EmpoderaData: Sharing a successful work-placement data skills training model within Latin America, to develop capacity to deliver the SDGs

Jackie Carter[a,*], Rafael Alberto Méndez-Romero[b], Pete Jones[a], Vanessa Higgins[a] and Andre Luiz Silva Samartini[c]
[a]*University of Manchester, Manchester, UK*
[b]*School of Engineering, Science and Technology, Universidad del Rosario, Bogota, Colombia*
[c]*Fundação Getulio Vargas, Brazil*

Abstract. *EmpoderaData* – from the Spanish word empoderar 'to empower' – is a partnership research project between the University of Manchester (UK), Fundação Getulio Vargas (Brazil), Universidad del Rosario (Colombia) and Data-Pop Alliance (US and France). The project builds upon a successful data-driven, research-led paid internship programme in the UK (Q-Step) which enables undergraduate social science students to practise data skills through immersion in the workplace. Two-hundred and fifty students have benefited from the Q-Step programme in six years, many graduating into analytical careers in civic society and industry. EmpoderaData aims to build on this experiential learning initiative by developing a data fellowship programme in order to foster and develop data literacy skills in Latin America, led by the need to address society's most pressing issues and using the framework of the Sustainable Development Goals (SDGs).
EmpoderaData Phase 1 explored whether the internship model would have relevance and usefulness within the context of three Latin American case study countries (Brazil, Colombia and Mexico). The team set out to establish a baseline of the state of data literacy and existing training programs in Brazil, Colombia and Mexico. As part of a 'Big Data for the Common Good' event, a workshop was held in São Paulo with thirty participants representing data literacy advocacy or policy formation and drawn from civil society, academia, the private and public sector. The main conclusions from this first phase are: (1) the most requested data literacy training need is for basic skills, including introductory statistics, foundation data analysis and methodological skills; (2) paid data fellowship models are acknowledged as a useful intervention; and (3) the notion of a 'hybrid' professional to build data literacy capacities for 'social science' purposes provides a practical way forward.
In the EmpoderaData Phase 2 project our focus was on Colombia to explore the challenges and opportunities of developing a pilot data fellowship model there. Engaging with national, regional and international capacity development efforts, this highlighted a demand for partnerships between universities and organisations working on the social challenges represented by the SDGs. Partnerships ensure that the in-country data literacy pipeline is strengthened in a home-grown, self-sustaining way, producing a steady flow of data literate graduates into the institutions and sectors where critical data skills are most needed. We report on how the EmpoderaData project is exploring working with students studying Science, Technology, Engineering and Mathematics (STEM) degrees at the Universidad del Rosario, to improve the application of statistical methods to the social sciences. The aim is to strengthen STEM skills and develop youth empowerment across Colombia, urban and rural areas, to improve the quality of statistical education at the national level, and support the skills needed to deliver the SDGs.
In parallel, the Fundação Getulio Vargas (FGV) Business School in São Paulo agreed to trial the work-placement programme in their undergraduate business and public policy degrees through a programme entitled 'The FGV Q-Step Center to improve

---

*Corresponding author: Jackie Carter, University of Manchester, UK. E-mail: jackie.carter@manchester.ac.uk.

quantitative skills in undergraduate business students'. This two-year-long funded study will enable us to explore the transferability of the internship model from the UK to Brazil. The paper will discuss how the programme was established (following the lessons learned from EmpoderaData), explain how this model will be implemented in FGV, especially paying attention to how the curriculum will develop to support it, and how the impact of the programme will be monitored. The knowledge exchange generated from this study will complement the research conducted through the EmpoderaData project.

The paper will cover the progress of the EmpoderaData project and FGV-Q-Step Center to date and explore how we are developing these initiatives, the challenges we have faced, and how through partnership working we are developing capacity building in statistical and data skills training.

1. Introduction

This paper introduces and outlines the EmpoderaData project, a transnational collaboration to develop data fellowship programmes within Latin America. These data fellowships are conceived as a tool for building statistical capacity to help deliver the UN's Sustainable Development Goals (SDGs) [1] by strengthening the quantitative skills pipelines in countries. EmpoderaData builds on the success of The University of Manchester Q-Step Centre's innovative programme (part of a larger UK-based initiative) which develops the quantitative skills of undergraduate social science students through both classroom-based instruction and workplace learning through a summer paid internship programme. As we describe below, this model has been successful on a number of fronts, creating partnerships and graduate employment opportunities between the university and many organisations with statistical research skills needs, such as polling organisations and local and national government. Through EmpoderaData, we are working with partners within Latin America to take practical steps towards developing a version of this model which will help achieve similar results, with a focus on the SDGs. The SDGs require an increase in levels of statistical capacity across the board [2], including at the level of individual statistical and data literacies, and innovative solutions are needed to help deliver this. We report in this paper on two different approaches that are in development by Universidad del Rosario in Bogotá, Colombia and FGV in São Paolo, Brazil, outlining their strategy, progress and emerging challenges.

The structure of the paper is as follows. In Section 2, we provide some background to the project, contextualising and explaining in more detail the Q-Step programme, its experiential learning-based approach to quantitative skills development, and how the present EmpoderaData project came about in relation to this. In Section 3 we discuss the role of the SDGs in this framework, outlining the unprecedented statistical capacity needs they create and why we believe a holistic notion of data literacy is necessary to address them. In Section 4, we focus on the two practical implementations currently in progress in Colombia and Brazil, each of which adopts a different strategy. The Brazilian intervention focuses on an immersion-based programme similar to that of the Q-Step model, while the Colombian initiative focuses on bringing real practical social research projects into the university classroom. Finally, we conclude by reflecting on the challenges and opportunities that have emerged through this work to date and outline the next steps for the project.

2. The background to the EmpoderaData project

2.1. The Q-Step Centre at University of Manchester

The EmpoderaData project was conceived in 2018, following the development and successful roll out of a paid internship programme at The University of Manchester UK, which began in 2014 and enabled us to build an annual cohort of 'data fellows'. Through a UK nationally funded programme strategically focused on developing data and statistical literacy skills (we expand on these definitions below) in the undergraduate social science population, the University of Manchester established a Q-Step Centre [3], one of 15 in the UK. The aim of the Q-Step programme was to create a step-change in teaching quantitative research methods to social science undergraduates. Q-Step centres were encouraged to be innovative and experimental in designing resources and educational and practical experiences, to add value to the research methods curriculum and develop a cross-national cohort of graduates who could enter the labour market equipped to enter 21st century careers that require statistical competencies. With the wealth of experience in the University of Manchester Q-Step Centre team in data-driven teaching [4–6] our approach was two-fold. First, we created course modules that ensured

numbers and statistics were a normal part of the social science curriculum, from the point at which students enter their degree courses. We also introduced specialist 'with Quantitative Methods' pathways through degrees for students who wished to specialise in data analysis and statistics. This approach involved employing new staff and developing substantively-led teaching materials and activities that reflected the subject of the degree, in our case this covered sociology, criminology, politics and international relations, English language and linguistics, philosophy and economics. In other words, the statistics and data analysis teaching was *embedded* in the subject teaching [7]. Second, the team developed a paid internship programme, working with external organisations across the public, private and voluntary sectors to co-create data-driven, two-month long research projects to enable students to take their learning from the classroom and PC lab and put this into practice in workplace environments, with research that matters to those host organisations. The interns' immersion into applied social research projects provides the opportunity to develop their analytical and research skills, as well as their professional skills, alongside colleagues (sometimes themselves in their early career stages). Collectively these developments enable undergraduates to learn statistics, using real-world data from official sources (mainly) and hone their skills in a professional environment. Carter [8] expands in her book on the approach taken, describes how this is underpinned by experiential learning theory, and illustrates through case studies of former interns, and vignettes of current social researchers, the benefits of 'learning by doing'. The initiative has produced 250 data fellows in just six years, with a further 62 taking place in 2021.

The University of Manchester's strong track record for teaching quantitative research methods to undergraduate students on social science and humanities degree courses, is accompanied by a strategic focus on the Sustainable Development Goals (SDGs), the University's third core goal being Social Responsibility. The 2019 University of Manchester Social Responsibility report made explicit reference to the Q-Step internships [9], noting that: 'the Q-step paid internship scheme has placed 200 students in 60 public, private and third-sector organisations to undertake social research that makes a difference and helps to identify and progress social issues locally, nationally and globally. The projects have addressed gender data gaps in developing countries, food poverty, recycling, immigration, sociodemographic factors affecting university admission rates, violence against females, bilateral spend-

ing on HIV/AIDS, and modelling UK deprivation. In 2019, the scheme will be extended to three Latin American countries to develop a data programme around the global SDGs'. In 2021, just two years after this report, the University of Manchester was ranked first in the world for its impact as measured against the SDGs [10]. This paper draws on the combined strengths of the data fellows training programme (delivered through the Q-Step internships) and the commitment to delivering on the SDGs, demonstrated by the university.

By way of illustration, as noted in the extract from the University's SDGs report, the interns – or data fellows as we go on to call them throughout this paper – undertook projects that provided them with real world opportunities to develop their data and statistical literacy skills. Here we draw from two of the project outputs (all students are required to produce a poster to evidence their learning), chosen to reflect the research projects' focus on the SDGs. Both examples were carried out in consultation with Open Data Watch (an organisation that works at the intersection of open data and official statistics, see https://opendatawatch.com/) and Data2X (a UN-led initiative working to improve the availability, quality and use of gender data, see https://data2x.org/), with the interns hosted at the first of these organisations in Washington DC.

Niamh, a second year politics and international relations student, worked for the summer on a project called Bridging the Data Gap assessing the availability of data about the lives of women for selected SDG participating countries. Her role was to find, record and assess gender-relevant development indicator data. Figure 1 was produced for the poster (required on completion of the project) revealing that the human development theme (domain) that had the highest percentage of sex-disaggregated indicators for all countries included in the study was Education, but that even for this theme (noting SDG4 is Quality Education) a quarter of the indicators included had no sex-disaggregated data.

Niamh's project, conducted in 2018, followed a project undertaken by another student, Grace, the previous year which also explored data gaps for a range of SDG indicators. The findings of that project[1] were:

> 'Within the 15 countries [included], 48% of all gender indicators were missing. The health domain with the greatest number of indicators (32) only has 58% of indicators available with sex disaggregation. This is the most complete coverage of all categories.'

---

[1] See summer intern's project poster at http://hummedia.manchester.ac.uk/faculty/qstep/student-stories-2018/desouza.pdf.
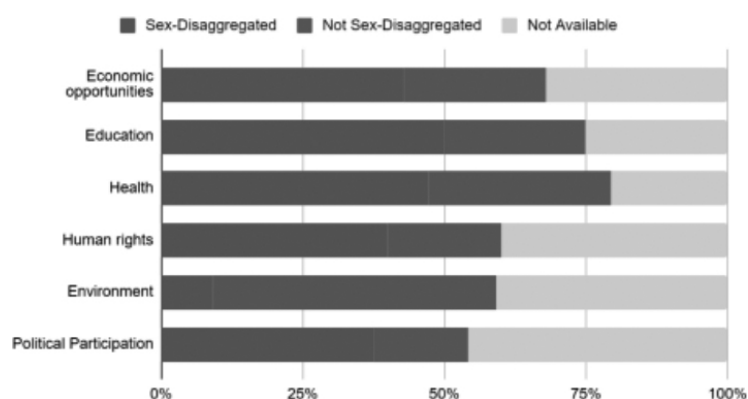
Fig. 1. Sex-disaggregation of indicators by domain. Source: Q-Step intern's project poster at [11].

and in her final poster she reflected that:

> 'I have learnt many skills ... I was taught data literacy, in order to find and reference the data needed for the project. From looking through hundreds of data sources, the importance of appropriate data presentation has been emphasised, and I will use these tools in my university research in the future. I also garnered softer skills of working in a professional environment, an invaluable experience. For each of the 15 countries five aspects of the 105 indicators were recorded: availability, source, metadata, consistency of data with indicators, and degree of sex disaggregation.'

Both students have contributed knowledge and empirical findings to the host organisations, and Niamh's project built directly on Grace's. Grace's finding appears in [12] 'In total, sex-disaggregated data for gender-relevant indicators are unavailable in any year for 48 percent of the possible observations in international or national databases' [11, p. 7]. Although just two examples of students' work, they contributed to the catalyst for the EmpoderaData project. Moreover, they gave the interns an opportunity to learn through doing, and resulted in an improved understanding of the data associated with measuring the SDGs.

2.2. The EmpoderaData project

In 2018 an application to develop a pilot project in three countries in Latin America – Colombia, Brazil and Mexico – building on an existing international network with colleagues overseas working in the area of data skills training for Sustainable Development – was developed. This built on the data and statistical capacity building expertise of two of the authors of this paper,

who co-led a Data Skills and Training Research Group at the University of Manchester. The project was entitled 'Developing data and statistical literacy capacity to achieve the SDGs: a pilot project in three Latin American countries'. The pump-priming funding (from the UK Government's Global Challenges Research Fund, GCRF) supported *collaborative research activities with a view to preparing academic colleagues to respond to future opportunities*. Critically, this project was developed to work alongside in-country activities in Latin America that were already underway in support of achieving the Sustainable Development Goals (SDGs) by 2030.

The project – which became known as EmpoderaData Phase 1 – built on earlier work that had developed a research relationship between the University of Manchester and DataPop Alliance (a global coalition on big data and development created by the Harvard Humanitarian Initiative, MIT Media Lab, and Overseas Development Institute that brings together researchers, experts, practitioners, and activists to promote a people-centred big data revolution through collaborative research, capacity building, and community engagement). In 2018 Carter and Higgins, as invited and paid-for fellows, presented their work and project ideas at the MIT-hosted DPA-led event on *Leveraging Big Data and Sustainable Development*. The purpose of the event was to 'strengthen the skills of UN staff and development practitioners in selecting, creating, using and interpreting data in support of 2030 Agenda for Sustainable Development'. Twenty-seven countries, all focused on developing capacity to deliver the SDGs, were represented at the event.

Immediately following the MIT event a strategy meeting was held where Carter, Higgins and members of the DPA team explored next steps with a view to

extending existing research and training. This resulting bid application – which became the EmpoderaData project – provided an opportunity to build on an established network, meet face-to-face in Latin America and Manchester and to begin to develop a large grant application to further the work of the group.

The team's primary aim was built on a belief that education and partnerships provide the cornerstone of achievable international development, and are required to support the delivery of the Sustainable Development Goals (SDGs). Our collective intention was to develop insight into the transferability of a UK-led initiative (the Q-Step Centre described above), by exploring the feasibility of the model in three pilot Latin American locations. We had two primary aims: first to work in-countries to explore the data and statistical literacy capacity to help achieve the SDGs; and second to explore the extent to which the University of Manchester Q-Step paid internship programme might be useful working with in-country partners to help achieve the first aim, through a data fellowship programme.

The first phase of EmpoderaData (further described in Section 4) was designed to enable the team to: (a) baseline the data and statistical capacity of those selected countries for delivering the SDGs (working in partnership with their statistical offices, and cross-checking with the UNDP and others) and (b) explore their attitudes and willingness to having an intervention SDGs data fellowship programme developed with their relevant stakeholders, building on the success of the University of Manchester's Q-Step work placement programme. This fed directly into the second phase of the project, which involved working closely with Data-Pop Alliance colleagues in Colombia and the Universidad del Rosario in Bogotá to tailor a solution that would be able to achieve an impact in that context. We report on key takeaways from this process in Section 5.

3. The SDGs and the need for data literacy

3.1. The SDGs as a framework

The UN Sustainable Development Goals set an ambitious global set of targets to reach by 2020. These 17 goals, which comprise 169 associated targets and 231 indicators with which to measure social and economic progress, are the central framework for the 2030 Agenda for Sustainable Development [1]. There is a global need at the level of statistical capacity to deliver these goals [1] accompanied by scepticism of whether they are indeed achievable [13–17]. Nonetheless these goals set a broad backdrop to the complex global challenges we face. We argue in this paper that in order for countries to be in a position to deliver on these goals, a fundamental starting place is to have well-trained citizens who are data literate.

The SDGs create a global need to deliver on an agenda which relies heavily on quantitative data. This creates a demand for human capital with the requisite statistical and data skills to work on the challenges represented by the SDGs. The country-led model we report on here is designed to address this need, rooted in the belief that "capacity building is most effective when it is home-grown, long-term in perspective and managed collectively by those who stand to benefit" [18, p. 897]. Through a focus on strengthening the data literacy pipeline in countries through a programme combining statistical education and workplace learning, this model aims to provide a steady home-grown flow of the skills and competencies needed to tackle the SDGs into the sectors and institutions where they are most sought after.

The SDGs present enormous measurement challenges, as has been well documented in recent years [14, 19]. Not only do they require an unprecedented amount and breadth of data for populating the many indicators, they also require an increase in human capacity across the board to ensure that there are enough trained individuals to work on these data and the complex measurement problems they involve [20]. As such, there is increasing recognition that data alone will not be sufficient for meeting the challenges of the SDGs, and that countries require investment in data literacy and skills in order to sustainably improve the capacity for effective use and understanding of data [21,22]. While the benefits of investing in data production, infrastructure and technology are relatively well-known and understood, there is still comparatively little emphasis on the need to invest in the individual capabilities necessary to make effective use of these data and technologies. In a recent survey of national statistical offices (NSOs) around the world on statistical capacity assessment, the Partnership in Statistics for Development in the 21st Century (PARIS21) found that "while only 2% of assessed capabilities target the individual, 32% of countries expressed that individual capabilities need to be improved to rise to new data ecosystem challenges" [23, p. 16]. As such, care must be taken to ensure that the short-term desire to populate the SDG indicators does not divert resources away from the longer-term development of the overall system capacity [18] or promote "the produc-

tion of data and information at the expense of statistical capacity" [23, p. 10]. Improving data skills and literacy in the population is a key element in this long-term human-centred approach to capacity building.

This more skills-oriented perspective on statistical capacity development is a more central focus of the Capacity Development 4.0 framework developed by PARIS21. This framework acknowledges that "too little investment in people and skills" [21, p. 19] has slowed progress towards the so-called data revolution for development [20], and thus aims to "go beyond the traditional production-side interventions to also include the strengthening of data use, literacy and results" [21]. In line with this recognition, we want to emphasise that an important aspect of ensuring home-grown and long-term capacity development requires connecting the big picture measurement challenges of the SDGs with the earlier stages of the statistical literacy pipeline. By focusing on interventions at the level of statistical education, and by centring partnerships between statistical stakeholders and the in-country educational system, we can ensure that there is a strong and sustainable supply of human capital with the necessary skills and competencies for working on the complex social challenges represented by the SDGs.

3.2. What is data literacy?

The EmpoderaData project focuses on creating data literate graduates ready to work on SDG problems. As such, it is worth unpacking what we mean by data literacy and which skills and competencies this encompasses. This task is made trickier by the various ways in which the term is deployed, and how these overlap with other related terms such as numerical literacy, statistical literacy and information literacy [25]. We therefore aim to clarify here our understanding of data literacy and why this term best reflects the skillset EmpoderaData aims to develop.

The technical, analytical and conceptual aspects of quantitative analysis have traditionally been considered under the term *statistical literacy*, broadly understood as a combination of "the ability to produce, analyse and summarise detailed statistics in surveys and studies" and "the ability to read and interpret summary statistics in the everyday media: in graphs, tables, statements and essays" [26, p. 135]. However, the demands of the current digital age and the rise of more holistic data science approaches to statistical work have led to an increased emphasis on a broader set of competencies for working with data which includes the "ability to

collect, manage, evaluate, and apply data, in a critical manner" [27, p. 2]. As the amount of data permeating our day-to-day lives has increased and the boundaries between producers and users of data have been eroded, critical data skills are now often seen as necessary parts of the statistical curriculum, with the term 'data literacy' increasingly being used to refer to this inclusive set of skills [28]. The current set of data challenges we face, including the daunting quantitative requirements of the SDGs, require a holistic skill set which includes and goes beyond understanding the basic formulas, principles and analytical aspects of statistical reasoning. In order to both leverage the power and understand the limits of the varied data sources which now permeate our world, a kind of literacy is needed which includes "a critical appreciation of data provenance and quality" [29, p. 12], "understanding issues of data privacy and ownership" [28, p. 3], and "understanding how data are stored" (ibid.). This is reflected in the UK Government Data Task Force's recent emphasis "[t]o make the best use of data, we must have a wealth of data skills to draw on. That means delivering the right skills through our education system, but also ensuring that people can continue to develop the data skills they need throughout their lives [30]."

These data skills are in addition to, rather than instead of conventional statistical literacy skills. We therefore follow Data-Pop Alliance [31] in understanding data literacy as "the desire and ability to constructively engage in society through or about data", a broad definition which "interacts with and builds on [other related literacies such as statistical and information literacies] and requires a combination of the technical, critical, quantitative and conceptual skills on which they are based". The conceptual, mechanical and critical aspects of doing statistics with real data all remain important elements of the skillset required to work on understanding and addressing social challenges such as those represented by the SDGs [28,32,33]. However, the inclusive notion of data literacy we have outlined here is important for developing the next generation of social statisticians to work on practical official statistics challenges as, unlike textbook data, "real data about society are often more complex and messy" [32, p. 46]. If learners are to use their statistical skills to work on SDG problems, they need to be prepared to work with such real-world data and develop the various data literacy requirements that this entails. This is particularly important for ensuring that data sources which fall outside the traditional sphere of official statistics are adequately understood when these are utilised towards the SDGs.

Likewise, the ability to draw insights from such data also requires conventional statistical training: the social problems represented by the SDGs are highly complex, and understanding them therefore "requires the ability to explore, understand, and reason about complex multivariate data, because social phenomena do not happen in a vacuum, and their understanding requires awareness of how variables co-vary, or affect each other, or are situated in a network of causal factors that may change over time in manifold ways" [32, p. 45].

As the EmpoderaData model is premised on the application of statistics in the workplace, we regard the need to have the full set of data and statistical literacies to provide an inclusive set of skills which will have most benefit in strengthening the skills pipeline nationally. As discussed above, the value of the Q-Step paid internship model is that it exposes students not only to classroom-based instruction in the basic statistical principles and analytical techniques which underpin statistical literacy, but also an appreciation for critical data skills through hands-on experience working with real data. A key challenge for statistics education is to ensure that students develop more than just the ability to recall and apply formulae and calculations from textbooks, but also a higher-level understanding of the principles of making sense of data. The experiential approach represented by EmpoderaData can help to bridge the gap between the formulae learned and the conceptual understandings underpinning them by putting statistics into the context of front-line social research [34]. In the next section, we describe how EmpoderaData has set about identifying the particular skills and data needs of its partner countries in relation to the SDGs, and how this feeds into the development of its data fellowship model.

4. EmpoderaData pilot research

EmpoderaData Phase 1 aimed to establish an understanding of the following within Brazil, Colombia and Mexico:

- unmet needs in terms of data literacy skills;
- the extent to which a data literacy fellowship model might help to develop these skills;
- the extent to which the SDGs are monitored, and the SDG goal that might be relevant to work on within each country's context.

These three countries were chosen as pilots because of their strong official statistical and academic systems that could serve as a backbone for the "data revolution",

their wide penetration of mobile and internet technology, their established open data movement and their active and vibrant civil societies.

The research methodology was qualitative and was undertaken in three stages:

1. May 2019: a workshop was held in São Paulo as part of a 'Big Data for the Common Good' event. Thirty participants, who were involved in data literacy advocacy or policy-making, attended the workshop. The participants represented a range of sectors including civil society, academia, private and public sector. The workshop aimed to (1) map the needs for training, (2) identify partners for future fellowship programs, and (3) present the current University of Manchester Q-Step internship model and explore its applicability in Brazil, Colombia and Mexico.

2. June-July 2019: Eighteen qualitative interviews were undertaken with stakeholders within Brazil, Colombia and Mexico. A semi-structured interview guide was used and interviews were conducted online (Zoom) in either Portuguese, Spanish or English. Fifteen interviews focussed on the data literacy needs and Q-Step internship model (6 in Brazil, 5 in Mexico and 4 in Colombia). The sample of interviewees was selected to represent a diversity of sectors (academia (private and public), civil society and public sector), and as much as possible provide a gender and regional (within countries) representation. The interviews were structured around (1) the availability of, and need for, training in traditional quantitative skills such as such as survey, census or official aggregate data (2) the availability of, and need for, training in data sciences/artificial intelligence, focussing on big data/new sources of data and artificial intelligence analytics, i.e. programming, machine learning, etc. (3) the interest in the adoption of the Q-Step internship model. The three countries are structured around pervasive inequalities in access to education, employment, healthcare etc. Therefore, the interviews included questions on inequalities in the need for data literacy training among sub-groups of the population (for example, inequalities related to income, and gender).

The remaining three interviews focussed on understanding the progress of Brazil, Colombia and Mexico in measuring progress towards the SDGs. One interview was conducted with a relevant professional working directly with the SDGs for the government of each country. Interviewees were

asked about (1) their impression of the national monitoring and evaluation system re: SDGs in the context of their country (2) which goal they think it is relevant to work in the context of their country (3) key resources/sources for monitoring and evaluation of the SDGs in their country.

3. October 2019: a workshop was held at the University of Manchester in October 2019, to present and discuss the preliminary findings with 31 invited stakeholders and potential partners.

The full results from the research are reported in [35] through the EmpoderaData project website[2] and the main conclusions follow:

Data Literacy Training: There is a clear need for more data literacy training across academia (particularly undergraduates), the public sector and civil society within all three countries. The biggest training needed is for basic skills such as introductory statistics, foundational data analysis, basic methodological skills and also basic data science skills. This was highlighted in all three countries. Interviewees and attendees at the workshops expressed that the end-goal is to foster critical analyses of data, rather than the development of pure mathematical competencies, in order to create leadership that can think critically about data.

Hybrid Fellowship Models: In all three countries, paid data fellowship models were acknowledged as a useful intervention. The participants in the research recommended that the target audience for fellowships should be students and potentially young professionals. However, different adaptations of the Q-Step model were suggested for each country; it is very important that the context of the country is considered when developing a fellowship model. Another recommendation was that a training curriculum should be based on non-proprietary software to ensure sustainability and facilitate access.

A key recommendation from the research was the need to foster 'hybrid' professionals that can understand, use and analyse data for social science purposes, i.e. evidence-informed policy, journalism, activism etc. A hybrid model would bring together people with complementary backgrounds (data science and social science) to work together at the host organization, on one specific sustainable development challenge. It was felt that this would lead to a truly critical analysis of data and would have a higher impact in terms of measuring achievement of the sustainable development goals.

_____
[2]https://datapopalliance.org/empoderadata-project/.

Since this research was undertaken, the field of computational social science has rapidly emerged and this may be an alternative to the hybrid model, with computational social science lying at the intersection of the social sciences, statistics and computer science [36,37].

Sustainable Development Goals: The research explored which SDG(s) might be used within each country to base the data literacy training around. The outcome was that training should not be restricted to a particular SDG within a country; the SDG content of the training should remain flexible to incorporate the host organizations' interests, as well as sectoral funding opportunities within each country. However, Colombia and Brazil interviewees suggested offering fellowships to students from poor backgrounds who were in need of a stable source of income. A specific outreach strategy to involve typically excluded subgroups may, therefore, generate an impact in terms of SDGs 10 (reduced inequalities) and 4 (quality education). However, it was noted that, paradoxically, the individuals would also need to have an acceptable educational background at entry level. Therefore, there is a need to find the balance to avoid the risk that a fellowship program would end up working with those with more privileged backgrounds.

In summary, the main conclusions from EmpoderaData Phase 1 were: (1) the most requested data literacy training need is for basic skills (2) paid data fellowship models are acknowledged as a useful intervention; and (3) the notion of a 'hybrid' professional to build data literacy capacities for 'social science' purposes provides a practical way forward. The full results from Phase 1 [35] were used to inform and design EmpoderaData's second phase – the implementation of the Q-Step model in Colombia and Brazil.

5. Implementing the data fellowship model in Colombia and Brazil

Building on what was learned in the EmpoderaData pilot research, the next steps for the project were to build the necessary partnerships to develop and deploy a data fellowship programme in the region. Two parallel projects emerged from the São Paulo workshop to this end; the first led by Universidad del Rosario in Bogotá, Colombia, and the second at FGV Business School in São Paolo, Brazil. In order to ensure that the data fellowship model serves the needs of the particular country in which it is implemented, the process must be led by partners within those countries. Local partners are uniquely positioned to understand both the needs of their societies and the challenges and opportunities for

achieving the most transformative impact in the local statistics education pipeline. In this section, we report on how these two parallel initiatives will work and how we will learn from them in designing future regional interventions.

5.1. Data fellowships in Colombia

Here we describe the actions in Colombia that aim at strengthening statistical capacities, the mobilization of knowledge (quantitative) and the impact of knowledge transfer in the communities. The Universidad del Rosario, a 367-year-old Colombian higher education institution, materializes its academic commitment to the common good through high-level educational offerings and impact in the country. The first college mathematics course was held at this university several centuries ago by José Celestino Mutis, who also led the first botanical survey of the country. That commitment to science continues today not only through academic programs in the context of mathematics and science but also through a dynamic curriculum, in courses of study such as the social sciences. That includes academic products for the strengthening of quantitative competencies for the resolution of complex problems (which requires interdisciplinary and diverse reading), such as hackathons, seminars, and collaborative projects.

This action that we refer to above does not work solely with numerical expertise. It is undoubtedly necessary to possess a social sensitivity that allows, among other things, engaging with the communities for the detection of social problems that can potentially be solved by whoever is learning.

This idea materializes on two fronts. The first is offering interesting courses that allow students to generate discussions mobilising what is typical in their disciplines but with a discourse permeated by mathematics, approaching what is typical of critical mathematics [38]. The second is a type of living laboratories, called the mINNga Labs [39], due to their closeness to the spirit of the indigenous *mingas*, as certain Andean communities called collective agricultural work carried out for the benefit of the tribe. "We understand living laboratories to be an innovation model where all the actors actively participate appropriating innovation tasks, clearly open and collaborative, in scenarios of co-creation and validation of solutions that they need themselves, in real life contexts, using, to a large extent, ICT as a medium, thus forming a research and innovation ecosystem that permanently and explicitly enables social innovation" [39, p. 255].

These laboratories are governed by five fundamental principles:

1. Continuity: Building trust and exploring an approach that leads to efficient, long-term sustainable innovations, involving users rather than restricting them [40].
2. Openness: Allowing the presence of multiple perspectives of thought, in user-driven research scenarios.
3. Realism: Providing solutions to real problems in the daily lives of those concerned.
4. Empowerment of users: This takes into account (and builds from) human needs and desires. The social dynamics of the living laboratories approach guarantees the rapid propagation of innovative solutions (viral adaptation) through social-emotional intelligence mechanisms.
5. Spontaneity: Allowing flexible and robust models.

5.2. The FGV Q-Step initiative

After the 2019 EmpoderaData workshop in São Paulo, a project to implement the Q-Step model at FGV Business School was submitted to a Brazilian funding agency (FAPESP). In collaboration with The University of Manchester, the two-year project, beginning in March 2021, is described below.

The Q-Step Center at FGV aims to develop and improve the quantitative skills of undergraduate Social Science students and increase general interest in quantitative research among students. We expect the program to encourage students to work in quantitative careers after graduating and to apply to undertake quantitative studies in postgraduate courses.

The program consists of basic disciplines, already in the curriculum, internships and advanced disciplines. The internships and the advanced disciplines will be offered in the following fields: public administration, finance, marketing, accounting and data science. Other fields may be offered, depending on student demand. Students will be selected from two undergraduate programmes at FGV-EAESP: Business Administration and Public Administration.

Students from the second year can apply for the Q-Step program and they must be approved in basic quantitative disciplines in the first year (Mathematics 1 and 2, Statistics 1, Information Technology and Programming Logic). Half of the students who apply for the program will be randomly selected to participate in the program. These students will participate in internships offered by the university in partnership with companies. As we select the students randomly, among those interested, we are able to compare those who participate in

the Q-Step program with those that do not and in terms of quantitative skills, and, thus, assess the efficacy of the program.

As part of the curriculum, FGV undergraduate business students must participate in two "immersions" or internship activities. These week-long "immersions" take place twice a year (April and October). The students have no classes during this period, but spend one week at a company developing a project or idea. As a result, they present a proposal to the company team and then are evaluated. These "immersions" range from trips to "favelas" to developing ideas for e-commerce companies. For Q-step participants, new immersions that require and develop quantitative skills will be offered and students can choose among these to complete Q-Step credit requirements.

Besides immersions, students can take summer internships, lasting from 3 to 4 weeks, to be developed by the Q-Step coordinator at FGV, in partnership with companies, and in consultation with the Q-Step internships lead from the University of Manchester.

Besides the basic disciplines (Statistics, Mathematics, Introduction to IT), FGV offers more than 60 elective disciplines, some of which are quantitative. In addition, as part of the Q-Step program, FGV will elaborate advanced quantitative-driven disciplines specifically to the Q-Step program. The Q-Step participant will have to complete at least one of these advanced disciplines.

The learning outcomes of the FGV Q-Step program are:

- Improve quantitative skills of the students; this will be measured by summative tests applied at the end of the program. We expect that the selected students of the Q-step program will score higher on these tests than the ones that were not selected;
- Increase the use of quantitative approaches and methods in students' works and research; it will be measured by a survey of both selected and not selected students;
- Increase the interest in using quantitative methods after graduating (in post-graduate courses or workplace); this will be measured by a survey of both selected and not selected students.

Some of these outcomes are aligned with the ones of the undergraduate program and are described in the pedagogic project. FGV is one of the few Latin American schools that participates in AACSB (Associate of Advanced Collegiate School of Business) and must evaluate whether the students are achieving the learning goals through the "Assurance of Learning" system. At different points of the course, students are evaluated in respect to the learning objectives of the course.

5.3. Collaborative work to materialize the alliance with Latin America

Having said all the above, we wonder how to continue materializing strategies that allow strengthening inter-institutional alliance and thus generating real and achievable actions within the framework of learning quantitative skills for solving social problems, through learning-by-doing in the business sector, but always aligned with the SDGs.

These initiatives should nurture the development of global citizen competencies in our students and guide strategies in the Internationalization of the Curriculum. Through these competencies, the aim is for students to be able, for example, to interact and communicate effectively in another language with people from other cultures and countries, to think globally and consider issues from different cultural perspectives and to lead work teams in culturally diverse environments, but, particularly, focusing on the mobilization of quantitative knowledge for the solution of real problems.

An optimal outcome is the Collaborative Online International Learning (COIL), as a privileged pedagogical setting where internationalization goals are actively developed in the teaching-learning process. This strategy is aligned with the learning to learn, which articulates the development of disciplinary contents with international virtual educational environments to enrich the training of students through collaborative and meaningful learning with students from different cultural and linguistic contexts. More specifically, we propose a COIL-type subject, which links students with different cultural and geographical experiences, sensitising them to the global world and deepening their understanding of themselves, their culture, how they are perceived by others and how they perceive the "others". A good COIL experience design engages students in learning the subject matter through their own cultural perspective and by exchanging their cultural and experiential perspective as they move through learning with international students. This scenario will give many students the opportunity to have an international experience and to develop linguistic and digital competences that are very valuable in a global and interconnected world.

The subject that we propose to work collaboratively in takes into account the germinal aspiration of responding to the SDGs. In particular, we propose, through a shared subject (with the use of COIL methodologies), to develop statistical competencies in the exercise of problem solving within the framework of SDGs. This subject will be based on experiential learning, with the

particularity that the problems that will be solved in class will correspond to real problems of the communities, using living laboratories. In this way, it will not only allow the construction of mathematical knowledge, but it will be developed from social justice and critical thinking.

6. Conclusions

The paper draws on the combined strengths of the data fellows training programme (delivered through the Q-Step internships) and the commitment to delivering on the SDGs, demonstrated by the University of Manchester in the UK. We have explored whether the Q-Step internship model has relevance and usefulness within the context of Brazil, Colombia and Mexico. We have outlined the strategy, progress and emerging challenge of two different approaches that have been developed via partnerships between the University of Manchester and the Universidad del Rosario in Bogotá, Colombia and FGV in São Paolo, Brazil.

There is no doubt that, within Brazil and Colombia, the data fellowship model is perceived as a tool for building statistical capacity to help deliver the UN's SDGs. The results from the EmpoderaData project gives a very clear narrative that a data fellowship model can be flexibly adapted to different disciplines/subjects (traditional social science, business studies and mathematics), within different country contexts and with different curriculum design. For example, the Brazilian intervention focuses on an immersion-based programme similar to that of the Q-Step model, while the Colombian initiative focuses on bringing real practical social research projects into the university classroom. The project also gives support to MacFeely and Barnat's argument of capacity building being "most effective when it is homegrown, long-term in perspective and managed collectively by those who stand to benefit" [18, p. 897].

We have identified some key challenges to building a successful data fellowship model within Brazil and Colombia. The first is that there are SDG data system/infrastructure deficits that restrict the capacity of data fellows to work with real data to measure the SDGs. The second challenge is the pervasive inequality in education among subgroups of the population. To overcome these inequalities, efforts must be made to target the recruitment of marginalised sub-groups (such as women, those with lower incomes and those in rural communities) into a data fellowship programme [41].

The early data fellowship initiatives within the Universidad del Rosario and FGV are a positive step forward in striving to build statistical capacity to help deliver the UN's SDGs. As these initiatives develop further, they will be monitored and evaluated to identify the successes and the improvements needed to strengthen the data fellowship model and work further towards the development of quantitative skills pipelines in Colombia and Brazil.

Acknowledgments

References

[1] United Nations. Transforming our world: The 2030 agenda for Sustainable development [Internet]. Sustainable Development GOALS Knowledge Plataform; 2015 [cited 2021 May 20] p. 41. Report No.: A/RES/70/1. Available from: https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf.

[2] Pfeiffer A, Middeke F, Tambour M. 2030 agenda for sustainable development: implications for official statistics. Stat J IAOS [Internet]. 2017 [cited 2021 May 20]; 33(4): 911–8. Available from: doi: 10.3233/SJI-170360.

[3] Nuffield Foundation, Economic and Social Research Council (ESRC), Higher Education Funding Council for England (HEFCE). Aims and activities of the Q-Step centres [Internet]. 2014 [cited 2021 May 20]. Available from: https://www.nuffieldfoundation.org/sites/default/files/files/Aims%20and%20Activities%20of%20the%20Q-Step%20Centres(1).pdf.

[4] Carter J, Noble S, Russell A, Swanson E. Developing statistical literacy using real-world data: investigating socioeconomic secondary data resources used in research and teaching. Int J Res Method Educ [Internet]. 2011 [cited 2021 May 20]; 34(3): 223–40. Available from: doi: 10.1080/1743727X.2011.609553.

[5] Wathan J, Brown M, Williamson L. Increasing Secondary Analysis in Undergraduate Dissertations: A Pilot Project. In: Teaching Quantitative Methods: Getting the Basics Right [Internet]. London: SAGE Publications Ltd; 2011 [cited 2021 May 20]. pp. 121–41. Available from: doi: 10.4135/9781446268384.

[6] The University of Manchester. Enriching Social Science Teaching with Empirical Data (ESSTED) [Internet]. ESSTED. 2013 [cited 2021 May 20]. Available from: https://sites.manchester.ac.uk/essted/.

[7] Buckley J, Brown M, Thomson S, Olsen W, Carter J. Embedding quantitative skills into the social science curriculum: case

studies from Manchester. Int J Soc Res Methodol [Internet]. 2015 Sep 3 [cited 2021 May 20]; 18(5): 495–510. Available from: doi: 10.1080/13645579.2015.1062624.

[8] Carter J. Work Placements, Internships & Applied Research [Internet]. 1st ed. United Kingdom: SAGE Publications Ltd; 2021 [cited 2021 May 20]. 304. Available from: https://uk.sagepub.com/en-gb/eur/work-placements-internships-applied-social-research/book253221.

[9] The University of Manchester. The University of Manchester and the Sustainable Development Goals [Internet]. 2019 [cited 2021 May 20]. Available from: https://documents.manchester.ac.uk/display.aspx?DocID=43121.

[10] THE World University Ranking. Impact Ranking 2021 [Internet]. Times Higher Education (THE). 2021 [cited 2021 May 20]. Available from: https://www.timeshighereducation.com/impactrankings.

[11] Evans N. Open Data Watch: Bridging the Gap [Internet]. [cited 2021 May 20]. Available from: https://documents.manchester.ac.uk/display.aspx?DocID=46665.

[12] Open Data Watch, Data2X. Bridging the Gap: Mapping Gender Data Availability in Africa. [Internet]. 2019 [cited 2021 May 20] p. 49. Available from: https://data2x.org/wp-content/uploads/2019/06/Bridging-the-Gap-Technical-Report-Web-Ready.pdf.

[13] Gennari P, Kalamvrezos Navarro D. Are We Serious About Achieving the SDGs? A Statistician's Perspective. ISSD: SDG Knowledge Hub [Internet]. 2020 [cited 2021 May 20]; Available from: https://sdg.iisd.org:443/commentary/guest-articles/are-we-serious-about-achieving-the-sdgs-a-statisticians-perspective/.

[14] MacFeely S. Measuring the sustainable development goal indicators: an unprecedented statistical challenge. J Off Stat [Internet]. 2020 [cited 2021 May 20]; 36(2): 361–78. Available from: doi: 10.2478/jos-2020-0019.

[15] MacFeely S, Nastav B. "You say you want a [data] revolution": a proposal to use unofficial statistics for the SDG global indicator framework. Stat J IAOS [Internet]. 2019 [cited 2021 May 20]; 35(3): 309–27. Available from: doi: 10.3233/SJI-180486.

[16] Moyer JD, Hedden S. Are we on the right path to achieve the sustainable development goals? World Dev [Internet]. 2020 Mar 1 [cited 2021 May 20]; 127: 104749. Available from: doi: 10.1016/j.worlddev.2019.104749.

[17] Editorial. Time to revise the Sustainable Development Goals. Nature [Internet]. 2020 Jul 14 [cited 2021 May 20]; 583(7816): 331–2. Available from: https://www.nature.com/articles/d41586-020-02002-3.

[18] MacFeely S, Barnat N. Statistical capacity building for sustainable development: developing the fundamental pillars necessary for modern national statistical systems. Stat J IAOS [Internet]. 2017 Nov 24 [cited 2021 May 20]; 33(4): 895–909. Available from: doi: 10.3233/SJI-160331.

[19] Dang H-AH, Serajuddin U. Tracking the sustainable development goals: emerging measurement challenges and further reflections. World Dev [Internet]. 2020 [cited 2021 May 20]; 127: 104570. Available from: doi: 10.1016/j.worlddev.2019.05.024.

[20] United Nations Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG). A World that Counts: Mobilising the Data Revolution for Sustainable Development [Internet]. United Nations; 2014 [cited 2021 May 20]. Available from: https://repositorio.cepal.org/bitstream/handle/11362/40319/AWorldThatCounts.pdf?sequence=1&isAllowed=y.

[21] PARIS21. A road map for a country-led data revolution [Internet]. OECD; 2015 [cited 2021 May 20]. 52 p. Available from: doi: 10.1787/9789264235250-fr.

[22] Stuart E, Samman E. The data revolution: finding the missing millions [Internet]. odi.org. 2015 [cited 2021 May 20]. Available from: https://odi.org/en/publications/the-data-revolution-finding-the-missing-millions/.

[23] PARIS21. Statistical Capacity Development Outlook 2019 [Internet]. Paris: PARIS21; 2019 [cited 2021 May 18] p. 75. Available from: https://paris21.org/flagship/2019.

[24] Keijzer N, Klingebiel S. Realising the Data Revolution for Sustainable Development: Towards Capacity Development 4.0. In: SSRN [Internet]. 2017 [cited 2021 May 20]. pp. 1–27. Available from: doi: 10.2139/ssrn.2943055.

[25] Gal Iddo, Ograjenšek Irena. Official Statistics and Statistics Education: Bridging the Gap. J Off Stat [Internet]. 2017 [cited 2021 May 20]; 33(1): 79–100. Available from: doi: 10.1515/jos-2017-0005.

[26] Schield M. Chapter 11: Assessing Statistical Literacy: Take CARE. In: Bidgood P, Hunt N, Jolliffe F, eds. Assessment Methods in Statistical Education: An International Perspective [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2010 [cited 2021 May 20]. pp. 133–52. Available from: doi: 10.1002/9780470710470.ch11.

[27] Ridsdale C, Rothwell J, Smit M, Ali-Hassan H, Bliemel M, Irvine D, et al. Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report [Internet]. Dalhouse University; 2015 [cited 2021 May 20]. Available from: https://DalSpace.library.dal.ca//handle/10222/64578.

[28] Gould R. Data literacy is statistical literacy. Stat Educ Res J [Internet]. 2017 [cited 2021 May 20]; 16(1): 17–21. Available from: http://iase-web.org/documents/SERJ/SERJ16(1)_Grant.pdf?1498680907.

[29] Ridgway J, Nicholson J, McCusker S. "Open Data" and the semantic web require a rethink on statistics teaching. Technol Innov Stat Educ [Internet]. 2013 [cited 2021 May 20]; 7(2): 1–12. Available from: https://escholarship.org/uc/item/6gm8p12m.

[30] United Kingdom Government: Department for Digital, Culture, Media & Sport. Policy Paper: National Data Strategy [Internet]. GOV.UK. 2020 [cited 2021 May 20]. Available from: https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy.

[31] Data Pop Alliance. Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data [Internet]. White Paper; 2015 [cited 2021 May 20]. Available from: http://datapopalliance.org/wp-content/uploads/2015/10/BeyondDataLiteracy_DataPopAlliance_Sept30.pdf.

[32] Engel J. Statistical literacy for active citizenship: a call for data science education. Stat Educ Res J [Internet]. 2017 [cited 2021 May 20]; 16(1): 44–9. Available from: http://ez.urosario.edu.co/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-85021060913&lang=es&site=eds-live&scope=site.

[33] Prodromou T, Dunne T. Statistical literacy in data revolution era: building blocks and instructional dilemmas. Stat Educ Res J [Internet]. 2017 [cited 2021 May 20]; 16(1): 38–43. Available from: http://ez.urosario.edu.co/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-85021077944&lang=es&site=eds-live&scope=site.

[34] Carter J, Brown M, Simpson K. From the classroom to the workplace: how social science students are learning to do data analysis for real. Stat Educ Res J [Internet]. 2017 May 1 [cited 2021 May 20]; 16(1): 80–101. Available from: http://ez.urosario.edu.co/login?url=http://search.ebscohost.com/login.aspx?

direct=true&db=eric&AN=EJ1152506&lang=es&site=eds-live&scope=site.

[35] Higgins V, Casasbuenas V, Ricard J, Carter J. Empoderata Data Literacy Assesment and Sustainable Development Goals Data Gaps. Brazil, Colombia and Mexico [Internet]. University of Manchester and Data-Pop Alliance; 2019 [cited 2021 May 20]. Available from: https://datapopalliance.org/wp-content/uploads/2020/09/EMPODERADATAREPORT_final_oct2019.pdf.

[36] Lazer DMJ, Pentland A, Watts DJ, Aral S, Athey S, Contractor N, et al. Computational social science: obstacles and opportunities. Science [Internet]. 2020 Aug 28 [cited 2021 May 20]; 369(6507): 1060–2. Available from: doi: 10.1126/science.aaz8170.

[37] Edelmann A, Wolff T, Montagne D, Bail CA. Computational social science and sociology. Annu Rev Sociol [Internet]. 2020 Jan [cited 2021 May 20]; 46(1): 61–81. Available from: doi: 10.1146/annurev-soc-121919-054621.

[38] Gates P, ed. 18. Critical Mathematics Education. In: Issues in mathematics teaching. 1st ed. London; New York: Routledge; 2001.

[39] Méndez-Romero RA, Gauthier-Umaña V. mINNga Labs: una innovación pedagógica-tecnológica para Colombia. In: En las regiones de Colombia. La Universidad del Rosario piensa el pais. Bogotá: Universidad del Rosario; 2021. pp. 92–107.

[40] Liedtke C, Jolanta Welfens M, Rohn H, Nordmann J. LIVING LAB: user-driven innovation for sustainability. Int J Sustain High Educ [Internet]. 2012 Jan 1 [cited 2021 May 20]; 13(2): 106–18. Available from: doi: 10.1108/14676371211211809.

[41] Jones P, Carter J, Renken J, Arbeláez Tobón M. Strengthening the Skills Pipeline for Statistical Capacity Development to Meet the Demands of Sustainable Development: Implementing a Data Fellowship Model in Colombia [Internet]. University of Manchester: Centre for Digital Development Global Development Institute, SEED; 2021 [cited 2021 May 20]. Available from: https://www.gdi.manchester.ac.uk/research/publications/di/di-wp89/.