EMPODERADATA:

DATA LITERACY ASSESSMENT

AND

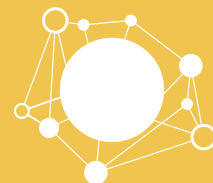## SUSTAINABLE DEVELOPMENT GOALS DATA GAPS

Brazil, Colombia and Mexico

Vanessa Higgins
Valentina Casasbuenas
Julie Ricard
Jackie Carter

October 2019

MANCHESTER
1824

The University of Manchester

DATA-POP
ALLIANCE

EmpoderaData Report

Data literacy assessment and Sustainable Development Goals data gaps

Brazil, Colombia and Mexico

October 2019

Vanessa Higgins
Valentina Casasbuenas
Julie Ricard
Jackie Carter

A collaboration between University of Manchester and Data-Pop Alliance.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

List of Tables

List of Figures

EXECUTIVE SUMMARY & NEXT STEPS

The goal of the EmpoderaData pilot research project was to establish a qualitative baseline of data literacy and existing training programs in three Latinamerican countries (Brazil, Colombia and Mexico). Moreover, the project sought to explore whether the Q-Step internship model would be applicable within the context of Latin America. Therefore, the purpose of this report is threefold: (1) understand the unmet needs in terms of data literacy skills, (2) recognize to what extent might a data literacy capacity building model can be helpful to develop these skills and last to dig up (3) in the current state regarding data availability for monitoring and evaluation of the SDGs in the three countries.

The research was undertaken in three stages. First, a workshop was held in São Paulo in May 2019 with thirty participants, involved in data, advocacy or policy-making, representing civil society, academia, private or public sector. The second involved eighteen qualitative interviews undertaken with stakeholders from the three countries and desk research. Last, a workshop was held at the University of Manchester in October 2019, to present and discuss with attendees preliminary findings. The recommendations listed below are the result of this 3-step pilot, and summarize outputs from research and workshops' discussions that will inform the next steps of EmpoderaData:

- **Hybrid Model:** Emphasis was made on building a 'hybrid' approach, bringing together participants with complementary backgrounds (data science and social science) to work together at the host organization, on one specific sustainable development challenge.

- **Content:** Based on research about the current state of Data Literacy on the three countries, and as a result of discussions with data experts coming from the public, private and civil sector, the recommendation is to focus the curriculum on **foundational skills in statistics and data science**. The goal is to foster critical analysis of data, rather than developing pure mathematical competencies and foster leadership that can think critically about data.

- **Audience:** Target audience will be students and potentially young professionals. This requires defining the basic skills potential participants should have (prior to the fellowship) and developing a selection process to assess them.

- **Sustainable Development Goals (SDGs):** At this stage, the recommendation is that training content should remain flexible, instead of restrained to (a) specific SDG(s). Host organizations interests, as well as sectoral funding opportunities, should be taken into account to inform the choice of SDGs.

- **Country of Focus:** For the next phase, it was decided to focus on a single country, instead of approaching three countries (Brazil, Mexico, Colombia) with their own contexts and complexities. After careful consideration, **Mexico** was chosen for several reasons including network, language, and possibilities of expansion.

- **Impact:** The project aims to promote a virtuous cycle of social transformation by fostering data literacy applied to addressing our society's most pressing issues, using the framework of the Sustainable Development Goals (SDGs). To achieve such a goal will require defining more concretely what are the pathways towards short-term vs medium/long-term impact, especially in countries structured by pervasive inequalities. This overall goal will be grounded into the project features, in particular target audience and modality.

BACKGROUND AND RESEARCH

a. Research for EmpoderaData scoping phase

i. EmpoderaData: project rationale

EmpoderaData, from the Spanish word empoderar for "to empower", builds upon the success of the Q-Step paid internship programme from the University of Manchester. Q-Step, for "Quantitative Step Change", was developed as a strategic response to the shortage of quantitatively-skilled social science graduates in the United Kingdom. The Q-Step Centre at the University of Manchester (UoM) is one of 18 centers and affiliates taking part in a £19.5 million Q-Step programme designed to promote a step-change in quantitative social science training.

Together, the University of Manchester and Data-Pop Alliance aim to expand upon the program's excellent results, exploring this model in the Global South as the "EmpoderaData Project". The project aims to promote a virtuous cycle of social transformation by fostering data literacy applied to addressing our society's most pressing issues, using the framework of the Sustainable Development Goals (SDGs).

This scoping phase is funded by the University of Manchester through GCRF (Global Challenges Research Funding) and carried out in collaboration with Data-Pop Alliance, with the goal to advance knowledge around two main areas of exploration:

(1) What are the unmet needs of the project's three pilot countries in terms of data literacy skills?
(2) To what extent might a data literacy capacity building model – developed and tested in the UK through the Q-step data fellow program – be helpful to develop these skills in our three pilot countries?

ii. Why promoting data literacy in (those) ODA countries?

From citizens to decision-makers, data literacy is increasingly a necessary (although not sufficient) condition for those aiming to promote social transformation in the 'data revolution' era. In 2014, the report "A World that Counts" called for an UN-led effort to mobilize a "data revolution" for all people and the whole planet in order to monitor progress, hold governments accountable, and foster sustainable development. "Data Literacy" was one of the 5 pillars of the suggested action plan.

Worldwide, new data and technologies are challenging and shaping our individual and collective capacities to learn, communicate, and make decisions. The hope is for data and analytics to allow policymakers, businesses, civil society organizations, and citizens to make better, more evidence-informed decisions. This requires improving digital inclusion, advocating for open data and civic technology, but crucially **developing data literacy across countries and population subgroups**.

Countries in the Global South need to tackle shortages in data literacy to position them to strive for social transformation and pursue of the Sustainable Development Goals, to the same extent as 'Development Assistance Committee (DAC) countries'. It is within that context, that Data-Pop Alliance and the University of Manchester decided to pilot a joint-programme aiming to uncover

and improve data literacy skills to unlock a virtuous cycle of social transformation.

Mexico, Colombia and Brazil were chosen as subjects of this scoping study for several reasons. Focusing on one region was appropriate given the project resources (including the team), timeline and logistics (in particular regarding the in-person workshop). Against that background, Latin-America and the Caribbean (LAC) presented certain advantages. The region has relatively strong official statistical and academic systems that could serve as a backbone for the so-called "data revolution," along with a wide penetration of mobile and internet technology and an established open data movement. Moreover, civil societies in these countries are active and vibrant.

In addition, conducting this work in close collaboration with the relevant local stakeholders was at the heart of the project proposal, which led the team to prioritize countries where we had a strong network. Data-Pop Alliance has a rich portfolio of activities and partnerships across sectors (i.e. academia, civil society, public and private sectors) in the region, particularly in Colombia, Mexico and Brazil.

iii. Why adopting the Sustainable Development Goals (SDGs) framework?

The 2030 Agenda for Sustainable Development adopted in 2015, is a global action plan for advancing social justice, inclusion and fighting poverty. It revolves around the adoption of 17 Sustainable Development Goals (SDGs) with 169 targets ranging from poverty and hunger, gender equality and climate action, to peace and justice. As such, it provides a common framework for action towards social transformation.

The 17 SDGs cover a broad range of social, economic and environmental development issues (see below list of SDGs), in which 1-3 indicators were set for every target in order to hold countries accountable for their advancement towards the SDGs by monitoring progress. The 230 indicators are classified into three tiers (Tier I, II and III) based on their level of methodological development and data availability. Tier 3 indicators are the hardest ones to measure, given the lack of internationally established methodology or standards for measurement.

Figure 1. Sustainable Development Goals

Source: https://www.un.org/development/envision2030.html

By reviewing each indicator, a strong emphasis on 'measurement' can be clearly seen throughout the 2030 agenda. For example, SDG 17 explicitly seeks to foster a 'Data Revolution' – to create a comprehensive and systematic indicator framework that includes improved data capacity, monitoring and accountability.

University of Manchester itself is committed to tackling the SDGs and has recently released a document (available here) showing how the institutions is addressing SDGs in four ways, namely "research impact, learning and students, public engagement activity and responsible campus operations". Moreover the University of Manchester was recognised as first in Europe and third in the world for the impact of its teaching, research and social responsibility activities as recognised through the THE University Rankings.

b. Partners and previous work
i. University of Manchester: Q-Step program

Q-Step was a £19.5 million programme (2013-2019) designed to promote a step-change in quantitative social science training in the UK. Funded by the UK's Nuffield Foundation, Economic and Social Research Council and the Higher Education Funding Council for England, Q-Step was developed as a strategic response to the shortage of quantitatively-skilled social science graduates. Q-Step funds eighteen universities across the UK to establish Q-Step Centres and Affiliates that support the development and delivery of specialist undergraduate programmes, including new courses, work placements and pathways to postgraduate study.

The Q-Step Centre at the University of Manchester, offers many ways for undergraduate students to develop their quantitative skills including five new undergraduate degree pathways that enable students to combine a specialism in a social science discipline with a rigorous training in the use of quantitative data and methods (for example Criminology and Quantitative Methods or Politics and Quantitative Methods). Undergraduate students on one of the Q-Step pathways degrees can apply for a paid internship for 8 weeks over the summer months (between the 2nd and 3rd year of their undergraduate degree). This model offers students an exciting way to develop and practice their data literacy skills in a workplace setting while being paid and gaining work experience. Students are matched with short-term quantitative data-driven projects within organisations such as the BBC, the World Bank, Universities, Local Government, market research organisations and many others. The Q-Step internship model has just completed its sixth year and has been a great success: 210 students have been placed in more than 60 organisations. Short films of student stories are being made available on the Q-Step website, and student outputs can be found under the 'student stories' section.

ii. Data-Pop Alliance: data literacy

Data-Pop Alliance (DPA) is a global coalition on Big Data, AI and sustainable development created in 2014 by the Harvard Humanitarian Initiative (HHI), the MIT Media Lab and the Overseas Development Institute (ODI), to "promote a people-centered Fourth Industrial revolution." DPA brings together researchers, experts, practitioners, and activists working around the globe across

different disciplines and industries around a common vision: making Big Data and new analytics techniques a force of positive social change in the 21st Century. Data-Pop Alliance works through 3 main modalities, deployed on the ground in various regions of the world: (1) collaborative research, (2) data-literacy training, and (3) policy and community engagement, including through our direct involvement with several in-country partnerships, activities and international network practitioners, such as the Global Partnership for Sustainable Development Data (GPSDD) and the UN World Data Forum.

DPA's "Data Literacy" pillar of work is designed in a multi-disciplinary approach to answer a growing call for resources and a framework to define Data Literacy. It has defined and seeks to promote "data literacy" as "the desire and ability to constructively engage in society through or about data." To this end, DPA has developed a framework and tools to establish core competencies towards becoming data literate. At the center of DPA's rationale and attention around data literacy promotion is "the goal of empowering citizens and communities as free agents. This can only be achieved by considering data literacy as a significant means and metric for social inclusion—where data literacy as defined and conceptualized above is promoted for and via greater social inclusion—or, more appropriately, data inclusion"[1]. More information about DPA's data literacy work is available here.

The creation of DPA Latin America was enabled by the consolidation and expansion of related research and projects in the region, in particular in Colombia, Chile and Mexico. Leveraging the resources of its team, founders, network, and partners, DPA has developed extensive experience in successfully developing research, capacity building and technical assistance projects in the region over the past 5 years.

---

[1]	E. Letouzé, Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data. 2015.

DATA LITERACY

a. Overview and definition(s)

i. Definition(s)

Data literacy is a broad and increasingly popular concept. As such, it has several definitions and is often used to refer to a variety of data-related competencies. One of the most common definitions, is that of data literacy as "the ability to use and understand data", focusing on the specialized quantitative skills to 'crunch' data, ranging from statistics to data science or machine learning competencies.

DPA proposes to conceive 'data literacy' as "being literate in the age of data". This definition includes quantitative skills but equally importantly the ability to blend knowledge from different academic disciplines (statistics, social science, data science, ethics, etc.) to use data in a responsible and inclusive manner.

In the data revolution era, 'data literate' individuals or stakeholders (policy makers, researchers, journalists, civil society and citizens) should ultimately be able to use reliable data (and analysis based on this) to inform and support research, policy-making, political debates, journalistic investigations, etc. They should also be able to question current and future ethical codes, regulatory frameworks, and legal systems governing data collection, control and use. The 'data revolution' we must strive for is one of data inclusion and empowerment.

For the purposes of this research, we have divided the 'technical' competencies of data literacy into two types: (1) traditional quantitative skills (i.e. statistics, econometrics etc.) and (2) data science and artificial intelligence skills. Our goal is to tease out data literacy needs for both dimensions in order to inform the scope of the training in the next phase of the project.

Although the Manchester Q-Step program is built around the term 'data-driven, research-led internship', for the purposes of this report and since the educational level of the beneficiary population has not been chosen, the use of the term data literacy is proposed throughout the report, and data fellowships rather than internships will be explored.

ii. Regional background

1. Access to higher education / STEM

Access to higher education remains as a big challenge for each of these countries, which impacts access to STEM careers. As shown in Table 1 below, Colombia, Brazil and Mexico all have a similar status in terms of access to higher education. Brazil has the lowest percentage of young adults accessing tertiary education (17% compared with 20% in Mexico and 22% in Colombia) and also the lowest percentage choosing STEM careers as a professional path (17% compared with 23% in Colombia and 28% in Mexico). The latter, may be related to the science score attained in the 2015 PISA test (see Table 2), which is also particularly low for Brazil.

When comparing these countries with the Latin American and Caribbean region, it is clear that all

three countries face the great challenge of reducing school drop-out during secondary education. Such a reduction is needed to achieve a more competent, prepared population in order to face such a dynamic job market. Regarding access to tertiary education, although rates in Colombia and Mexico are higher than in Brazil, they are still much lower than the average of OECD countries (60% in 2017).

Table 1. Access to Higher Education

	LATAM average (2007)	Brazil (2015)	Colombia (2015)	Mexico (2017)
% of young adults (24-34 year olds) who finished high school*	75%	64%	62%	66.7%
% of young adults (24-34 year-olds) with higher education	18%	17%	22%	20%
% of STEM graduates	-	Brazil has one of the lowest shares of graduates in STEM fields in LATAM: 17%	23%	28%

*Average age to graduate from high school is between 16-18 years in the three countries.
Source: http://gpseducation.oecd.org, https://data.worldbank.org

2. Mathematics and Science Skills - PISA Test Results (2015)

As a proxy for the educational level of all three countries regarding Mathematics and Science, we used the Programme for International Students Assessment (PISA), developed by the OECD. This Assessment is conducted every three years, and it tests 15-year-old students from more than 70 countries on reading, mathematics and science.

Table 2. Averages for PISA Mathematics Scale: overall mathematics, age 15 years (2015)

Year	Jurisdiction	All students	
		Average	Standard Error
2015	**International Average (OECD)**	**490**	**(0,4)**
	Chile	423	(2,5)
	Uruguay	418	(2,5)
	Argentina	409	(3,1)
	Mexico	**408**	**(2,2)**
	Costa Rica	400	(2,5)
	Colombia	**390**	**(2,3)**
	Peru	387	(2,7)
	Brazil	**377**	**(2,9)**
	Dominican Republic	328	(2,7)

Source: https://pisadataexplorer.oecd.org

As can be noted in Table 2, even though Mexico is the fourth highest ranking Latin American country (after Argentina, Chile, and Uruguay), still all three countries show a low level of mathematical skills, in comparison with the international average for OECD countries. For mathematics, all three countries are in Level 1a, while the OECD average is level three. At this mathematical level, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are almost always obvious and follow immediately from the given stimuli (OECD, 2015).

Table 3. Averages for PISA Science Scale: overall science, age 15 years (2015)

Year	Jurisdiction	All students	
		Average	Standard Error
2015	**International Average (OECD)**	**493**	**(0,4)**
	Chile	447	(2,4)
	Uruguay	435	(2,2)
	Argentina	432	(2,6)
	Costa Rica	416	(2,4)
	Mexico	**416**	**(2,1)**
	Colombia	**416**	**(2,4)**
	Brazil	**401**	**(2,3)**
	Peru	397	(2,4)
	Dominican Republic	332	(2,6)

Source: https://pisadataexplorer.oecd.org

Regarding scientific literacy, understood as the use of scientific knowledge to identify questions, acquire new knowledge, explain scientific phenomena, and draw evidence-based conclusions about science-related issues, Colombia and Mexico appear to be in Level 2, while Brazil remains in Level 1a. All three countries still have important educational challenges in order to reduce the gap between them and the OECD average which is again level 3, and increase the number of students who are literate in science and mathematics. At level 3, students are able to draw upon moderately complex content knowledge to identify or construct explanations of familiar phenomena. In more complex situations, they are able to construct explanations with relevant cueing. They can also draw on elements of procedural or epistemic knowledge to carry out simple experiments in a constrained context. Students at this level are also able to distinguish between scientific and non-scientific issues and identify the evidence supporting a scientific claim (OECD, 2015).

However for science in particular, Colombian and Mexican students are able to draw on everyday content knowledge and basic procedural knowledge to identify an appropriate scientific explanation, interpret data, and identify the questions being addressed in a simple experimental design. Students can also use basic, everyday scientific knowledge to identify a valid conclusion from a simple data set. They also demonstrate basic epistemic knowledge, by being able to identify questions that can be investigated scientifically (OECD, 2015).

In the case of Brazil, students are able to use basic or everyday content to recognize or identify explanations of simple scientific phenomena. With support, they are able to undertake structured scientific enquiries using no more than two variables. They are also able to identify simple causal or correlational relationships and interpret graphical and visual data that require a low level of cognitive demand. Students can also select the best scientific explanation for simple data in a familiar, personal, and local context (OECD, 2015).

3. Inequalities

The three countries are structured around pervasive and cross-cutting inequalities, which define access to opportunities and services, i.e. education, employment, healthcare etc. Although it is highly complex to quantify inequality, we use the World Bank GINI Index as an indication of the above. As evidenced in Table 4, Brazil is still the most unequal country in Latinamerica, and Colombia is listed as the fourth after Honduras and Panama. Overall, the three countries remain amongst the most unequal countries as per the OECD (2017) with a GINI index of 0.49 for Colombia, 0,43 for Mexico, and 0.53 for Brazil, where a higher score indicates a higher inequality amongst the population.

Table 4. Gini Index - World Bank Estimates

Year	Country	Index
2017	**Brasil**	**0.533**
	Colombia	**0.497**
	Costa Rica	0.481
	Chile	0.456
	Ecuador	0.447
	Bolivia	0.441
	Mexico	**0.433**
	Peru	0.432
	Argentina	0.406
	United Kingdom	0.332

Source: https://data.worldbank.org

Despite an overall increase in the access to academic opportunities for people from a lower socioeconomic stratum, this gradient is still strong on educational outcomes for all of the three countries. For instance, students from wealthier backgrounds are close to three times more likely to finish upper secondary education than their less privileged peers (OECD, 2017).

Therefore, breaking the vicious intergenerational cycle of poverty remains particularly challenging, due to lack of opportunities for social and economic progress. Indeed, certain subgroups of the population, in particular women, non-white and those living in rural areas, are particularly affected by cross-cutting layers of discriminations. For example, according to OECD's Education at a Glance Report, in Colombia non-indigenous, non-black communities, hold a steady advantage, as they are almost twice as likely to complete upper secondary education.

Surprisingly, in the case of gender inequality, there seems to be a difference between the perception of the interviewees and the 2018 Global Gender Gap Report data. According to this source, the differences among graduates in mathematics, statistics, and science was of less than 0.1 for the three countries, showing 96% equality in Brazil, 95% in Colombia and 100% equality amongst men and women.

Moreover, there are contradictions between the Global Gender Gap Report and other sources consulted. For example, the 2017 Education at a Glance reports from the OECD states that in Colombia 78% of young women are expected to graduate from upper secondary education at some point in their lives, compared to 62% of men. Furthermore, the female upper secondary graduation rate in Colombia is higher than in other Latin American countries including Brazil, and

Mexico (OECD, 2017). According to these reports of the OECD countries, Colombia is closer to achieve gender balance in almost all fields of study at the tertiary level. For example, two men graduate for every woman in STEM careers in Colombia, compared with an average of three men for every woman in OECD countries (2017).

As it is showed, the above contradicts the supposed gender equality in terms of STEM careers and graduation rates described by the Global Gender Gap Report published by the World Economic Forum. This is probably related to the complexity of quantifying these inequalities, which could be a challenge for the project in the future. In conclusion, across the three countries, access to education opportunities are conditioned by structural inequalities. In each country, the intersection of inequalities results in different exclusion / inclusion dynamics that need to be taken into account when studying access to education and science.

b. Methodology
i. Objectives and research questions

The goal of this pilot research is to establish a qualitative baseline of the state of data literacy and existing training programs in the three countries, and explore whether the data-driven internship intervention model pioneered by the University of Manchester would be applicable within the context of Brazil, Colombia and Mexico. Overall, the research should enable us to:

- A1: baseline data and statistical capacity of those selected countries for delivering the SDGs
- A2: explore their attitudes and willingness to having an intervention SDGs fellowship programme developed with their relevant stakeholders, building on the success of the UoM Q-Step internship programme.

The qualitative research on the topic of data literacy was undertaken in two stages:

- A working group session at the Fundação Getúlio Vargas in São Paulo (May 16-17, 2019), more information here. The workshop gathered approximately 20 participants from different sectors, all of them involved in data advocacy either from the civil society, academia, private or public sector. This session aimed to (1) map the needs for training, (2) identify partners for future fellowship programs, and (3) present the current UoM Q-Step internship model and explore its applicability in the three pilot countries (more information about the WG is available here and a summary of the discussions can be found here).
- Fifteen interviews were conducted with stakeholders from Brazil (6), Mexico (5) and Colombia (4). Interviews were conducted between 11 June-8 July 2019, in Portuguese, Spanish and English. The sample of interviewees was selected to represent a diversity of sectors (namely academia (private and public), civil society and public sector), and as much as possible provide a gender and regional (within countries) representation. The interview guide used for all interviews is available here. To understand the monitoring and evaluation system for the SDGs in each country, additional three interviews were conducted with professionals working for the NSO in charge of the 2030 Agenda.

ii. Interview guide for the qualitative interviews

The interviews were structured around three topics:

- The availability of, and need for, training in traditional quantitative skills. Focussing on traditional quantitative data, such as survey, census or official aggregate data). This included questions on inequalities in the capacity.

- 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. This included questions on inequalities in the capacity.

- The interest in the adoption of the Q-Step internship model.

The full interview guide can be found here.

c. Context and results (per country)
i. Brazil
1. Background information

- Access to higher education and employment is still conditioned by structural inequalities:

  - Higher education enrollment rates for students from the private secondary school system were more than double that for students from the public school system. Whites have higher admission rates than blacks or browns in both public and private systems (IBGE, 2018). According to PISA, 36% of students do not have access to a computer at home (PISA, 2015).

  - Considering the same levels of instructions, the unemployment rate is always higher non-whites (black, indigenous). Having higher education is a factor that contributes to access to the labor market with more intensity for black or brown population, but not enough to match them with white peers (IBGE, 2018).

  - Moreover, women are overrepresented in the salary range from 0 to 1.5 minimum wages and in precarious jobs. This disparity reflects gender stereotypes and social discrimination that still relegates women to less favored jobs and positions.

2. Results / insights from interviews

Overall capacity in data literacy (both 'traditional quantitative skills' and data science') varies across sectors:

- Academia concentrates the highest capacities, particularly in specializations that traditionally include mathematics or statistics. In social sciences courses, quantitative training is mostly weak (particularly at the undergraduate level but also at postgraduate level). There is less quantitative training in private universities than in public universities. There is an emerging offer of Data Science specializations (University of Pernambuco, Fundação Getúlio Vargas SP) as well as courses outside universities (but these are often expensive and thus limited to the private sector).

- ⬜ Overall, there are good capacities in "data visualization", but much less in techniques such as machine learning or parallel computing.
- ● Overall, civil society organizations outside the 'tech bubble' do not have strong quantitative skills (i.e. lack of capacity to use data, to set up, and even to answer matters that are more analytical). They lack the capacity to understand/ discuss more in-depth the functioning and implications of algorithms. There are (few) organizations focused on working with data and evidence (Escola de Dados, Escola de Ativismo, Data Lab, Social Good Brasil).
- ● In the public sector, capacities seem to be concentrated only in specific teams, mostly at federal level and in the biggest (richest) municipalities (e.g. Rio de Janeiro or São Paulo) or states. An additional issue raised is the challenge of talent retention in the public sector as salaries are not competitive with the private sector.
- ● Additionally, the cost of the necessary hardware infrastructure, in particular for 'big data' analysis, was raised both by public sector and academia interviewees; related to the need to develop better data management competencies in the public administration.

The structural inequalities in Brazil (gender, social class and race) permeate the data literacy eco-system and make access to training unequal for certain subgroups (women, non-white, poor). Both lack of awareness and representativity further exacerbate inequalities in terms of access.

- ● Inequalities increase in more advanced knowledge/skills that require more years of schooling etc.
  - ⬜ There is a need for better data about inequalities in the STEM sector, in particular about representativity of women, black women and non-whites.

Overall, civil society, public sector and academia (particularly undergraduates and graduates in applied social sciences) would greatly benefit from better training in statistics and methodology to understand and use data in a methodological and sustainable manner.

- ● There is a need for 'hybrid' profiles, that are able to understand the question / issue from a social science / policy point of view and translate it into a data question / project.
- ● Rather than developing very advanced technical skills, there would be more value in providing basic training in statistics and methods, towards solving problems / questions and project development and implementation.
- ● Outstanding question: should we teach statisticians/programmers about social sciences or policy-makers to count/code?

About the Q-Step internship model:
- ● In June 2019, Open Knowledge Brasil launched a call for applications for the first cohort of the "Frictionless Data Reproducible Research Fellows Programme". To our knowledge, there aren't other data fellowship programs following the Q-Step model. The Data Science for Social Good Fellowship at the University of Chicago has been mentioned by a few interviewees as a good example to follow.
- ● One of the main challenges is ensuring inclusion of the less privileged subgroups of the population. This will require an appropriate outreach strategy, potentially motivation work and considering a structure compatible with daily routines. Otherwise, the risk is that only the elite will participate.

- The fellowship should be undertaken in partner with academia, private and public sectors. In doing so, short fellowships may not work because it needs longer term partnerships to get public administration to use the results of projects.
- Suggested vision to build a group of people that are very qualified to work with data on specific campaigns. e.g. gender data, mobility data, environmental data, public health data. Think about specific sectors, and their particular challenges. Also think about different audiences e.g. faster training for a determined audience, to reach a larger number of people or a more specific training for a group that is going to have a bigger impact, that is going to be able to multiply more that knowledge etc. Also, take advantage of the university structure that can facilitate the movement to multiply a project.

The main actors from the data literacy ecosystem mentioned by interviewees include:

Sector	Organization	Description*
Academia	Universidade Estadual de Pernambuco (UFPE)	Public university located at the northeastern state of Pernambuco. Data literacy program (MA, BA)
	Fundação Getulio Vargas (FGV)	One of the most important private universities in Brazil with presence in Rio de Janeiro, São Paulo and Brasilia.
	International Association for Statistical Education (IASE)	International umbrella organization for statistics education
	Universidade de São Paulo (USP)	Brazil's biggest public university
	Instituto Brasileiro de Pesquisa e Análise de Dados (IBPAD)	Brazilian institute of data research and Analysis, they offer courses in data analysis and visualization
	Universidade Federal da Bahia (UFBA):	Largest public university in the state of Bahia
	Escola Nacional de Ciências Estatísticas (ENCE)	Public statistics school based in Rio de Janeiro, part of IBGE (Brazil's NSO)
	Universidade Federal de Minas Gerais (UFMG)	Public university of Minas Gerais city
	Universidade Federal do Rio de Janeiro (UNIRIO	Rio de Janeiro's biggest public university
	International Political Science Association (IPSA)	Co-leads in partnership with the Department of Political Science from USP the Summer School "Concepts, Methods and Techniques in Political Science, Public Policy and International Relations" (in english).
	Instituto de Estudos Sociais e Políticos (IESP-UERJ)	Holds a "Winter School" on research methods and techniques for social sciences.

Civil Society	Olabi	Social organization that seeks to democratize technology production and access
	Escola do Dados - Open Knowledge	Global network in data literacy. They have activities in Brazil since 2013.
	Data for Good	Social organization focused on data for social good
	Data Lab	Private data lab from the Experian group focused on solving problems through big data and AI.
	Women Who Code	Non-profit organization which inspires women to excel in technology careers
Public Sector	Instituto Brasileiro de Geografia e Estatística (IBGE	Brazil's National Statistical Office

* Blue colored boxes means the actor offers a data literacy program. For this report, we are including any type of education on the topic (undergraduate, post-graduate and non-formal education).

ii. Colombia

1. Background information

- 39% of students do not have access to a computer at home (PISA, 2015).

- There are only 75 formal education programs on Data Science and analytics offered in the country, being only 5 of them for a postgraduate level and only one offer in PhD (Ministerio de Educación Nacional 2017).

- In 2017, only 2.45% of the scholarships offered by the colombian government were on data skills related topics (Documento 3, Estrategia nacional de Big Data para el Estado Colombiano, Data-Pop Alliance, 2018).

2. Results / insights from interviews

- Overall, the level of data literacy seems to vary across all sectors, but in general is perceived as limited, especially for data science skills.

  - Within academia, the offer in traditional quantitative training seems to be different across public and private universities and between applied sciences and social sciences; where it was described as weak. There is a tendency to focus on theory in social sciences rather than the real use of data, Universities are making an effort to expand their offer in data management in their undergraduate and postgraduate courses, focusing in data science for business decisions. On the other hand, Data Science was described as being very weak in public universities, lacking not just the offer for students but also teachers.

  - Civil society organizations were described as having a low level of working with data in general. Journalists and media professional were described as having the

lowest level of data skills, lacking basic analysis skills, which are key skills in their scope of work.

- The public sector was described as lacking capacity to make decisions based on data science analysis. Regarding traditional methods, the public sector was also described as having needs in basic data skills, lacking an understanding of basic methodologies for creating or analysing indicators.

- On a country level, big deficiencies and delays were listed for public universities and entities regarding Data Science. Teachers and professionals from the public sector, were described as being very focused in theory, sometimes even refusing to use data science tools or learn new methodologies in their research and work. A change in which education and public policies are conceived was listed as necessary, being the key goal a data-driven government and educative system.

- As for Mexico and Brazil, Colombia's structural inequalities, such as socioeconomic status, race, and gender impact on the capacity in both traditional and non-traditional data literacy. Region of origin was also listed as an inequality factor, impacting students and professionals from municipalities far from the center of the country.

- The biggest need for social science professionals is in basic traditional quantitative skills, especially analysing and processing basic data. Advanced skills are not described as being crucial to change how data is being used. The biggest needs are in basic knowledge of data management and software.

- Both civil society and public sector were described as the sectors that would benefit the most from data literacy training. Also undergraduate students from private non-validated / public small regional universities would particularly benefit from this opportunity.

- One of the barriers listed for the project is that data literacy is not perceived as an unmet need in the public sector, so awareness raising is recommended for working with government professionals.

The Q-Step internship model:
- There is nothing like the fellowship model in Colombia. The only similar course is the Maths course at Universidad del Rosario
- The main barriers will be the motivation to do a fellowship because of the lack of perception of the need.
- Alternative models suggested were:
  - Run 6 month placements, rather than summer placements, as many students are from low socio-economic backgrounds so need the financial help.
  - Focus placements on the groups with unequal capacity such as those with low income - this would also mean focussing on more basic topics
  - Focus on public sector professionals instead - an executive training scheme for professionals working in the public sector would be more impactful.

The main actors from the data literacy ecosystem mentioned by interviewees include:

Sector	Organization	Description*
Academia	Universidad Nacional de Colombia	Colombia's biggest public university with a presence in more than 9 colombian departments
	Universidad del Valle	Public University located in Valle del Cauca region
	Universidad de los Andes	Colombia's top ranked private university
	Universidad Santo Tomás	Catholic private university with a presence in Bogotá, Villavicencio, Medellìn, Tunja and Bucaramanga
	Universidad del Rosario	Private colombian university based in Bogotá
	Alianza CAOBA	Public-private association for Big Data excellence and analytics
Civil Society	DataSketch	Civil society initiative focused on open data and data journalism
	Random Monkey	Data analysis and visualization initiatives
	Sociedad Colombiana de Estadística	Colombian Statistics Society
Public Sector	Departamento Administrativo Nacional de Estadística - DANE	Colombia's National Statistical Office.
	Departamento Nacional de Planeación - DNP	National Planning Department
	RutaN	Innovation Agency for the municipality of Medellín

Blue colored boxes means the actor offers a data literacy program. For this report, we are including any type of education on the topic (undergraduate, post-graduate and non-formal education).

iii. Mexico

1. Background information

- 42% of students do not have access to a computer at home (OECD, 2015).

- Mexico has made great progress in increasing tertiary educational attainment from 16% in 2008 to 23% in 2018. (OECD, 2018).

- In 2016, expenditure per student from primary to tertiary level was USD 3 6001 per year, the lowest across OECD countries (OECD, 2017)

- According to the 2017 Education at a Glance Report, Mexico has one of the largest shares shares of students entering science related tertiary education across OECD countries. In 2015, Mexico's share of women among new entrants to tertiary STEM (32%) was slightly above the OECD average (27%) (OECD, 2017).

- In 2016, 32% of new entrants to tertiary education chose the science, technology, engineering and mathematics (STEM) fields of education, 5 percentage points more than the OECD average, placing Mexico in the top six OECD countries for this measure.

2. Results / insights from interviews

Similarly to Brazil, the overall capacity in data literacy (both 'traditional quantitative skills' and data science') seems to vary across sectors:

- In academia, there is a solid offer in traditional quantitative skills training (both at undergraduate and master's levels), and to a lesser extent an emerging offer in data science (e.g.: ITAM offers undergraduate and masters, UNAM too). In both cases, this training is primarily in traditionally quantitative careers. For the latter, demand (and awareness) is still evolving. There is a lack of social science training focused on basic statistics / data science for public policy or 'implementation' (not necessarily complex statistics).

- In civil society and public sector, there is an emphasis on the need to develop a quantitative/ logical mindset (rather than in-depth programming skills), basic training in statistics and core ethical principles related to data use.

- Journalists and social scientists, in particular those working with government, as well as the public sector, would greatly benefit from introductory data literacy training. Most professionals working in those sectors have a very limited data background. There is a recommendation to prioritize women, indigenous groups and marginalized populations (i.e. rural).

In Mexico, just as in Brazil and Colombia, there are inequalities in access to higher education for certain subgroups (low socio-economic status, gender, race, region and age).

- Going to university implies a certain economic level, even if it is a public university.

- Curricula that require several years of studies (e.g. data literacy-related) represent a high opportunity cost. There is a feeling that the data science field is/will be more equal (both in terms of gender and social class) than traditional quantitative curricula (such as statistics), because it comes with a higher promise of return and impact.

- Demand for data literacy has grown among the medium and higher income communities.

- According to Edgar Barroso from Universidad Tecnológica de Monterrey, "it is necessary to understand that especially in developing countries such as Mexico, empowering people

should be the key objective, otherwise it is just imposing... elites will remain as the owners of knowledge and then the rest of the communities will always be left behind".

- The key to the problem seems to be the lack of numeracy rooted in the education system, which permeates in the whole society, but especially affects the poor, the indigenous, and the vulnerable.

In conclusion, Mexico's data literacy ecosystem is very regionalized, with more developed skills in the private sector of big cities like Mexico City and Guadalajara. The majority of these skills are not impacting government and/or public policies, therefore most interviewees suggest that working with the public sector, or people pursuing a career in social sciences or journalism would be very impactful.

- As Mexico decided to remove mathematics and philosophy from technology and engineering majors, professionals working with data in these fields do not know how to ask the right questions, therefore, there is an unmet need from social science professionals in data, but there is also a challenge with professionals that know how to work with data but lack how to analyze data from a research point of view.

The Q-Step internship model:

- There were a few limited examples of similar models given by interviewees:
    - ITAM data science masters graduates do internships in companies. It was noted that there is something similar starting in Autumn for undergraduates but the interviewee was unsure of the details.
    - Data Science for Social Good at the University of Chicago. summer-long internship process in which four highly skilled recently graduated or undergraduates or during their postgrad studies, focus on a project that was given by the government.
    - A "Data Lab" in govt was launched and partnered with CIDE, "Laboratorio Nacional de Políticas Públicas". A fellowship program was launched for recent graduates. To work for 8 months in a government team that had a specific problem. It was very successful but difficult to find people with the right skills to do the project.
- It was noted that most universities do have placements and projects for the students. By law every student before graduating needs to do a number of social service and apprenticeships but none (that the interviewee knew of) were data and government linked.
- Barriers to implementation of a Q-Step internship models were: whether companies would sign up to it; getting funding for a scheme; getting people with the right skills; a fellowship would need to be longer than 8 weeks (minimum 12 weeks); it would work in private universities and good public universities but not in 'bad' public universities because of the lack of English, social skills etc; projects have to benefit the officials as well as the students.
- Suggestions/alternative models were:
    - work with NGOs, rather than the public sector, as they have better quality and more accessible data;
    - open it to wider groups of people, not just students
    - internships are not normally paid so it would be good to have some investment from the company to work on a project that solves a problem they have.

The main actors from the data literacy ecosystem mentioned by interviewees include:

Sector	Organization	Description*
Academia	Universidad Nacional Autónoma de México- UNAM	Mexico's largest public university
	Instituto Tecnológico Autónomo de México - ITAM	Private tech-focused university
	Instituto Tecnológico y de Estudios Superiores de Monterrey	Also known as Tecnológico de Monterrey: Mexico's biggest private university
	Centro de Investigación y Docencia Económica- CIDE	Public university based in Mexico City and Aguascalientes focused in social sciences
	Colegio de Educación Nacional Técnica para el Estado de México - CONALEP	Public technical and technological school
	Centro de Investigación en Matemáticas - CIMAT	Mexican center for mathematical research
Civil Society	Data Cuatro	Data analysis and visualization
	SocialTIC	Digital technology for social change NGO
	Sociedad Mexicana de Ciencia de Datos	Mexican Data Science Society. They run a three day data literacy program.
Public Sector	Instituto Nacional de Estadística y Geografía -INEGI	Mexico's National Statistics Office

Blue colored boxes means the actor offers a data literacy program. For this report, we are including any type of education on the topic (undergraduate, post-graduate and non-formal education).
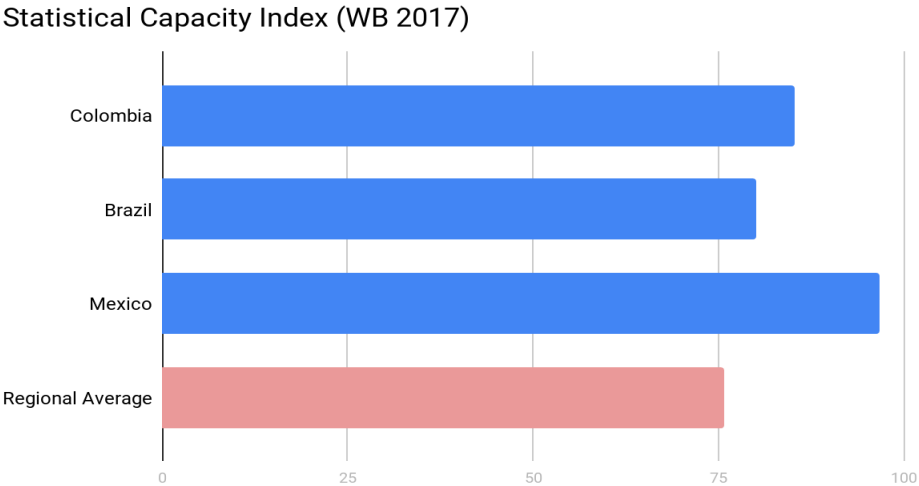
SUSTAINABLE DEVELOPMENT GOALS

a. Rationale and relevance

The 2017 UN Report on Sustainable Development emphasized "the need for reliable, timely, accessible and disaggregated data to measure progress, inform decision-making and ensure that everyone is counted". The Agenda 2030 will be all but impossible to achieve without greater availability of quality, timely data which governments may leverage to make evidence-based decisions, and for citizens to hold them accountable.

Countries across the globe are currently thriving towards achieving and measuring the SDGs indicators. As part of their effort, countries are dedicating statistical resources to measuring the set of more than 230 SDG indicators. Overall, Mexico, Colombia and Brazil are recognized in the region for their high statistical capacity. According to the Statistical Capacity Index, developed by the World Bank, Mexico has the higher result in the region, followed by Chile and Perú. The Index is a composite score assessing the capacity of a country's statistical system. It is based on a diagnostic framework assessing the following areas: methodology; data sources; and periodicity and timeliness.

Figure 2. Statistical Capacity Index

Statistical Capacity Index (WB 2017)

Source: Statistical Capacity Indicator Dashboard
http://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx

Data-Pop Alliance has ongoing research and programs around measurement and Sustainable Development. Our work in this field aims to contribute to the international community's debate on the role of innovative methodologies, including new approaches based on artificial intelligence and Big Data, in ensuring global accountability towards the UN Agenda 2030 for Sustainable Development. Our research is presented in papers, focused discussions and workshops. More information can be found here.

b. Methodology
i. Data collection and visualization

The purpose of the research is to understand the progress of Brazil, Colombia and Mexico in measuring SDGs. To do so, we have collected information from official sources in Brazil, Colombia and Mexico and used it to build an interactive visualization (see below). This visualization tool should support the choice of SDGs on which the EmpoderaData project can / should focus. An overview of the methodology used to build the interactive visualizations can be found below:

- We used the "Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development". The table can be downloaded (excel version) here.

- We collected information (either manually or by web scraping) about the status of measurement of each indicator from each country (Brazil, Colombia, Mexico) official page for tracking SDGs measurement (see 'Source' row in table below).

    - .csv Brazil can be found here.
    - .csv Mexico can be found here.
    - .csv Colombia can be found here.

- Based on the official information for each country, we coded each indicator with a "score" as per table below:

Score	Brazil	Colombia	Mexico
3	"Indicator produced" (IBGE, 2019)	"There is trustable, updated, quality data available" (DNP, 2017)	Data available for 2000-2016. Next update (2017-2018) on 06/2019. (SIODS, 2018)
			Data available for 2013-2017 (SIODS, 2018)
			Data available for 1990-2017 (SIODS, 2018)
			Data available for 1990-2017 (SIODS, 2018)
			Data available for 2010-2017. Next update on 02/2020. (SIODS, 2018)
			Data available for 2000-2018 (SIODS, 2018)
			Data available for 2003-2018 (SIODS, 2018)
			Data available for 1995-2018 (SIODS, 2018)
			Data available for 1996-2018 (SIODS, 2018)
			Data available for 2014-2018 (SIODS, 2018)

2	"Indicator under analysis / construction" (IBGE, 2019)	"Some data available / Needs to be updated" (DNP, 2017)	Data available for 1994-2014, it needs to be updated. (SIODS, 2018)
			Data available for 1992-2016 (SIODS, 2018)
			Data available for 2000-2016 (SIODS, 2018)
			Data available for 2007-2016 (SIODS, 2018)
			Data available for 2011-2016 (SIODS, 2018)
			Data available for 1998-2016 (SIODS, 2018)
			Data available only for 2016 (SIODS, 2018)
1	"No data" (IBGE, 2019)	To date there are no comprehensive official figures (DNP, 2018)	Data available only for 1993-2011 (SIODS, 2018)
			Data available only for 2013 (SIODS, 2018)
	"No global methodology" (IBGE, 2019)		Data available only for 2014 (SIODS, 2018)
			Data available only for 2015 (SIODS, 2018)
na	"Does not apply to Brazil" (IBGE, 2019)	Goal will only be measured in financial / resource mobilisation indicators (DNP, 2018)	"Indicator not included in the National Monitoring System" (SIODS, 2018)
		"Considered a Global Responsibility Goal, so its measurement is not included in the National Monitoring System" (DNP, 2018)	
		Not included in the National Monitoring System	
Source	IBGE, 2019	DNP 2018	SIODS, 2018

Caption:
- **Colors for indicators:**
  - Red = 1 = indicator not measured (either lack of data or global methodology)
  - Yellow = 2 = indicator under analysis / construction
  - Green = 3 = indicator produced
  - White = na = does not apply to country or is not included in the National Monitoring System
- **Colors computed for targets and SDGs:**
  - At levels of target and SDG, the colors are computed as an average of their 'parent' (indicator and target respectively).
  - How to read: the darker the orange of the SDG (goal, inner circle) the less advanced is the country towards measuring it **vs.** SDGs colored in 'light green', which have a higher proportion of indicators measured / being measured.
  - NB: Indicators colored white, are ponderated 'red' for the average of their 'parent'.

Color scale:

ii. Interviews

Three interviews were conducted with relevant professionals, working directly with SDGs for the government of each country. In the case of Colombia, an interview was held with the Department of National Planning, with an officer directly involved with the designing of the national monitoring system for the SDGs. In Mexico, an interview was realized with an officer from INEGI (Mexico's NSO). In the case of Brazil, we interviewed two project officers, leading a project focusing on SDGs monitoring at Abrinq Foundation, as well as a member from ABRAPS ("Associação Brasileira dos Profissionais pelo Desenvolvimento Sustentável"). The interview guide used for this interviews is available here.

iii. Countries' progress towards measuring SDGs indicators
1. Brazil

Figure 3: Overview of progress towards SDGs Indicators - Brazil

Source: own elaboration based on data from https://ods.ibge.gov.br
For full and interactive visualization, including targets and indicators: (1) download this .html file: https://drive.google.com/open?id=14WlAQvvBg79WCZYdDlN7CzuHB1-lz0CF, and (2) open with your browser.

Overall the platform produced by the IBGE [https://ods.ibge.gov.br/] is said to work well, for the purpose of visualizing progress on measuring the indicators. For professionals (from civil society for e.g.) working on SDGs monitoring, the platform itself is not the best source of information, as it doesn't allow to download the data per se. Interviewees have clarified that the indicators labelled "under construction" in the platform, are indicators for which no methodology is currently available.

According to interviewees, as other countries in the region, Brazil has done an exercise to "adequate" some indicators to the national context, a process lead by IPEA (Instituto de Pesquisa

Econômica Aplicada). According to [IPEA's report](#) about the exercise, "in some cases, Brazil has already achieved set goals; in others, these refer to problems that are not observed internally, or do not address issues of major relevance to the country. Moreover, in Brazil, commitments to the SDGs and the implementation of public policies required to achieve them can be assumed and implemented in their three spheres: federal union, states and municipalities. Therefore, it is necessary to adapt as goals so that the federates feel contemplated in national exports" (IPEA, 2018).

In that sense, the platform produced by IBGE provides a partial overview of the country's progress towards measuring SDGs, as it only displays the 'international' indicators. According to interviewees, it is important to bear in mind that for certain SDGs, there are still no metrics for measurement (for e.g. SDG 5). Given the country's institutional structure, and the particular autonomy of municipalities in a variety of matters, SDG 11 is of special relevance in the case of Brazil. Given widespread inequalities, SDG 10, and any goal / indicator related to inequality is extremely relevant. In particular, there are important data gaps in terms of quantifying racial inequalities.

Additionally, it is important to note that in April 2019, the current government has dissolved the "National SDGs Commission" (Comissão Nacional para os ODS), which had been created in October 2016. The Commission gathered government and civil society organizations, to discuss and support SDGs measurement. This 'verticalization' of monitoring and evaluation hinders SDGs measurement, in particular at the level of local governments.

2. Colombia

According to the Colombian and Mexican interviewee, when speaking of SDG measurement and evaluation, the Colombian system seems to be an example at a regional level. In general, Latinamerican countries have few methodologies developed for measuring Tier II and Tier III indicators. To address this challenge, Colombia took the lead in 2016, and developed complementary, country indicators for each goal to measure progress in each of them. This has allowed the government to measure each goal in an integral way, understanding and integrating the countries characteristics and needs into the framework. The platform developed by the national government works pretty well, but as stated by the interviewee, public agencies need to adhere to delivery dates, because it is common to see outdated information with available updates pending.

Regarding goals in specific, indicators related to environment and climate change (SDGs 13-15) are the ones with more data gaps and lack of methodological frameworks. Particularly SDG #14 has been really challenging, and as told by the Government, support with environmental SDGs measurement would be really impactful. Other goal in which more data is available but more methodologies were listed as needed are the indicators related to gender (SDG #5). Institutionalism for gender equality in Colombia is weak, so precisely, support this gender needs for equality with data driven policies would really impact the progress towards this goal.

3. Mexico

Figure 4: Overview of progress towards SDGs Indicators - Mexico

Source: own elaboration based on data from http://agenda2030.mx
For full and interactive visualization, including targets and indicators: (1) download this .html file: https://drive.google.com/open?id=14ejGKiadYbyWRCy0vvJ-A4J1gr-KR7LI, and (2) open with your browser.

According to the Mexican interviewee, the new government elected (2018) has impacted the SDG evaluation and monitoring system, since the plan that was presented at a technical level was rejected, to choose one not based on data and with a more political objective.

Although the platform seems to work well, two areas of improvement stand out. Although visualization is good, it is not possible to compile or compare indicators or data, making downloading and organizing quantitative analysis much more difficult. Beyond this, the underlying problem seems to be the outdated data, which is caused by the lack of commitment of national agencies in charge.

Just as in the case of Colombia, under the motto of "leave no one behind", the country listed 54 indicators for the country framework, adapting each goal to the national reality. However, after evaluating with the Office of the President a few months ago, some of the indicators will be reopened for discussion, as they have methodological flaws that need to be revised.

Apparently, the project should focus on SDG 10 and/ or 5 (gender equality and the reduction of inequalities). As stated by the interviewee, Mexico has an important challenge to reduce gender violence and femicide, and unfortunately government has not been able to develop a methodology to measure it towards public policies in violence prevention and reduction. On SDG # 10, just as in the case of Colombia and Brazil, inequality has been very light measured by the government, lacking updated data and a comprehensive methodology, and unfortunately it is still one of the greatest challenges faced by latinamerican countries.

Finally, it is worth mentioning that the Mexican state has decided not to monitor any indicators for the SDG #12. This could pose an interesting challenge to the project, since it would certainly generate a lot of impact to be able to gather information on sustainable food production and consumption, which is currently not being measured by public entities in the SDG framework.

CONCLUSIONS AND INPUTS FOR EMPODERADATA

a. Main takeaways

Data literacy needs:

- In all three countries there is a need for more data literacy training across academia, (particularly undergraduates) the public sector and civil society.
- All three countries emphasize the need to build data literacy capacities towards fostering 'hybrid' professionals that can understand, use and analyze data for 'social sciences' purposes, i.e. evidence-informed policy, journalism, activism etc.
- The biggest training need is for basic skills such as basic statistics, basic analysis skills and basic methodological skills. This was highlighted in all three countries.

Relevance and interest of the 'fellowship' model:

- Training + paid fellowship model considered a promising set-up in the three countries. Several interviewees have stated their interested in partnering for such a program.
- Country context needs to be considered when developing an fellowships model. Different ideas were suggested for each country, in terms of how the Q-Step model might need to be adapted. This would need much further exploration in any future work.

Cross-cutting inequalities:

- We would need to consider a specific outreach strategy to involve typically excluded subgroups. Colombia and Brazil interviewees listed as a concern the fact of offering a short term fellowship to students coming from poor backgrounds in need of a stable source of income.
- Interviewees described a possible paradox: prioritizing the most excluded subgroups would generate the biggest impact but 'ideal' target audience would have an acceptable educational background. Need to find the balance to avoid risk that the program would end up working with those with more "privileged background".

Additional considerations:

- Overall, agreement that there is danger to unsupervised online courses - as it is not possible to enforce standards on learning, and ensure that core statistics (and ethical) principles are learnt.
- In the public sector, prior to considering data literacy, it's important to note that data system / infrastructure deficits affect governments' capacities to work with data.

b. Country of focus: Mexico

Strategically, the project team, supported by the Steering Group made the decision to focus on one country only for the deployment of the next phase. After careful consideration, Mexico was chosen. The choice was oriented by the strength of our (DPA + UoM) network, language (Spanish eases potential expansion of the project to Spanish speaking countries in Latin-America) and in-country presence of DPA, which is expected to facilitate project deployment.

As stated above, Mexico has an overall developed statistical system, which facilitates the development of this pilot, which seeks to use traditional and non-traditional sources to complement the current measurement of the SDG indicators. On the other hand, interviewees identified an interesting data literacy ecosystem in the country, and possible technical partners with whom we could reach smaller cities. This would facilitate the goal of creating a virtuous cycle of social transformation by working with vulnerable populations coming from non developed regions.

c. SDGs and area of focus

The goal of the visualizations and the interviews was ultimately to shed insights on which SDGs, or broader development topics, the deployment of EmpoderaData should focus on, for the selected country. Suggested SDGs of focus, include gender equality (SDG 5, 10), social progress / equality (SDG 1, SDG 10), cities (SDG 11), climate change (SDG 13). At this stage, further research is still needed to define the SDG(s) for Mexico, which is one of the goals of the workshop to be conducted in the University of Manchester on October 16th, 2019.

The criteria and considerations to be taken into account for selecting the SDG(s) include:

- Topic / areas of (most) relevance to the country, which can be defined in terms of main data gaps (SDGs measurement) or areas where country is falling behind (SDGs achievement).
- In the selection, it will need to be ensured that:
  - ☐ Sufficient data is available to work on the topic.
  - ☐ Sufficient "host organizations" available to host fellows working on this topic.
- An additional consideration that may be taken into account is donors interest and potential opportunities for funding around this SDG.

d. Takeaways and next steps

The next phase of the EmpoderaData project consists in building the structure of the fellowship, including modality, scope of training, audience, as well as consolidating partners for training delivery and host organizations. The main takeaways to inform next phase are:

- **Audience:** based on inputs from interviews, one of the most relevant target audience seem to be (young) professionals, from civil society and public sector in particular. Criteria for selection should include:
  - ☐ affirmative action to reach and include traditionally excluded subgroups of the population,
  - ☐ professionals with potential to generate change and become 'multiplicators' of knowledge

- **Scope of training:**

  ☐ The biggest training need is for basic skills such as basic statistics, basic analysis skills and basic methodological skills.

  ☐ The training curriculum should be based on non proprietary-software to ensure sustainability and facilitate access.

- **Modality of the 'Data Literacy and SDGs' (name tbd) fellowship:**

  ☐ Training + paid fellowship model considered a promising set-up in the three countries.

      ° If target audience is confirmed as young professionals, the 'training' phase need to be compatible with professional activity hours (for e.g. night or week-end shifts).

  ☐ Considering the need to build data literacy capacities towards fostering 'hybrid' professionals that can understand, use and analyze data for 'social sciences' purposes, it could be relevant to consider a fellowship in 'duo', in which one 'data scientist' and one 'social scientist' would be paired at the same host organization to undertake together one 'data for social good' project. The goal will be to leverage their expertise while promoting the consolidation of the 'new' skills learnt during the training phase. In this set-up, separate trainings for 'data scientists' and 'social scientists' may be considered.

Immediate next steps: at the EmpoderaData workshop, on 16 October 2019, at the University of Manchester, the project consortium will gather, along with potential partners, to work on the above, decide on SDG to focus on and consolidate the core partnerships. and ideas for a funding bid. Alongside building a broad project proposal for the deployment of EmpoderaData in Mexico, the project consortium will keep undertaking research on data literacy, i.e. a literature review.

BIBLIOGRAPHY

Analysis and diagnostic of the current situation in Colombia structured around six dimensions, Data-Pop Alliance. 2019. Available online: https://datapopalliance.org/colombia-big-data-strategy/

"Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data". Rahul Bhargava, Erica Deahl, Emmanuel Letouzé, Amanda Noonan, David Sangokoya, and Natalie Shoup with guidance from Mark Frohardt, William Hoffman, Alex "Sandy" Pentland, and Alessia Lefébure. September 2015. Available online: https://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data/

Education at a Glance: Colombia OECD Indicators (OECD, 2019). Available online: http://gpseducation.oecd.org/Content/EAGCountryNotes/COL.pdf

Education at a Glance: Mexico OECD Indicators (OECD, 2019). Available online: http://gpseducation.oecd.org/Content/EAGCountryNotes/MEX.pdf

Regions and cities at a Glance (OECD 2018). Available online: https://www.oecd-ilibrary.org/sites/reg_cit_glance-2018-en/index.html?itemId=/content/publication/reg_cit_glance-2018-en

The Global Gender Gap Report 2018. World Economic Forum. Available online: https://www.weforum.org/reports/the-global-gender-gap-report-2018

ODS – Metas Nacionais dos Objetivos de Desenvolvimento Sustentável, IPEA, 2018. Available online: http://www.ipea.gov.br/portal/images/stories/PDFs/livros/livros/180801_ods_metas_nac_dos_obj_de_desenv_susten_propos_de_adequa.pdf

PISA 2015 results (volume I). OECD. Available online: https://read.oecd-ilibrary.org/education/pisa-2015-results-volume-i_9789264266490-en#page1

PISA 2015 results (volume II). OECD. Available online: https://read.oecd-ilibrary.org/education/pisa-2015-results-volume-i_9789264266490-en#page1

PISA 2015 Technical Report. OECD. Available online: PISA 2015 results (volume II). OECD. Available online:
https://read.oecd-ilibrary.org/education/pisa-2015-results-volume-i_9789264266490-en#page1

Síntese de Indicadores Sociais: indicadores apontam aumento da pobreza entre 2016 e 2017, Agência IBGE Notícias. Editoria: Estatísticas Sociais, 05/12/2018. Available online: https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/23298-sintese-de-indicadores-sociais-indicadores-apontam-aumento-da-pobreza-entre-2016-e-2017

The University of Manchester Sustainable Development Goals. 2019. Available online: http://documents.manchester.ac.uk/display.aspx?DocID=43184

A World that Counts, Mobilising the Data Revolution for sustainable Development,
Report prepared at the request of the United Nations Secretary-General by the Independent Expert Advisory Group on a Data Revolution. November 2014. Available online: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf