Bringing about the data revolution in development: What data skills do aspiring development professionals need?

Rachel Bennett, University of Gloucestershire, UK

Gianmarco Alberti, L-Università ta' Malta, Malta

Aytekin Çibik, Gazi Üniversitesi, Turkey

Tatiana Eremenko, Universidad Nacional de Educación a Distancia, Spain

Saviour Formosa, L-Università ta' Malta, Malta

Janice Formosa-Pace, L-Università ta' Malta, Malta

María Jiménez-Buedo, Universidad Nacional de Educación a Distancia, Spain

Kenneth Lynch University of Gloucestershire, UK

Leire Salazar, Universidad Nacional de Educación a Distancia, Spain

Paloma Ubeda, Universidad Nacional de Educación a Distancia, Spain

Abstract

The post-2015 agenda calls for a 'data revolution in development' and recognises that statistical capacity amongst the workforce is a prerequisite for achieving it. Universities have a critical role to play in building this capacity. This paper reports insights from in-depth interviews with development professionals in Malta, Spain, Turkey and the UK about current and future data skills needs. It presents recommendations on data skills training for aspiring development professionals informed by this evidence, with a particular focus on curriculum in social science related undergraduate programmes at universities.

Key words: data revolution; global development; statistical capacity; data skills; universities

Introduction

The High-Level Panel appointed to advise on the post-2015 agenda in development called for a 'data revolution' (United Nations, 2013). Data is crucial for informing decision-making and the design of policy and programmes, enabling rapid and continuous monitoring and evaluation, as well as supporting accountability in all stages of development work (United Nations, 2014). Whilst efforts to monitor progress towards the Millennium Development Goals (MDGs) were hampered by weaknesses and blind spots in statistical systems (Espey et al., 2015), the Sustainable Development Goals (SDGs), as well as being directly measurable and requiring fine-grained data, include a target to increase statistical capacity in itself.

Data brings opportunities to help tackle major issues, such as the COVID-19 pandemic, but also carries justice risks which need to be recognized and addressed (Oldekop et al., 2020). Thus, human capacity and judgement play a critical role at all stages of working with data from the data collection and decisions over categorisation, to interpretation and visualisation, to communicating often complex findings to non-specialists (United Nations Global Pulse, 2012). PARIS 21 and Open Data Watch (2017) argue that support for statistical capacity building must be much more expansive than previous efforts, focusing on building partnerships and on enabling a wider range of users to understand and promote data (cited in Kejzer and Klingebiel, 2017).

Academia has a leading role to play in building efforts to strengthen data skills of professionals aspiring to work in development (UN, 2014). Among the three core competencies for international development graduates identified by Woolcock (2007; 57), one directly relates to the more technical data skills – the 'detective' (locating, generating, analysing, and interpreting information) – alongside the other two key competencies of the 'translator' (mediating a dialogue between very different constituencies—policymakers, managers, field staff, villagers, local officials, academics, donors), and the 'diplomat' (brokering differences, doing deals, moving agendas, negotiating agreements).", which, in the context of data work, relate to the softer skills associated with successful use of data. 'Soft' skills are understood as an umbrella of attributes and competencies including communication, leadership, teamwork, negotiation and broader emotional intelligence (Deepa and Seth, 2013). This paper adopts the holistic definition of data skills used in the UK Government's' Quantifying the UK Data Skills Gap' 2021 report that data skills are 'any skills that involves the gathering, interpretation and communication of data, often as part of a team'. This definition encompasses both technical skills in manipulating and analysing data and the softer skills involved in understanding data and its implications and expressing findings to a range of audiences. Thus this definition relates to all three of the core competencies of International Development graduates identified by Woolcock (2007).

Despite being well-placed to support the future development sector workforce to enhance their data skills, universities face challenges around how to enthuse students about this area of skill development. A persistent theme in the pedagogic literature on students not specifically enrolled in data or statistics degrees is the need to overcome their statistical anxiety: "the apprehension that occurs when an individual is exposed to statistics content or problems and instructional situations, or evaluative contexts that deal with statistics" (Macher et al., 2015, p.1). Such anxiety has been linked to a lack of confidence, particularly amongst students with a limited mathematical background and/or who have had bad experiences of mathematics teaching earlier in their education (Hodgen 2010). An additional barrier to students' engagement in data education can be a simple lack of interest and perceived relevance to their discipline and career aspirations (Willams et al., 2015, Carter et al., 2017, Chamberlain et al., 2015).

This paper seeks to understand how the "data revolution" in development is viewed and experienced by development actors, how data skills are used within their organisations and projections on this for the future, as well as the eventual implications of these evolutions for professionals in this sector. The paper draws on expert interviews with development sector professionals across Malta, Spain, Turkey and the United Kingdom, and based on these insights makes recommendations about the content and style of data skills curriculum in support of the data revolution in development. The focus is predominately on the curriculum for social science related undergraduate programmes, which are commonly selected by students aspiring to work in development, and where students are often particularly reluctant to develop data skills (Ralston, 2020). However, the paper also has relevance for undergraduate data skills curriculum in other disciplines not specifically focused on statistics and for current and aspiring development professionals who are not currently undergraduate students.

Debates in Data Skills Education & their Relevance to Global Development

The choice of issues covered in the interviews in this paper were informed by core debates in data skills education in relation to the development sector. Firstly, the level of knowledge and competence in data science required amongst future graduates. Much of the debate about data in global development has been focused on the potential of new advances, particularly the use of 'big data' and sophisticated, cutting-edge methodologies (Aiken et al., 2020; Jacobson, 2018; Tatevossian, 2013). However, critiques of this narrative have argued that using big data may not necessarily produce higher quality analysis than existing data sources (censuses, surveys, administrative statistics), and that a clear understanding of the issues being investigated should be the priority (Latif et al., 2019). In terms of this debate in the pedagogic literature, MacInnes et al. (2016) cites the case of the team running the University of Auckland's first year statistics course who advocate for utilising advances in data science to equip non-statistical specialist students with a broader range of technical competencies. However, others, including Schuff (2018), argue that *data literacy,* rather than statistical or technological proficiency per se should be prioritised as a core skill which can be offered to all undergraduates.

Secondly, and related to the previous debate, is the identification of software/s development professionals view as most useful in the sector. University instructors have often used licenced software (e.g. STATA or SPSS) which are well supported and resourced. However, there has been a move towards open-source software such as R and Python (amongst others) in both academic and non-academic settings. In a survey of analytic communities conducted by Lavastorm Inc in 2013, R was a more popular self-service analytic tool than SPSS (35.3% vs 22.8%). R and Python have powerful and flexibility utility, but predominately rely on users writing code which often means a steep learning curve, although there is some point and click functionality too. Moving aside from specific statistical programmes, Microsoft Excel is widely used in the 'real world'. Microsoft Excel is a general-purpose programme with statistical functions rather than being primarily designed for data analysis which

leads to a number of limitations in its utility (Li, 2019). However, a fundamental advantage of Microsoft Excel is its availability as it is produced by Microsoft and part of the Microsoft Office package. In the Lavastorm Inc survey of analytic communities mentioned previously Microsoft Excel was by far the most popular self-service analytic tool (selected by 75.6% of respondents). Thus, there is a bigger question about whether students aspiring to work in development would benefit from competence in simpler and very widely used programmes such as Microsoft Excel, or with an eye to the future, whether more advanced software where users code is to be preferred in their training.

Finally, there is a debate on the balance between technical and soft data skills. Often data skills training in social science degrees forms part of research methods modules which focus heavily on technical data skills in analysis and very little on softer data skills and are run in isolation to modules on the discipline subject matter. For example, in their report on quantitative skills training in UK social science, MacInnes et al. (2016) highlight poor integration between technical data analysis training and substantive modules and limited space on the curriculum devoted to developing and consolidating all types of data skills. Global development is a field where collaborative working across cultures is key (Villarino, 2014). Indeed, Jütting et al. (2019) identify reducing co-ordination challenges between stakeholders as one of three key priorities for strengthening statistical capacities in the SDG era, for which they mention the need for leadership, negotiation and communication skills. Therefore, there may be a particular need for soft data skills to be prioritised in support of technical data skills for global development.

Methods

In order to address the exploratory nature of this investigation, semi-structured interviews were carried out with development professionals based in Malta, Spain, Turkey and the United Kingdom, as part of a project supported by the ERASMUS+ programme of the European Union. The policy, economic and institutional contexts vary across the four nations in this study. Mangen (1999) argues that there are strengths in the application of qualitative interviews when exploring cross cultural perspectives on employment, work and social issues, including in multilingual intra-European projects such as this one.

Ethical approval for the research was granted by the School Research Ethics Panel of the School of Natural & Social Sciences, University of Gloucestershire (NSS/2002/001) for the interviews in the UK, the Faculty for Social Research Research Ethics Committee, University of Malta for the interviews in Malta, the Ethics Committee of the Universidad Nacional De Educación a Distancia for the interviews in Spain and the Ethics Committee of Gazi University (document Gazi Üniversitesi Etik Komisyon Onay Belgesi 2021-25) for the interviews in Turkey.

National contexts

This section provides a brief overview of Overseas Development Assistance spending and distribution, as indicators of government led work on development, and the nature of development non-governmental organisations (NGOs) in each country in order to contextualise development work by professionals based in each country.

Malta contributed 0.3% GNI to ODA in 2019 (OECD, 2020). This is largely directed to tackling the refugee crisis that affects the country due to its proximity to the Libyan coast, and a substantial component is also assigned to multilateral agencies (Malta Ministry of Foreign Affairs, 2018). A total of 1,479 NGOs has been registered to date in the Maltese Islands. The largest proportion of voluntary organisation employees work in the Health and Disability Sector, followed by the Education and Sports, Social and Humanitarian, Environment and Animal Welfare and lastly the Arts and Culture.

Research exploring cost effectiveness within the voluntary organisation sector showed that voluntary organisations employees are paid less than employees in the same position working in the private or state sector and are more willing to work after hours without renumeration (MS Advisory Services, 2016).

Spain's contribution to ODA is lower than other OECD countries (0.2% of the GNI in 2019) but has been increasing in recent years after the budget cuts during the 2008 economic crisis (OECD, 2020). This aid is distributed to organisations based in Spain and abroad, with the largest cooperation efforts in Latin America and Western Africa. Most Spanish-based NGOs are small or medium-sized, but the number of larger NGOs (more than 250 employees or income of more than €50 Million) is growing: 31% in 2017 compared to 24% in 2010. The 9 largest NGOs concentrate 78% of the overall expenditure in the sector (Coordinadora de Organizaciones para el Desarrollo, 2017).

In the early 2000s, Turkey's ODA spending as a percentage of GNI was the lowest of the four countries but has grown dramatically since 2011 and by 2017 exceeded the 0.7% of GNI on ODA UN target (OECD, 2020). The rapid rise in Turkey's ODA is linked to the country's response to the refugee crisis originating in neighbouring Syria – 70% of Turkish ODA was spent on this in 2016 (OECD, N.D.). Further, Turkey's impressive economic and social development since 2000, has provided employment and increased income levels. However, economic vulnerabilities have increased in the last few years and a more challenging external environment has emerged. Turkey is classified as an upper middle-income country and thus is also eligible to receive ODA. In 2017, Turkey ranked 7th for the amount of ODA received (OECD, 2019). The number of active non-governmental organisations in Turkey approached 112,000 as of January 2018, although this is not specific to the development sector (Ministry of Treasury and Finance and the Presidency of Strategy and Budget, 2020). The most prominent activities are vocational solidarity, humanitarian aid and education (Ministry of Treasury and Finance and the Presidency of Strategy and Budget, 2020).

The UK has met the 0.7% of GNI on ODA UN target since 2013, and a commitment to honouring this target was put into legislation in 2015 (Booth and Lun, 2016). However, this was reduced to 0.5% from November 2020 in response to the COVID-19 pandemic (UK Government, 2020). UK governmental organisations have been recognised as leaders in innovative use of data across development agencies internationally (Powell and Orton-Vipond, 2019). Development NGOs in the UK are a significant sector - collectively spending £7 billion in 2015, which is equivalent of just over half the Official Development Assistance budget in that year (Banks and Brockington, 2019). Expenditure is highly uneven with a number of very large players responsible for most of the spending: 8% of development NGOs were responsible for nearly 90% of spending in 2015, and there are many more smaller organisations (Banks and Brockington, 2019).

Interviews

The initial target was to obtain one interview from professionals working in three types of organisations in the development sector for each county: governmental organisation, larger non-governmental organisation and smaller non-governmental organisations. Non-governmental organisations were defined to be those that were not-for-profit and self-identified as working on development issues. Table 1 summarises the characteristics of the participants. It is important to note that the positions/occupations of the respondents in each country does not match. This is in part a result of the differing make-up of the development sector in each country. However, since the focus for this research is on the career opportunities for the graduates in each country, the research is more concerned with how each respondent represents the employing organisation than their basis for operation.

A common interview guide was used across countries and participants to ensure that broadly comparable topics were covered. This included questions on data science advances, software choice, and the relative importance of technical and softer data skills (highlighted previously as core debates in data skills education), as well as questions seeking insights about how data skills are used within their organisation and projections on this for the future. The interviews were completed between April 2020 and September 2020 via video call, lasted approximately 45 minutes each, and were undertaken in the usual business language of the country. In the UK, Malta and Spain all participants gave their consent for the interviews to be recorded, and these were later transcribed by the research team and translated to English. In Turkey, participants requested that detailed notes be taken in lieu of recordings, and these were also translated to English. The analyses were carefully considered by native speakers amongst the research team to ensure the messages from the translated interviews were reflective of the original interviews (see Table 2). The interview notes and transcriptions were then coded according to themes raised in the literature or additional themes identified by the research team during data analysis. Illustrative coded text was then selected and tabulated according to the themes. The national teams convened to discuss the findings and comparisons were drawn out between the countries.

Development actors at different stages of the "data revolution"

Heterogeneity in the integration of data work in current operations

There was significant diversity in the extent to which data are currently used by participants' organisations across the four countries, which is an important factor for interpreting the interview findings. Participants in Turkey, Malta and Spain discussed how data skills were less likely to be a primary requirement in recruitment and selection of candidates. Participants from the NGOs in Turkey and Spain noted that technical data skills were not integral to their day-to-day work and advanced analysis was rarely needed. This made it more cost-effective for them to employ consultants, leading to a culture where technical data-related work is often externalised:

> *"I think that both employees and management have not perceived, let's say, the need to do this work with their own staff and, therefore, to develop the competencies"* [SPAIN SMALLER NGO PARTICIPANT]

Although participants from government organisations in Turkey and Spain noted that technical data skills were not actively sought in recruitment, they reflected on how high-quality data was nevertheless important for their day-to-day work in terms of informing decision-making and enabling them to do their jobs, thus relating to softer data skills. A Maltese participant from a governmental organisation explained that they would primarily seek soft skills at the recruitment stage but would be willing to invest in training in technical skills if needed:

> *"We can't expect people to be proficient in so many different areas, before they come to us otherwise they might go to the private sector which offers a nice salary package, But training is available, whether it's in house or outsourced training"* [MALTA, GOVERNMENTAL ORGANISATION, PARTICIPANT 2]

In the UK, data and both technical and soft data skills were utilised in all of the participating organisations though only specific employees within these organisations actively used their data skills. Participants from UK governmental organisations manifest their firm commitment to using data in development work. It is also noteworthy that the smaller NGO had a specific data focus in its mission, so the extent of data used may not be typical of other smaller NGOs in the UK.

Consistent acknowledgement of the potential of data & that greater use of data 'is coming'

Although data was not integral to the day-to-day operations of many of the interviewees' organisations, there was a general sense of change in how data was used in their work, echoing a broader narrative of the increasing availability and possibilities of data. For example, several participants noted that they were experiencing a step change from relying on anecdotal evidence to quantitative evidence in support of campaigns:

> *"At least since I joined (ORGANISATION NAME) I think it there is more appetite, yes, to report these things and to be evidence based, and before my understanding, it was (ORGANISATION NAME)'s reports were based more on, like, anecdotes or some sort of interviews with a few people. But now it needs to be something bigger – some global coverage and that you need both data and data analysis to crunch all those numbers and come up with messages"* [UK LARGER NGO PARTICIPANT]

> *"So, the present provision of data is in the form of charts and graphs of our service users, reports and presentations. We would like to focus more on inferential statistics and other applicable statistics. We are being ambitious because we want to employ a statistician"* [MALTA GOVERNMENTAL ORGANISATION PARTICIPANT 1].

> *"I don't know what the trend is in terms of public behaviour and the social base, but this will give us the information we need to better design our proposals for engagement, whether we ask them to read a report or sign a campaign or ask them to volunteer with us, buy from fair trade or give money occasionally or more frequently. or to participate in '[a large-scale fundraising] campaign, right? These are the kinds of profiles that are being sought more and more"* [SPAIN LARGER NGO PARTICIPANT 1]

Even participants from organisations where data is less prominent in their current operations reflected on how this was starting to change. The participant from a smaller NGO in Spain eloquently expressed how the fact that their current workforce has less experience in this area might be holding them back from developing in this area:

> *"We know less about [these] tasks so we choose to externalise them. But it doesn't necessarily have to be like that, no. And if you have someone with skills in that area, you could leave them a little more flexibility, do analyses …which we sometimes don't do because we don't want to bring someone from outside"* [SPAIN SMALLER NGO PARTICIPANT]

Further, there were examples across country contexts of how recently hired individual employees with data skills had been able to bring about step changes in their use of data:

> *"they [NEW EMPLOYEE WITH DATA SKILLS] has generated a discipline and an institutional behaviour, right, linked to data analysis that we did not have….we generally do not work with data, but always work with secondary data, we obtain information from other institutions. But no, it was the first time that we had started with this, well 2-3 years ago more or less, that we wanted an institutional culture based on the management and linking of data. And all this with the support of this professional"* [SPAIN GOVERNMENTAL ORGANISATION PARTICIPANT]

> *"In the past, or before I joined there were some data analysis that we do every year for some reports, and they were not in STATA so there was not the coding part there that you can actually see and reproduce and see what has been done in the past, so it was a matter of kind of guessing what someone made in Excel and these things, so we are trying to implement these*

> *to make everything more consistent in STATA but also having a good practice of recording everything into there"* [UK LARGER NGO PARTICIPANT]

There are also examples of very new initiatives. Further, both participants who worked for UK governmental organisations were in sections of their organisations which had been set up within the last 5 years.

Many of the UK participants (all directly based in departments using data) held senior management positions which meant they were not directly involved in 'hands-on' data analysis or using the latest software and techniques. However, they acknowledged that there were new expectations on professionals starting their careers now in terms of more advanced technical data skills:

> *"We've got a lot of, you know, bright new people who, you know, they code in Python and R, which is a world away to what I was doing 15 years ago when I was working in Tanzania"* [UK GOVERNMENTAL ORGANISATION PARTICIPANT 2]

> *"It's actually a generational thing....we [now] want subject matter experts with data science skills"* [UK SMALLER NGO PARTICIPANT]

No clear 'way in' to a career using data skills in development

The absence of a clear career path leading to a job using data skills in development work was another unifying them, an issue already identified for careers in this sector (Woolcock, 2007):

> *"[There is] such a wide diversity of backgrounds and people have fallen into it by accident almost. They hadn't set off thinking 'I want to be a statistician in international development''* [UK GOVERNMENTAL ORGANISATION PARTICIPANT 1]

> *"It is much more difficult to find a profile, of a social scientist specialised in data to work in an international NGO than to find a data analyst who has studied business administration for a... bank, for any private institution"* [SPAIN GOVERNMENTAL ORGANISATION PARTICIPANT]

The participant from a smaller NGO in Turkey reflected on how 'data analyst' is not a standalone role in their organisation, although it is possible that if an employee has particular strengths in this area they could develop their career around this specialism. In Spain and Malta there was an acknowledgement that development organisations, and more broadly the public sector, were behind the private sector in recognising and attracting data skills talent:

> *"I think that the private sector....has understood, or has incorporated earlier, data analysis and data management as a strategic issue"* [SPAIN GOVERNMENTAL ORGANISATION PARTICIPANT]

> *"Ability to communicate clearly and concisely and have good presentation skills and ability to understand the requirements of the client coupled with coming across as being able to take initiative and exposure to suppliers, online training and trips to factories to understand product are crucial to career progression"* [MALTA SMALLER COMPANY PARTICIPANT]

The participants from Malta explained this difference in terms of the different priorities of the organisations – for those not in the private sector these types of analysis are currently useful but not crucial to operations.

Data science competencies and tools: which are the most relevant?

Enthusiasm for data science but also clear need for broader data literacy

When respondents were asked about advances in data science informing their organisation's work, there was enthusiasm for its potential, particularly around being able to do new tasks and to enhance traditional work, but it was commonly very new:

> *"More and more of our programmes are about taking and using data science and using those skills…..data science skills are being used to improve and speed up processes and enable faster processing with less mistakes, etc."* [UK GOVERNMENTAL ORGANISATION PARTICIPANT 1]

> *"It is still at a very idea level, let's say it has come up in meetings with the coordinators, programme coordinator... To investigate a little more, in more middle-income country contexts, Guatemala or Bolivia, if there is a way to use those massive amounts of data, the Big Data, etc., in our projects. For the diagnosis, but also for our activities perhaps….it is a subject that we have raised, but well, it is about to be deepened"* [SPAIN SMALLER NGO PARTICIPANT]

In addition to considering how more advanced data analysis techniques are or could be used, several participants mentioned a need to raise basic data comprehension across their organisations including for colleagues who may not directly work with data:

> *"But one of the things is that we are convinced that more and more of these types of skills will have to be mainstreamed as more instrument techniques, including in human resources, no, that is, it is not something of the data specialist who is in Public Engagement, but things that should already be thought of as more universal skills"* [SPAIN LARGER NGO PARTICIPANT 1]

> *"One thing that COVID-19 has exposed, is that ….there's a huge part of the department that doesn't have these data skills"* [UK GOVERNMENT PARTICIPANT 2]

Thus, alongside the potential of specialist technical skills in newer techniques, there was a need for raising the data competencies of the workforce as a whole.

Varied views on preferred software, but emphasis on widely available programmes

Table 3 summarises the main statistical software used by participants' organisations. Amongst the UK interviewees, there was a preference for open-source software, particularly R and Python – with the key rationales given being that they are freely available and lend themselves to code sharing and being able to replicate analysis. These packages were also used by the research department of the larger Spanish NGO (participant 2). However, it was stressed that more than knowing one particular programme, it was important to have a broad understanding of statistical software. Further, Microsoft Excel was also mentioned, in particular for exploratory phases of data before considering more advanced analysis. Microsoft Office programmes were the most cited amongst the interviewees from Turkey and Spain, whilst not open source, these are very accessible. One participant from Spain also noted that the different features available are often underused:

> *"Excel is a very powerful and often underused tool, and I think that just by knowing it better and exploiting it more, you could go a long way"* [SPAIN LARGER NGO PARTICIPANT 2]

For more advanced analysis in Turkey and Spain, SPSS was mentioned - but this work was being conducted largely by external consultants so the participants' organisations did not have licences for this software. Further in Spain, where financing of operations and staff contracts were often temporary and linked to fixed-term projects it was difficult to justify investment in training staff in data analysis. In contrast, the participants from Malta noted that both SPSS and Microsoft Excel were

used in-house and there was support for training if employees did not have competencies in these programmes.

Broad knowledge and soft skills crucial to interpreting and communicating results

High value attached to effective data visualisation & communication to non-specialists

Across governmental and non-governmental organisations participants in all countries conveyed enthusiasm for investing in data visualisation as tools to communicate key messages.:

> *"Data is a kind of statistical universe, and if it's not displayed correctly, it's no good….For a data analyst, let's say her work tool is the data matrix, that is, the whole dataset, let's say the whole dataset, is like her big tool. But for the public manager or for the person who works with countries, or who works with the institution, this is unreadable"* [SPAIN GOVERNMENTAL ORGANISATION PARTICIPANT]

From the interviews in Turkey there was a particular emphasis on getting visualisations and presentation right when writing for external audiences and government decision makers.

However, the enthusiasm for data visualisation were cautioned with a message that it is not enough to have a good visualisation, the underlying data must convey a useful message –

> *"You can produce lots of things that look very nice, but the fundamental thing is that they convey important and useful information"* [UK GOVERNMENTAL ORGANISATION PARTICIPANT 2]

The importance of being able to communicate findings was a stronger theme throughout the interviews than organisations looking for competence in particular types of analysis:

> *"Focus on dissemination, understanding your user and principles around statistical production rather than necessarily saying you know how to do a T test"* [UK GOVERNMENTAL ORGANISATION PARTICIPANT 1]

> *"And even I mean, the production of the graphs and presenting things in a way that people can actually read; presenting things in a way that people can understand easily even if we're not producing peer reviewed documents"* (MALTA GOVERNMENTAL ORGANISATION)

> *"I have worked with many data experts who are very good at analyzing the whole thing, but they can't understand or make sense of it and filter all that information so that it meets the objective that is required. I had to do my own process from delivering 20-page reports that nobody read or understood, to building 3-page interactive dashboards that included what was relevant to each project and stakeholder group. It is very important to spend time on this, and to "educate" the people who use the information"* [SPAIN LARGER NGO PARTICIPANT 2]

Data skills grounded in real world contexts

Although there were examples of projects using more advanced data science techniques this was caveated with the message that what is really valued is an appreciation of where data has come from and the 'real world' relevance of results. Several participants made it clear that they would prefer individuals with a grounding in the application of data to very advanced data skills:

"Data scientists in their own right is this some kind of amoral nondirectional geeks. Personally, I think they are slightly overvalued. Yeah, you need to understand the world" [UK SMALLER NGO PARTICIPANT]

"You know that kind of enquiring mind that doesn't just naively take data and present it but understand it - a bit about where it might have come from and what the constraints would be…. We're also looking for people to have a good understanding of, if not the statistics itself but of what it takes to produce good quality evidence" [UK GOVERNMENTAL ORGANISATION PARTICIPANT 1]

"Then there is more strategic knowledge and skills... that have more to do with "making sense" of all the information, and that does not depend so much on precise tools or knowledge, but on a skill and a sense of smell that develops with practice" [SPAIN LARGER NGO PARTICIPANT 2]

In addition, the participant from the governmental organisation in Turkey stressed the value of having real experience of working in the field in order to interpret and communicate the findings.

More generally, the ability to think critically about the context of data and how it can be used to influence decisions most effectively, encompassing an appreciation of cultural, political and institutional factors was valued:

"Understanding of the motivations and the drive behind decisions that are made in an international development context, in a government context, and an NGO context in different situations, in fragile situations….I think there is a strong need for these particular skills is acting at the boundary of the producers and the users of statistics" [UK GOVERNMENTAL ORGANISATION PARTICIPANT 2]

"Quick understanding of complex contexts because, of course, when we are given projects, people often change or have to support countries that they don't know in formulations, a follow-up of the project, which is why we all need a lot of agility, let's say in understanding different contexts, natural and linguistic peculiarities" [SPAIN SMALLER NGO PARTICIPANT]

"I think there are far too many people in this field who don't appreciate that politics is central to all of this. You know that the Political Economy of data is largely overlooked in the attempt to sort of look at a lot, a lot of development in within apolitical environment - particularly from a data point of view" [UK SMALLER NGO PARTICIPANT]

"Something about the ethical aspects is important, because information is power and is very easy to manipulate" [SPAIN LARGER NGO PARTICIPANT 2]

Conclusions & Recommendations

Responding to calls for a 'data revolution' in global development, this paper provides new insights into how development professionals view the current and future data skills needs of their sector in four countries. Based on insights from the interviews, this paper proposes four recommendations for aligning data skills training to development sector needs, focusing on how universities can adapt their curriculum on social science programmes.

Firstly, learners should be trained in data strategy as well as data skills. The *potential* of data for development was a common theme in the interviews, and participants described how individual new initiatives or hires had been able to transform or create new work processes. The term 'Data Strategy'

is more synonymous with the private sector; the Beye NETWORK states that data strategy requires 'an acknowledgement that data is a corporate asset which requires management and protection like any other business asset' and realising such a strategy means having a combination of good management and technical capacity (Eckerson, 2011; p. 3). Applying this to development means acknowledging that data can be integral to enabling organisations to reach their social goals and proactively seeking to manage how data is used throughout their work. In their piece 'Driving Social Change with Data' Day (2020) argues that the potential of data for non-governmental organisations cannot be fully realised where it exists in siloes, and that there are efficiencies and improvements to be made once the role of data in the organisation has been invested in, making it crucial to invest in developing a data strategy. The 'data for development ecosystem' is likely to look very different in the future (Epsey et al., 2015) – an idea reflected in the interviews – meaning graduates need not only to be aware of current ways of working with data, but to know how to keep abreast of the latest developments and to plan strategically to harness their benefits. A focus on data strategy in university curricula will be useful both across countries where novel ways of working with data are already integrated into their workflows and where there is currently less use of data.

Secondly, learners should be supported to think critically about the quality of evidence and what it means for the real world, whilst having an appreciation of data science advances. A core theme in the interview data was the need to have employees who could fully understand and evaluate data and use it to inform decision-making, and that data science has potential for their work in the future. Thus social science degrees should include foundational training, so that students have a basic level of competence, but also the confidence, interest and motivation to work with data and to seek further training in the future where necessary. The focus on being able to critically evaluate data also responds to some of the risks associated with the data revolution in development. These include the limitations of trying to reduce the complexity of development to quantifiable measures (Rocha de Siqueira, 2019), challenges around equality of access to data (UN, 2014), and the need to respond to ethical, privacy and security challenges and to prevent the uncritical acceptance and careless application of new methods (Latif et al., 2019).

Thirdly, learners should be competent in readily accessible software and documenting their data processes. There is little evidence from the interview data that it will be useful for students to learn to use licenced products for working in this sector in the future. Organisations that are not already using licenced software in-house are unlikely to invest in costly software when the capabilities of open-source software make them an attractive option. Students should have a foundational training in basic programmes that allow them to 'see' their data and produce simple analysis, such as Microsoft Excel, the skills for which are readily transferrable to similar open access software. For more advanced analysis, the focus should be open-source programmes with capabilities to branch into new types of data analytics, for example R or Python amongst others. Further, an additional insight from the interviews is that emphasis needs to be put on how to document exactly how the data has been sourced, cleaned and analysed in order to enable collaborative and transparent working, which lends itself to software that supports users writing code. This is a further reason not to rely solely on Microsoft Excel because it does not lend itself to making a record of exactly how data has been processed and analysed. This represents a significant departure from how many universities currently teach data analysis where the emphasis is on learning analysis techniques rather than the entire process of working with data.

Fourthly, learners should be equipped with skills for collaborating & influencing, and for communicating data messages. It is clear from the interview data that there is much more to using data for development than producing the analysis, and another area where social scientists have the potential to make a strong contribution is their ability to collaborate with a diverse range of stakeholders, influence the way evidence is used and propose innovative ways to communicate and

visualise key messages from data. To develop these skills students should be given opportunities to collaborate, including with those from other disciplines and cultures, to formally learn about strategies to influence and advocate for change, and be taught up-to-date methods for visualising data. This stems back to the UK Government (2021) definition of data skills outlined in the introduction to this paper which explicitly highlights that data work is often teamwork - thus it is valuable to gain experience in collaborative data-related work. It is well acknowledged that using applied examples based on real data is good pedagogic practice for enthusing students and helping them understand the relevance of data-related skills (Henshaw and Meinke, 2018) – and the evidence in this paper supports the importance of that approach for development sector careers, with participants stressing they are interested in graduates that understand the real-world implications of their data work.

Each of these recommendations would help prepare students for careers using data skills in development sector careers. They also serve as strategies to help overcome 'statistical anxiety' and lack of interest in data skills education. Providing foundational training in key technical skills, but then also focusing on data strategy, critical thinking about data and skills for collaborating and influencing would significantly diversify the data skills education currently taught in social science programmes. This approach would provide opportunities to integrate data skills throughout degree programmes, rather than siloing data education to technical skills taught within research methods modules given limited space on the curriculum (McInnes et al., 2016). Constraining the amount of technical content also aligns with the challenges of overcoming 'statistical anxiety' and acknowledges what Robertson (n.d.) refers to as 'educational bandwidth' - limits on the amount of content that can be delivered in quantitative methods education given that each concept must be reinforced and revisited for this learning to be successful. Building students' abilities to master techniques is likely to serve them well for having the confidence to learn new and emerging data techniques across their careers and to comprehend data even if they are in a career where data is not a central focus – which speaks to the finding from the interviews that organisations are interested in building data literacy across the whole workforce. Further, a critical application of data skills is an area where social science students may have an advantage relative to their counterparts on statistics or data science degrees because of the nature of their core disciplines. Therefore, it may be a pragmatic approach to focus on developing these softer data skills in order to give social science graduates an advantage in the labour market. In terms of the recommendation to move towards readily available software and consider coding-based programmes, despite its challenges, teaching students to use software where they need to code can support deeper analytical thinking and conceptual understanding of what they are trying to achieve relative to a 'point and click' approach. Indeed, there are recent examples of successful transitions to these types of programmes on social science courses (Li, 2019, Hulsizier and Woolfe, 2009).

There are limitations to acknowledge with this research. Firstly, as an exploratory study the number of interviews was modest, however the research team hope larger future studies can be informed by these early results. Secondly, the research team recognise that that this paper presents perspectives on the issues from the Global North. Within the countries eligible to take part in the ERASMUS+ programme partner institutions were deliberately chosen to ensure social, economic and cultural diversity and indeed the finding show similarities as well as important contrasts between development professionals in different countries. The research team hope that by sharing these findings this paper will initiate conversations on data training for development sector careers and similarities and differences across other world regions.

The research team also have two broader reflections from the findings in order to develop data skills training for development sector careers. Firstly, despite a demand for specialists with data skills in the development sector, there was often no clearly identified "way in" for a recent social science graduate, even in organisations where data related positions exist. On the other side, students, especially at undergraduate level, aspiring to work in this sector may be primarily committed to causes

and campaigns and interested in 'hands on' roles, rather than taking a more analytical, career-pragmatic approach and developing their data skills. In this context, part of universities taking a 'leadership role' in realising the data revolution in development (UN, 2014) should focus on creating opportunities for students to gain experience in and enthusiasm for data skills and working with development organisations to explore the opportunities offered by recruiting social science graduates with a rich set of both technical and "soft" data skills. A second reflection is that whilst the primary interest here was data skills for development sector careers amongst social science undergraduates, many of the conclusions are relevant more widely for professionals already working in the development sector and for universities seeking to developing graduate employability attributes more broadly and contribute to the personal development of students as global citizens in an increasingly data driven era.

Data Availability Statement

The data that supports the findings in this study are restricted to the research team. This was the level of data sharing agreed with participants via the informed consent process.

References

Aiken, E.L, Bedoya, G., Coville, A., Blumenstock, J.E. 2020 *Extended Abstract - Targeting Development Aid with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan.* In ACM SIGCAS Conference on Computing and Sustainable Societies, June 15–17, 2020, New York.

Banks, N, Brockington, D. 2019. *Mapping the UK's development NGOs: income, geography and contributions to international development*. GDI Working Paper 2019-035. Manchester: University of Manchester.

Booth, L, Lunn, J. 2016. *The 0.7% aid target: June 2016 update*. House of Commons Briefing Paper.

Carter, J, Brown, M, Simpson, K. 2017. From the Classroom to the workplace: how social science students are doing data analysis for real. *Statistics Education Research Journal*, **16**(1), 80-101.

Chamberlain J, Hillier J, and Signoretta P. 2015. Counting better? An examination of the impact of quantitative method teaching on statistical anxiety and confidence. *Active Learning in Higher Education*, **16**(1), pp.51-66. DOI: 10.1177/1469787414558983

Coordinadora de Organizaciones para el Desarrollo. 2019. *Informe sobre las organizaciones de cooperación para el Desarrollo—2017*. https://informedelsector.coordinadoraongd.org/archivo/2017/

Day, A. 2020. *Technology for change supplement: driving social change with data*. Stanford Social Innovation Review. Available: https://ssir.org/articles/entry/driving_social_change_with_data

Deepa, S, Seth, M., 2013. Do soft skills matter?-Implications for educators based on recruiters' perspective. *IUP Journal of Soft Skills*, **7**(1), p.7.

Eckerson, W. (2011) *Creating an enterprise data strategy: managing data as a corporate asset.* BeyeNETWORK. Available: http://docs.media.bitpipe.com/io_10x/io_100166/item_417254/Creating%20an%20Enterprise%20Data%20Strategy_final.pdf

Espey, J. et al. 2015. *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development*. Sustainable Development Solutions Network. Available: 2017Data-for-Development-Full-Report.pdf (un.org)

Guzman, L.M, Pennell, M.W, Nikelski, E, Srivastava, D.S., 2019. Successful Integration of Data Science in Undergraduate Biostatistics Courses Using Cognitive Load Theory. *CBE—Life Sciences Education*, **18**(4) DOI: 10.1187/cbe.19-02-0041

Henshaw, A. L., Meinke, S. R. 2018. Data Analysis and Data Visualization as Active Learning in Political Science. *Journal Of Political Science Education.*, **14**(4), 423–439. DOI: [10.1080/15512169.2017.1419875](10.1080/15512169.2017.1419875)

Hodgen, J. 2010. *Is the UK an outlier? An international comparison of upper secondary mathematics education*. London: Nuffield Foundation.

Jacobson, R. 2018*. Opinion: 4 ways to apply data science skills to a career in development*. DEVEX. Available: https://www.devex.com/news/sponsored/opinion-4-ways-to-apply-data-science-skills-to-a-career-in-development-93396

Jütting, J, Avendano, R, Kuhm, M. 2019. Counting the invisible: Three priorities for strengthening statistical capacities in the SDG era. OECD Insights Blog Available: Counting the invisible: Three priorities for strengthening statistical capacities in the SDG era | Development Matters (oecd-development-matters.org)

Keijzer, N, Klingebiel, S. 2017. Realising the Data Revolution for Sustainable Development: Towards Capacity Development 4.0. P*artnership in Statistics for Development in the 21st Century Discussion Paper No. 9.*

Latif, S., Qayyum, Q., Usama, M., Qadir, J., Zwitter, A, Shahzad, M. 2019. Caveat Emptor: The Risks of Using Big Data for Human Development. *IEEE Technology and Society Magazine*. **38**(3).

Li, R. (2019) Teaching Undergraduates R in an Introductory Research Methods Course: A Step-by-Step Approach, *Journal of Political Science Education*, DOI: 10.1080/15512169.2019.1667811

MacInnes, J, Breeze, M, de Haro, M, Kandlik, M, Karels, M. 2016 *Measuring Up: International Case Studies on the Teaching of Quantitative Methods in the Social Sciences*. The British Academy. Available: https://www.thebritishacademy.ac.uk/documents/218/measuring-up.pdf

Mangen, S. 1999. Qualitative Research Methods in Cross-National Settings. *International Journal of Social Research Methodology*, 2(2), pp. 109–124.

Mascaró, M., Sacristán, A.I. and Rufino, M. 2014. Teaching and learning statistics and experimental analysis for environmental science students, through programming activities in R. In Constructionism and Creativity-Proceedings 3rd Intl. Constructionism Conf (pp. 407-416).

Mason, G, Nathan, M, Rosso, A. 2015. *State of the Nation: A Review of evidence on the supply and demand of quantitative skills.* London: The British Academy.

Malta Ministry of Foreign Affairs. 2018. *2018 ODA Statistics.* Available at https://foreignaffairs.gov.mt/en/pages/official%20development%20assistance.aspx

Ministry of Treasury and Finance and the Presidency of Strategy and Budget. 2020. *The role of NGOs on developmen*t, Ankara, Turkey Available: https://sbb.gov.tr/wp-content/uploads/2020/04/KalkinmaSurecindeSivilToplumKuruluslariOzelIhtisasKomisyonuRaporu.pdf

OECD (2020), *Net ODA (indicator).* Available: doi: 10.1787/33346549-en

Oldekop, J.A., Horner, R, Hulme, D, Adhikari, R, Agarwal, B, Alford, M, Bakewell, O, Banks, N, Barrientos, S, Bastia, T, Bebbington, A.J. 2020. COVID-19 and the case for global development. *World Development*, **134**. DOI: 10.1016/j.worlddev.2020.105044

Powell, J, Orton-Vipond, S. 2019. *Understanding How DFID Makes Decisions – Landscape Report on the Role of Data.* Development Gateway. Available: https://developmentgateway.org/blog/understanding-how-dfid-makes-decisions-landscape-report-role-data/

Rocha de Siqueira, I. 2019. Data in the context of intervention and state building. In: Lemay-Hebert (ed.) *Handbook on Intervention and State building*. Cheltenham: Edward Elgar.

Schuff, D, 2018. Data science for all: a university-wide course in data literacy. In Deokar, A., Gupta, A., Iyer, L and Jones, M. (eds) *Analytics and Data Science* , p. 281-297. Springer: Cham.

Ralston, K. 2020. 'Sociologists shouldn't have to study statistics': Epistemology and anxiety of statistics in sociology students. *Sociological Research Online*, **25**(2), pp.219-235. DOI: [10.1177/1360780419888927](10.1177/1360780419888927)

Robertson, D. n.d. *Don't teach students the hard way first* Available: http://varianceexplained.org/r/teach-hard-way/

Tatevossian, A. 2013. *'Big Data' for development: What is it and why you should care*. DEVEX. Available: https://www.devex.com/news/big-data-for-development-what-is-it-and-why-you-should-care-81453

UK Government. 2020. *Changes to the UK's aid budget in the spending review*. Available: https://www.gov.uk/government/news/changes-to-the-uks-aid-budget-in-the-spending-review#:~:text=UK%20aid%20will%20be%20overhauled%20to%20achieve%20greater%20impact.&text=The%20Chancellor%20of%20the%20Exchequer,0.5%25%20of%20our%20national%20income.

UK Government. 2021. *Quantifying the UK Data Skills Gap – full report. Department* for Digital, Culture, Media and Sport. Available: https://www.gov.uk/government/publications/quantifying-the-uk-data-skills-gap/quantifying-the-uk-data-skills-gap-full-report

United Nations. 2013. *A New Global Partnership: Eradicate Poverty And Transform Economies Through Sustainable Development-The Report of the High Level Panel of Eminent Persons on The Post-2015 Development Agenda*. Available: https://www.post2020hlp.org/wp-content/uploads/docs/UN-Report.pdf

United Nations. 2014. *A World That Counts* Available: https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf

United Nations Global Pulse. 2012. *Big Data: Development Opportunities and Challenges White Paper* Available: https://www.unglobalpulse.org/document/big-data-for-development-opportunities-and-challenges-white-paper/

Villarino, M. 2014. *5 things you should know about soft skills for global development*. DEVEX. Available: https://www.devex.com/news/5-things-you-should-know-about-soft-skills-for-global-development-83192

Woolcock, M. 2007. Higher education, policy schools, and development studies: What should masters degree students be taught? *Journal of International Development*, **19**(1), 55–73. https://doi.org/10.1002/jid.1300