as the wrapping layer of disciplinary-oriented sub-projects. Put differently, instead of joining up disciplinary projects, it would be more effective to jointly design them from the outset.

Such recommendations have been made directly in data science teaching [26], and doctoral training can apply them more broadly to interdisciplinary training tasks in the form of data study groups, consultancy projects or short placements [27]. These training tasks use data as a resource that brings together different disciplinary perspectives. Being involved in multiple and ideally diverse projects makes it possible for researchers to realise that even though there might be an overlap between the skill sets required to tackle two different projects, there may also be gaps. Filling small skills gaps from one project to the next could be more manageable than leaving skills development on the side for a long time and then having to make strides to catch up. This will help researchers develop an appreciation of the need to adopt a learning culture and to adopt processes that can sustain it. For example, skills in co-designing data infrastructures and practices are equally valuable for communities that attempt to scale up their data [28].

From *Learning* to *Leading*, we note the significant emphasis that training needs to put on supporting processes early in someone's academic career. The transition from doctoral studies to a first academic job and tenure is highly demanding for most graduates and not one of immediate consideration to researchers before completion of their degrees. Institutional inductions complemented by mentorship and other schemes (e.g. [29], [30]) provide the foundations for early career researchers to transition to a leading role with independence in selecting research projects and priorities. Time pressures usually lead early career researchers to prioritise immediate tasks like designing and delivering teaching material. Accordingly, research priorities are driven by the urgency to produce quick wins such as maximising the publication quality from one's doctoral thesis so as to establish their position and – in many academic systems – work towards securing tenure. During this intense period, publication efforts might be focused on conventional projects perceived as a safer proposition, with fewer opportunities to engage in continuous learning and skills development. Following a stage when someone's career settles down, returning to learning is not facilitated due to priorities like management roles and supervision duties.

To enable the *Leading* elements of data training practice, academic departments need to reflect on how workload policies support academic development. Managers could consider separating expectations of delivery (e.g. working towards publishing a paper) from academic development (i.e. working towards developing skills for publishing papers) and safeguard the latter. Similarly, they could introduce regular and structured programmes that academics can attend to upskill themselves. The rising role of academic networks is another significant source of knowledge exchange in data science practice. Academics should be encouraged to participate in these communities for the learning opportunities. In addition, mentoring schemes within academic networks could be reorganised to encourage collaborations among staff that have the necessary skills and those that need to develop them.

A researcher's transition from consolidating data skills and learning to a leading phase is the embeddedness of a research vision of data management as a key dimension for all research activities. We recommend that this should include comprehensive strategies that reflect the threats and opportunities of complex data management challenges in academic projects. For example, researchers at this phase of their career will be expected to submit applications to funding bodies, where data has to be appreciated both as a resource and an area of methodological innovation. It is common that early career academics acquire this experience via personal networks and informal contacts more than training that builds strategic awareness of data as a resource. Being aware of wider expectations set by external bodies is good practice and can help inform research design and how research projects are undertaken.

Academic impact is another area of change and opportunity in relation to data science practice. Academics are under revamped pressures to translate their research to have impact in ways that maximise universities' performance based on national evaluation frameworks; a notable example of this type of translation is the UK Research Excellence Framework [31], [32], [33]. Training that maximises the outreach opportunities of data skills and data science methods can enhance the potential to generate impact and value for stakeholders beyond academic publications. This context opens up multiple opportunities for academics with advanced data skills to collaborate with organisations that seek suitable expertise to enhance their data resources. Training, for example, could help academics learn how to publish datasets in specialised outlets that enhance the reach of their work.

In the framework's third area of attention, *Linking* provides consideration of data science methods outside the narrower boundaries of academic development and career progression. Academic engagement with data science can support researchers with training that creates awareness of diverse data ecosystems in areas like smart cities, sustainability, healthcare research and service ecosystems (e.g. [34], [35], [36], [37]). Organisations participating in such networks and ecosystems face challenges such as creating shared data agendas and addressing the factors that hinder the effective flow of data and knowledge. In these cases, well-trained academics can take a leading role in managing the perceived legitimacy and trustworthiness of data resources in the eyes of key stakeholders like intermediary organisations [37], [38]. In doing so they could become focal points of data ecosystems, which can result in significant benefits for the academic institutions involved. Academic institutions, however, cannot rely on the good faith of those involved but need to play an active supporting role themselves. When it comes to creating such a vision and operationalising it, an important difference between academic and commercial organisations is that academic institutions often rely heavily on individuals to act. Academics can choose what to study and how best to go about their research. This is why academic institutions need to focus more on providing incentives for colleagues to invest in data science development and at the same time create an environment that exposes them to data science and encourages them to actively practise it. At the individual level an awareness with regards to digital vision and the external environment is arguably even more important when operating in an academic environment versus a commercial one. On one hand academic institutions, their leaders and staff need to serve their own internal vision, which empowers researchers, as discussed above. At the same time as data is often made available by external organisations, institutions and researchers need to understand the implications of negotiating and accessing data [38]. Being able to relate to external stakeholders' agendas can make it possible to approach those agendas in ways that are aligned with their vision and goals. The value proposition put forward by the academic partner to the data sharing organisation cannot be one that revolves solely around publishing papers, which is a narrow view. Academic propositions need to have a significant impact dimension built into them, aiming to make a demonstrable positive contribution beyond academia. Such research output can make research a very exciting and fulfilling experience, not to mention that it strengthens external relationships, which in turn can have a wide range of effects on academic activities.

Finally, we note the continuous and critical need for academic development to incorporate responsible use of data science methods beyond traditional ethical considerations of research methods. Responsible data science is rapidly evolving as a concept and includes issues like data compliance, algorithmic fairness and diversity, transparency of data and algorithms, privacy, and data protection [39], [40]. Although training will include these aspects from the early stages of academic development, researchers need to both keep up with developments and good practice as well as understand how these issues actually apply to their research and its impact and public engagement.

IV. CONCLUSION

We have outlined the pragmatic issues in the development of data science research skills with a focus on doctoral and early career researchers, building on Cleveland's [8, p. 24] reflection from over 20 years ago: "*Education in data science does many things. It trains statisticians. But just as important it trains non-statisticians, conveying how valuable data science is for learning about the world"* Although much has changed, the development of data science practice is reaching important milestones as a widespread resource for researchers who would not traditionally engage with data and data science methods. As such, the time is ripe to reflect on our current position and ways forward in a positive manner. The challenges faced are not specific to doctoral or early career researchers. The stakes are higher, though, when it comes to them as they are the next generation leaders. Unless they have the time and space to develop themselves and form a positive attitude towards data science, the challenges identified will continue to affect research capacity and outputs in the long term.

Recommendations such as those put forward by the *Learning*, *Leading* and *Linking* elements of the framework might help address challenges in a holistic manner, while taking into consideration broader pressures and expectations in the academic environment. The framework emphasises the point that while some actions may be viewed as sitting principally within the boundary of a focal organisation such as a university (e.g. Learning: the development of core skills and capabilities), other aspects connect the internal and external environments (e.g. Leading: the need for vision, strategy, oversight etc.); and yet others are mainly driven from outside the focal organisation, and benefit from close working with a range of partners and stakeholders, including potential employers and policy makers across the for-profit and not-for-profit sectors (e.g. Linking: embedding the focal organisation within its wider ecosystem, and responding to societal issues). We hope that the application of the ideas presented in the framework will raise further practical considerations and issues for future research in academic development and training.

REFERENCES

[1] Turing Institute, "Defining data science and AI." Turing.ac.uk. https://www.turing.ac.uk/news/data-science-and-ai-glossary (accessed Jun. 13, 2022).

[2] G. George, M. Haas, and A. Pentland, "Big Data and Management," *Academy of Management Journal*, vol. 57, no. 2, pp. 321–326, 2014.

[3] J. Mahmoodi, M. Leckelt, M. W. H. van Zalk, K. Geukes, and M. D. Back, "Big Data approaches in social and behavioral science: four key trade-offs and a call for integration," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 57–62, 2017.

[4] X. L. Meng, "Data Science: A Happy Marriage of Quantitative and Qualitative Thinking?" *Harvard Data Science Review*, vol. 3, no, 3, 2021.

[5] B. Padmanabhan, xiao fang, N. Sahoo, and A. Burton-Jones, "Machine Learning in Information Systems Research," *Management Information Systems Quarterly*, vol. 46, no. 1, pp. iii–xix, 2022.

[6] A. L. Phelps and K. A. Szabat, "The Current Landscape of Teaching Analytics to Business Students at Institutions of Higher Education: Who is Teaching What?," vol. 71, no. 2, pp. 155–161, 2017.

[7] S. Schiller *et al.*, "Build Your Dream (Not Just Big) Analytics Program," *Communications of the Association for Information Systems*, vol. 37, no. 1, 2015, doi: 10.17705/1CAIS.03740.

[8] Cleveland, W. S. "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics." *International Statistical Review* 69, no. 1, pp. 21–26, 2001.

[9] K. Pan, A. I. Blankley, M. J. Mazzei, C. F. Lohrke, J. B. Marshall, and C. M. Carson, "Surveying industry advisors to select data analytics topics for all business majors," *The International Journal of Management Education*, vol. 16, no. 3, pp. 483–492, 2018.

[10] I. D. Dinov, "Modernizing the Methods and Analytics Curricula for Health Science Doctoral Programs," *Frontiers in Public Health*, vol. 8, p. 22, 2020.

[11] B. McKenna, M. D. Myers, and M. Newman, "Social media in qualitative research: Challenges and recommendations," *Information and Organization*, vol. 27, no. 2, pp. 87–99, 2017.

[12] S. Papagiannidis, E. W. K. See-To, D. G. Assimakopoulos, and Y. Yang, "Identifying industrial clusters with a novel big-data methodology: Are SIC codes (not) fit for purpose in the Internet age?," *Computers & Operations Research*, vol. 98, pp. 355–366, 2018.

[13] National Academies, "Facilitating Interdisciplinary Research", The National Academies Press: Washington D.C., USA, 2005.

[14] D. Boyd and K. Crawford, "Critical Questions for Big Data," *Information, Communication & Society*, vol. 15, no. 5, pp. 662–679, 2012.

[15] G. T. L. Brown, "Doctoral Education in Quantitative Research Methods: Some Thoughts about Preparing Future Scholars," *Frontiers in Applied Mathematics and Statistics*, vol. 3, p. 25, 2017,

[16] F. A. Hodis and G. R. Hancock, "Advances in Quantitative Methods to Further Research in Education and Educational Psychology," vol. 51, no. 3–4, pp. 301–304, 2016.

[17] M. Meadows *et al.*, "Tension in the data environment: How organisations can meet the challenge," *Technological Forecasting and Social Change*, vol. 175, p. 121315, 2022.

[18] K. Michael and K. W. Miller, "Big Data: New Opportunities and New Challenges", *Computer*, vol. 46, no. 6, pp. 22-24, 2013.

[19] S. P. Kemp and P. S. Nurius, "Preparing Emerging Doctoral Scholars for Transdisciplinary Research: a Developmental Approach." *Journal of Teaching in Social Work*, vol. 35, no. 1-2, pp. 131–50, 2015.

[20] The Alan Turing Institute. *"Skills at the Turing"*. Retrieved from https://www.turing.ac.uk/opportunities-turing/skills-turing, 2022

[21] Chatterjee, S. "Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents", *Decision Support Systems*, vol. 119, 14-22, 2019.

[22] Kang, Y *et al.* ,"Natural language processing (NLP) in management research: A literature review", *Journal of Management Analytics*, vol. *7*, no. 2, pp. 139-172, 2020.

[23] Kumar, S., Kar, A. K., & Ilavarasan, P. V. "Applications of text mining in services management: A systematic literature review". *International Journal of Information Management Data Insights*, vol. 1, no. 1, 100008, 2021.

[24] Börner, K. et al. "Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy." *Proceedings of the National Academy of Sciences* 115, no. 50, 12630-12637, 2018.

[25] Asamoah, D. A., Doran, D., Schiller, S. "Interdisciplinarity in data science pedagogy: a foundational design." *Journal of Computer Information Systems, vol.* 60, no. 4, 370-377, 2020.

[26] Navarro, Joan, Agustín Zaballos, David Fonseca, and Ricardo Torres-Kompen. "Master as a Service: A multidisciplinary approach to Big Data teaching." In *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, pp. 534-538, 2019.

[27] Lasser, J., Manik, D., Silbersdorff, A., Säfken, B., Kneib, T. "Introductory data science across disciplines, using Python, case studies, and industry consulting projects." *Teaching Statistics* vol. 43, S190-S200, 2021.

[28] Sterner, B., Franz. N. M. "Taxonomy for humans or computers? Cognitive pragmatics for big data." *Biological Theory* vol. 12, no. 2 99-111, 2017.

[29] Latifa, J., Acosta Maestre, H. "The Data Science Mentoring Fire Next Time: Innovative Strategies for Mentoring in Data Science." In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3597-3600, 2020.

[30] Kross, S., Guo, P. J. "Practitioners teaching data science in industry and academia: Expectations, workflows, and challenges." In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1-14, 2019.

[31] UKRI. *"Defining impact".* Retrieved from https://www.ukri.org/councils/esrc/impact-toolkit-for-economic-and-social-sciences/defining-impact/, 2022.

[32] Jensen, Eric A., Paul Wong, and Mark S. Reed. "How research data deliver non-academic impacts: A secondary analysis of UK Research Excellence Framework impact case studies", *PloS one* 17, no. 3, e0264914, 2022.

[33] Reed, M. S., Marie Ferre, J. Martin-Ortega, Rachel Blanche, R. Lawford-Rolfe, Martin Dallimer, and J. Holden. "Evaluating impact from research: A methodological framework." *Research Policy* 50, no. 4, 104147, 2021.

[34] Oliveira, S., Iury, M., Glória de Fátima Barros Lima, and Lóscio. B. F. "Investigations into data ecosystems: a systematic mapping study." *Knowledge and Information Systems*, vol. 61, no. 2, pp. 589-630, 2019.

[35] Gupta, A., Panagiotopoulos, P. and Bowen, F. "An orchestration approach to smart city data ecosystems." *Technological Forecasting and Social Change,* vol. 153, 119929, 2020.

[36] Beirão, G., Patrício, L. and Fisk, R. P. "Value cocreation in service ecosystems: Investigating health care at the micro, meso, and macro levels." *Journal of Service Management*, vol. 28, no. 2, pp. 227-249, 2017.

[37] A. L. Franklin, A. Grossman, J. Le, and M. Shafer "Creating Broader Research Impacts Through Boundary Organizations." *Public Administration Review*, vol. 79, no. 2, pp. 215 - 224, 2018.

[38] Tritter, J. "Negotiating access and establishing expertise." In *Reflections on research: the realities of doing research in the social sciences.* Open University Press, 2004.

[39] Lewis, A. and Stoyanovich, J. "Teaching responsible data science: Charting new pedagogical territory", *International Journal of Artificial Intelligence in Education*, pp.1-25, 2021

[40] Taylor, Linnet, and Nadezhda Purtova. "What is responsible and sustainable data science?", *Big Data & Society* 6, no. 2, 2053951719858114, 2019.