



Workforce Wanted

DATA TALENT FOR SOCIAL IMPACT

data.org



Patrick J McGovern
FOUNDATION

Dalberg

CONTENTS

| | |
|---|-----------|
| Acknowledgements..... | 3 |
| Foreword..... | 4 |
| Abbreviations..... | 7 |
| Executive Summary..... | 8 |
| Chapter 1: The Emerging Field Of Data for Social Impact..... | 18 |
| Chapter 2: The Talent Landscape..... | 24 |
| Chapter 3: Four Pathways to Expand DSI Talent..... | 32 |
| Pathway 1: New talent..... | 34 |
| Pathway 2: Existing social sector talent..... | 43 |
| Pathway 3: Transitional talent..... | 50 |
| Pathway 4: Leadership..... | 54 |
| Emerging opportunities across the DSI ecosystem..... | 60 |
| Chapter 4: Recommendations to Advance the Field..... | 62 |
| Annexes..... | 67 |
| Definitions..... | 68 |
| Literature review..... | 71 |
| Analysis of existing data talent for social impact initiatives..... | 73 |
| Market sizing methodology..... | 75 |
| Inclusion, Diversity, Equity, Accessibility (IDEA)..... | 76 |
| List of Interviews..... | 79 |
| Bibliography..... | 82 |

ACKNOWLEDGEMENTS

This report would not be possible without the contributions of the experts and ecosystem actors who shared their expertise through interviews. We are grateful for the time and effort they gave to this project, as well as for the work they do every day on behalf of the communities they serve and the systems they are trying to change.

In particular, we would like to thank Acumen Academy, the African Center of Excellence in Data Science (ACEDS), the African Institute of Mathematical Sciences (AIMS), the Amani Institute, Andela, Dasra Social Impact Leadership Program, DataKind, Educate Girls, eMobilis, Girl Effect, GiveDirectly, Harambee Youth Employment Accelerator, IFC, the India Leaders for Social Sector (ILSS), J-PAL, Laboratoria, the Mastercard Center for Inclusive Growth, Molengeek, Moringa School, Mozilla Foundation, the Namibia University of Science & Technology, NetHope, One Acre Fund, the Rainforest Alliance, Salesforce, Skoll World Forum, SoCieDat, Splunk Ventures, Strathmore University, Tableau Foundation, Talent Rewire, Tech Change, Tech4Good Community, the World Bank, and experts including Frank Mccosker and Lutz Ziob.

To the Project Team, including Dalberg, data.org, and the Patrick J. McGovern Foundation, that pushed this work forward, interpreted insights, and contributed to this final report: thank you.

Thank you to our partners in the data for social impact ecosystem for conducting qualitative and quantitative research that is the foundation of this report's insights.

On behalf of data.org and the Patrick J. McGovern Foundation, we look forward to continuing to work with all our partners to make our shared recommendations the expectation in the field.

FOREWORD



Data is everywhere; but how social impact organizations tackling the most complex problems of our day effectively unlock its potential remains a stubborn challenge worthy of our collective action.

Undoubtedly, data is an essential tool to better solve important global problems like climate change, healthcare disparities, food insecurity, humanitarian emergency response, lack of financial inclusion, and systemic discrimination in all its forms. To unleash the power of data across the social impact sector, we must advance how we identify, cultivate, support, and retain diverse data professionals.

Easier said than done. There is a persistent workforce shortage, with equally persistent barriers that exclude women and people of color from the field. Data infrastructure within social impact organizations is often underdeveloped due to lack of meaningful investment from philanthropy and organizations themselves. At the same time, competition is fierce for the skills needed to better use data, as sectors of every kind are competing to attract, upskill, and retain data talent.

Determination to overcome these challenges, coupled with the exciting opportunity to transform how the social impact sector uses data, is why data.org and the Patrick J. McGovern Foundation partnered on a landscape analysis. We ventured into this work to better understand the current data talent training landscape within the social impact sector, uncover opportunities for collective action, and develop a set of recommendations for global partners across academia, philanthropy, government, social impact, and the tech for good sector to build a diverse and inclusive data talent ecosystem. We have a unique chance to build pipelines to opportunity and impact that are filled with talented individuals representative of the societies and communities that have previously been excluded.

Spoiler alert: to be effective, the solutions and recommendations uncovered require all of us to act in concert.

Additional learning is needed, capital is required, and an authentic and unwavering commitment to building a diverse and inclusive data talent pool is essential. We remain energized by the work ahead. There are innovative and replicable ways that organizations that tackle entrenched problems are growing data talent. New talent is emerging, and they are eager to apply their skills and expertise to problem statements that matter. Talent within social impact is looking for pathways to their data skills. Finally, and perhaps most critically, leaders are beginning to understand the potential they can unlock if their organizations use data better to achieve impact.

That is where you come in. The purpose of this report is to shed light on the scale of the opportunity in front of us all, better understand its dimensions, and catalyze a wide array of people and institutions to work together and be deliberate about how we build a diverse data talent pool. Together, we can create an ecosystem that connects the right skills to the right problems and supports people across their careers to develop and strengthen the diverse skills necessary to effectively establish the field of data for social impact.

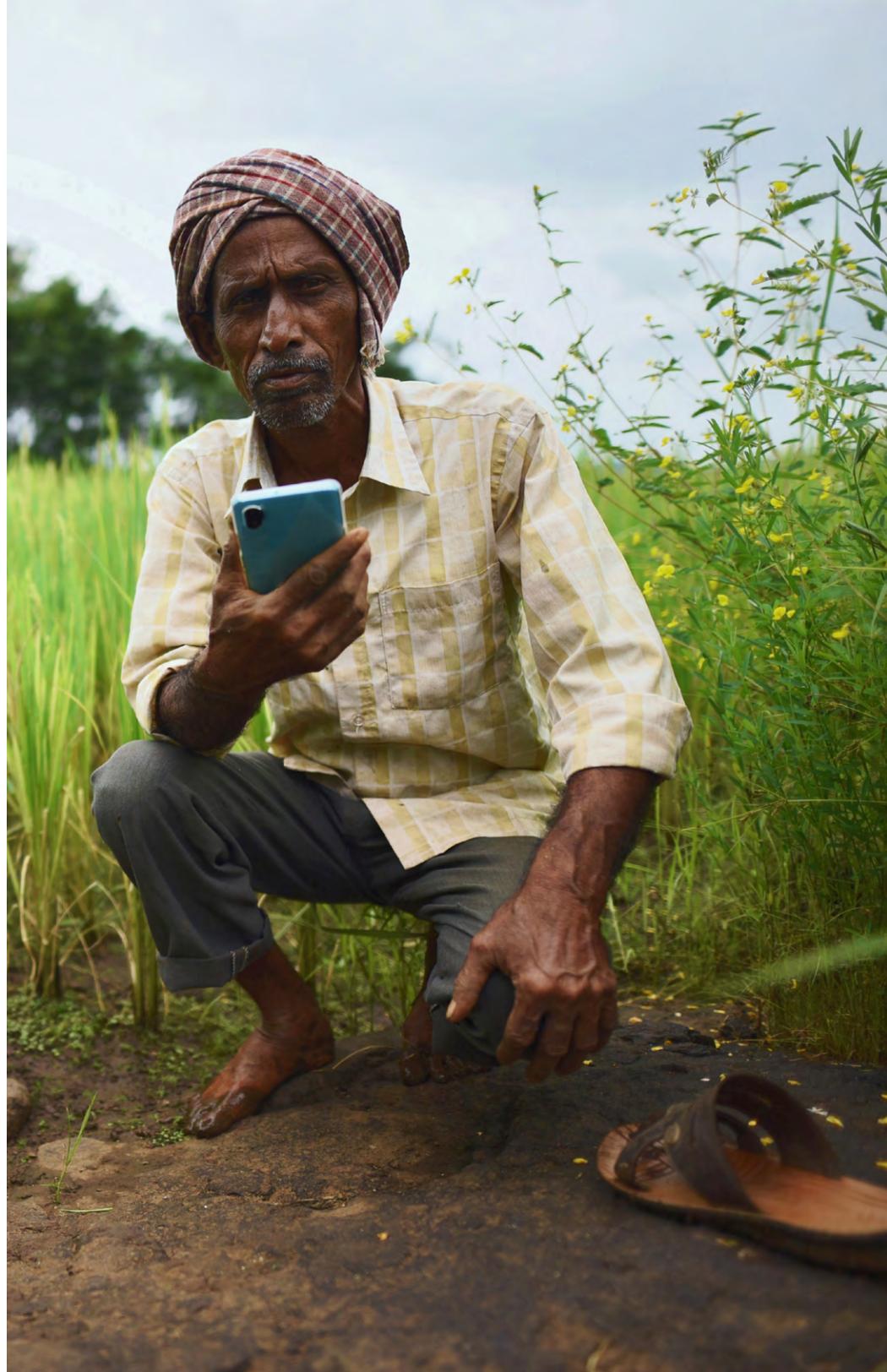
Claudia Juech
Vice President of Data and Society
Patrick J. McGovern Foundation

Ginger Zielinskie
Chief Strategy Officer
data.org

June 2022

ABBREVIATIONS

| | |
|-------------|--|
| AI | artificial intelligence |
| BSc | Bachelor of Science |
| CEO | chief executive officer |
| CIO | chief information officer |
| CTO | chief technology officer |
| DSI | data for social impact |
| DQM | data quality management |
| GDP | gross domestic product |
| HR | human resources |
| IDEA | inclusion, diversity, equity, and accessibility |
| MEL | measurement, evaluation, and learning |
| MOOC | massive open online course |
| MSc | Master of Science |
| SIO | social impact organization |
| STEM | science, technology, engineering, and mathematics |
| WEF | World Economic Forum |



EXECUTIVE SUMMARY





This is a complex and pivotal moment for harnessing the global impact of data. On the one hand, we see tremendous innovation across public, private, and social sectors, and broad recognition of the critical role data is playing in our daily lives. On the other hand, the industry struggles to understand, unlock, and keep pace with demand—all while navigating serious questions about who is considered when gathering, shaping, and using data, and how we ensure that the data field is enriching human lives and livelihoods, not diminishing them. Lack of resources, technical capacity, and data governance hamper the production of useful data for public policy.¹

Amid this complexity, data.org and the Patrick J. McGovern Foundation (PJMF) believe that there is significant untapped potential to shape the field of data more broadly to tackle the world's most pressing issues in developing contexts.

Quality talent is central to this goal. data.org and PJMF recognize that building the next generation of diverse data talent for social impact is essential for us to effectively address social challenges. This first-of-its-kind report focuses on understanding what is needed to cultivate a diverse, equitable, and inclusive talent base of data professionals that are driving social impact in developing contexts around the world.

¹ <https://www.worldbank.org/en/publication/wdr2021>

This study aims to do three things:

- 1.** Bring visibility to an emerging pool of talent: data professionals focused on social impact in developing contexts;
- 2.** Explore the potential to accelerate this labor market segment, particularly when it comes to inclusion, diversity, equity, and accessibility (IDEA); and
- 3.** Offer recommendations for like-minded efforts to dramatically grow and expand access to purpose-driven data professionals around the world.

To achieve these goals, this report focuses on a few key questions:

- What do we know about the current talent landscape and the challenges it faces—particularly in developing contexts, and with an emphasis on inclusion, diversity, equity and accessibility?
- What are the immediate-to-big-picture opportunities for this space?
- What are the pathways for attracting, building, and supporting data professionals focused on social impact? What systemic challenges do the data field, and data for social impact (DSI) ecosystems face? What support, enhancement, or new solutions could accelerate innovation, growth, and scale?
- How can we align interests and develop recommendations for a broad range of committed actors who will foster the growth and expansion of inclusive, diverse, equitable, and accessible data talent labor markets around the world?

KEY FINDINGS OF THIS REPORT

Opportunity

We believe there is an opportunity to shape and develop 3.5 million data professionals focused on social impact in developing countries over the next 10 years. This opportunity is based on multiple factors and includes variation based on different scenarios.²

- The current landscape for purpose-driven data professionals is nascent in terms of its overall size and organization, but shows momentum and growth, fueled by a number of intrinsic and extrinsic factors—including increasing access to the internet and data itself; acceleration of digital transformation efforts around the world; recognition of the value of inclusion, diversity, equity, and access; acceleration of digitization forced by COVID-19; elevation of and investment in advancing social issues like the SDGs, climate, public health, etc.; exceedingly high demand for data professionals globally; and many others.

- Alignment of efforts, partnership, and resilient digital infrastructure is required. There are several complementary efforts that offer the potential to galvanize shared goals when it comes to unlocking DSI talent.
 - » The momentum and investment in supporting digital transformation strategies around the world—particularly in low- and middle-income countries (LMICs)—offers opportunity to further advance structural investments in shaping labor markets for data-driven skills and professionals.
 - » The advancement of research institutions and academia in understanding the opportunities and realities of IDEA in the data ecosystem can help push the data field, including DSI, toward a systematic emphasis on IDEA.
 - » Digital transformation efforts focus on holistic, meaningful connectivity solutions (including cost of data, devices, and enabling infrastructure such as identity, payments, and asset registries), which are critical to addressing the digital divide and unlocking meaningful access to opportunities.

² In sizing the data talent opportunity, we are considering potential, i.e. the possibility of new jobs to be added to the current trajectory, rather than a current employment gap. In order to reach this potential—whereby the percentage of the workforce in the social sector in developing contexts that occupies data roles equals that of developed contexts — it will be necessary also to stimulate the demand for DSI talent. We recognize that these analyses have several limitations, e.g. attempting to size the opportunity across “developing contexts” and for the “entire social sector” within a nascent/emerging field. Most developing countries’ labor market data are limited and incomplete, and data are not aggregated across countries. Moreover, there is no common taxonomy of data professionals across countries. However, we believe that this analysis is critical to perform with the evidence available to us.

The opportunities highlighted above can be realized through the high-level prioritization of data, including long-term financing, investments in human capital, and laws conducive to the safe production, exchange, and use of data. Some investments in better data have paid for themselves.

Capacity

There are a number of ways in which organizations can access and build data skills, teams, and organizational strategy.

- As a framework for considering the different approaches to building and growing skills and organizational strength, we identified four pathways:
 - » *New talent.* Expanding exposure of learners through development of DSI use cases; integration of hands-on, practical learning; incorporation of applied learning into curriculum; and stronger alignment of training models with the needs and demands of the social impact sector.
 - » *Existing talent.* Models for upskilling and reskilling—such as in-house, outsourcing, and sponsorship

models—that recognize the value of existing talent committed to social impact and SIOs.

- » *Transitional talent.* Greater exposure and access to opportunities that allow for more agile flow of talent across sectors; examples include hands-on fellowships, short courses, volunteer opportunities, and rotational leadership programs.
- » *Leadership.* Enhancing and shaping new models to support design, experimentation, and advancement of data-driven strategies, initiatives, and talent acquisition; investment in allies, such as boards and funders, to advance understanding of data-driven solutions.

The DSI field is competing within the broader data skills ecosystem, meaning that DSI professionals are often disincentivized to choose the social impact pathway, particularly when considering public-private wage gaps and limited visibility into career growth opportunities.

A review of nearly 200 data talent initiatives, a literature review of approximately 90 articles and reports, and expert interviews with more than 30 leaders in the field suggest that training and talent initiatives struggle with several systemic issues, including low levels of organizational awareness of how data can be valuable and a need for increased, sustained financing to drive shared growth, as well as:

- *Limited capacity of traditional institutions.* Traditional education programs—specifically, university science, technology, engineering, and mathematics (STEM) programs—are insufficient in terms of both the number of institutions and volume of qualified data professionals; they also lack social impact orientation.
- *Proliferation of non-traditional training models, including massive open online courses (MOOCs) and other online training platforms, lack evidence of efficacy.* Program outcomes are often disconnected from longer-term results such as job placements, and many programs demonstrate a bias towards technical training rather than integrated translational skills and work readiness.
- *Mid- and senior-level talent as both a gap and driver of growth.* Intermediate and advanced skills are underserved relative to the need for growing the talent base; training or bringing in mid- and senior-level data talent can have a multiplier effect based on leadership’s ability to shape ecosystems.
- *Ecosystem constraints.* The ability of training programs to adjust their business models is hindered by a lack of accurate market demand data for skills. While there is a growing appreciation of the value of data for social impact across public, private, and social sectors, the understanding of demand and sourcing of talent remains limited.
- *Need for leadership programs that focus on supporting leaders’ evolution rather than one-off interventions.* Leadership programs and fellowships often focus on individuals rather than holistic interventions that affect the broader data ecosystem; lessons of leadership programs are not embedded into the existing activities of the professional environment through workplace experiments and nudges.



Diversity

- While the values of inclusion, diversity, equity, and accessibility (IDEA) are generally acknowledged as important, these values have the potential to be more deliberately and comprehensively embedded in the data field today, in order to ensure responsible ethical advancement and impact.
- There may be more in common than not when it comes to DSI in “developing contexts.” The use of the phrase developing contexts, rather than developing economies, recognizes that the characteristics of developed and developing economies may not be highly variable when it comes to DSI. While economic structure may be quite different at an aggregated national level, the high levels of variation led us to push for a more granular perspective on where and how DSI talent could be unlocked. This terminology also recognizes that many contexts share common characteristics and constraints (e.g., advancing diversity, competition with private sector for skills, etc.), and that countries often have a wide range of social, economic, technological, and structural realities. As a result, while “developing context” is difficult to cleanly delineate, it signals that our work attempted to focus not on geographic or economic delineations, but rather on countries, regions, sub-regions, or other “contexts” where the infrastructure, social, or economic context is still emerging.

- Based on the findings of the research, a few recommendations emerged for actors interested in advancing the data for social impact space:
 1. *Experiment early and evaluate often.* The nature of a nascent field requires actions that will crowd in others, draw attention to what works, quickly demonstrate limitations, and facilitate frequent pivots.
 2. *Prioritize inclusion, diversity, equity, and accessibility* when considering access to education and training, links between training and placement, and absorptive capacity of maturing data ecosystems (organizations and beyond).
 3. *Recognize the interdisciplinary nature of data for social impact,* where the depth of technological understanding and expertise is matched with the discipline and understanding of social sciences.
 4. *Move from individuals to ecosystems.* Recognize the role an individual leader plays within an organization, an industry or sector, or a broader ecosystem, and align efforts and investments accordingly. Recognize the individual incentive systems already in play and the potential tensions that may exist when seeking to build new data-driven strategies or decision-making processes.
 5. *Invest in applied learning and stronger links to professional placement and advancement,* shifting from a focus on “the number of people trained” as a critical result to “the number of people playing an active role addressing social issues and working within organizations.” Consider aligning funding with intended outcomes. Financing mechanisms that link training to sustainable employment could nudge the sector in an impact-focused direction.
 6. *Coordinate complementary efforts.* Looking across various efforts linked to data, digital transformation, and the advancement of data-driven strategies for nonprofit or social-impact-oriented organizations we see significant opportunities for greater coordination to advance DSI as a field.
 7. *Continuously invest in more and better visibility through data* to illustrate how the ecosystem is functioning—in order to build on what works, better understand gaps, and track the many factors that influence outcomes.

- Because this is a new sector, we *recognize the limitations of this research and the broader ecosystem constraints*. As we developed this report, we were conscious of several challenges in mapping, sizing, and analyzing this emergent field. These limitations include:

- » Availability of granular datasets that disentangle *information* linked to the positioning and movement of talent, particularly in developing contexts, across the social sector and with regard to advancing inclusion, diversity, equity, and access.

- » *Difficulty defining “social impact.”* We know it is not simply a legal construct, such as a nonprofit organization, and that social impact manifests across all sectors, including government and the private sector. However, as a starting point, where possible, we used the nonprofit sector as a proxy for a “social impact” focus.

- » *Lack of a common language* to identify necessary data skills and roles across the public, private, and social sectors.

- » *Lack of visibility* into current and future demand, particularly latent demand.

Recognizing these constraints, we have established a set of principles to guide the analysis:

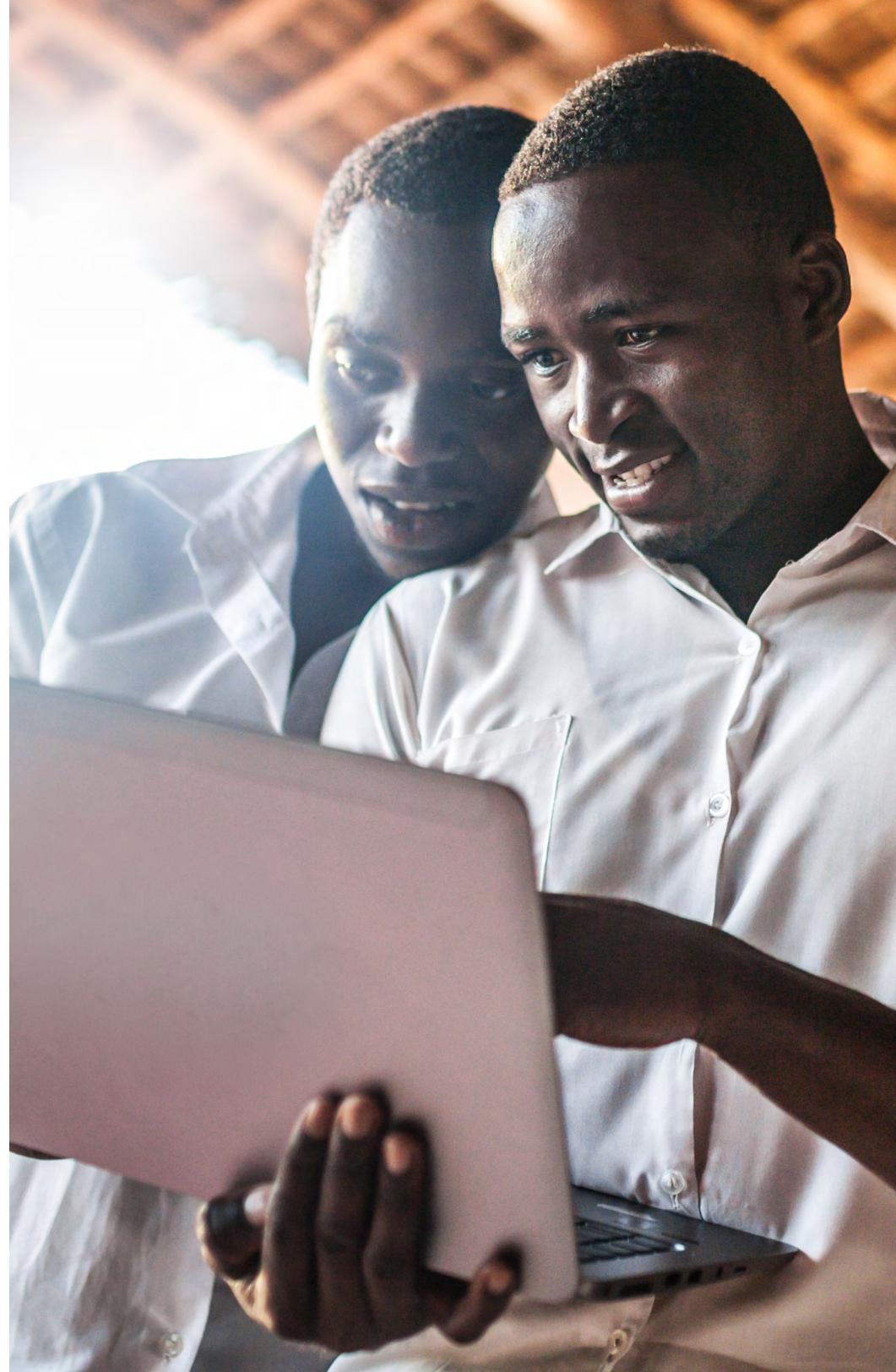


FIG 1 » GUIDING PRINCIPLES FOR ANALYZING THE TALENT LANDSCAPE

- » Reflecting on the important interplay between supply (number of DSI professionals) and demand (e.g., Social Impact Organizations' (SIOs') absorption capacity) in conjunction with limited data availability. Therefore, we use data maturity stages as a proxy for understanding demand
- » Taking a long-term view on building the DSI field, and using a portfolio perspective that combines scaling proven approaches, with active experimentation in order to iteratively learn and advance a strategy that works
- » Recognizing an overarching commitment and opportunity to build and scale a diverse community of DSI professionals by acknowledging deep systemic challenges that influence existing labor markets, as well as targeting opportunities to design new solutions that drive greater and more equitable access for all pathways
- » Anchoring on SIO leadership and senior level talent challenges and opportunities, in order to unlock the size of opportunity and build the field
- » Identifying and learning from outlier organizations (bright spots), that have managed to overcome key challenges in this space, and from analogous sectors that have faced similar barriers
- » Considering the direct and indirect results of investments, particularly as it allows for a multiplier effect on results and/or deeper, more systemic shifts

Because our intention is to grow and shape an emerging field, we believe that, in defining the talent base for DSI, we are offering a starting point from which others will enhance, build, and deepen the research. As one interviewee shared with us,

“We will never know as little about DSI as we do today; that’s the place to build from.” We look forward to the improvements and advancements that other stakeholders will develop in coming years.

1 THE EMERGING FIELD OF DATA FOR SOCIAL IMPACT



As the data field has progressed over the past decade, it has transformed and driven innovation across a wide range of industries around the world. With the aid of vast computational power, new ways of analyzing giant datasets have generated advanced predictive models and other forms of machine learning that have benefited a broad spectrum of industries. In the finance industry, for example, data capabilities have transformed both the customer experience and the way companies detect and prevent fraud.³ Across many industries, small and medium-sized businesses are able to collect data from a range of sources (e.g., loyalty programs, social media engagement, purchase history) to gain a 360-degree view of customers that enables significantly faster and more accurate decision making.⁴ And in healthcare, machine learning is driving rapid, low-cost, and accurate diagnostics of diseases.

The evolution of the data field is also informing and amplifying positive social impact on the world's most pressing challenges. There are three main ways in which the data field is helping social impact organizations (SIOs) to amplify their efforts: (i) gaining a better understanding of the broader context in which the organization works, (ii) strengthening operational efficiency and effectiveness, and (iii) designing better solutions to serve beneficiaries and address social issues. Harambee has built one of the largest datasets of South Africa's 1.4 million unemployed youth, which deepens understanding of

unemployment challenges in the country. Educate Girls developed a machine learning model that uses publicly available information, like census data, to predict the number of out-of-school girls in each village in India, which allowed it to reach between 50 and 100% more out-of-school girls for the same budget. Girl Effect leveraged advances in natural language understanding (NLU) to develop a chatbot that enables young girls in developing countries to anonymously write questions about relationships and sexual health. GiveDirectly has partnered with the government of Togo to use machine learning and mobile phone data to remotely identify, enroll, and pay households in poverty. These are just a few examples, and the library of case studies continues to grow.

While the significance of data for the global economy is uncontested, the potential of data to transform the work of social impact organizations is far from being realized.

There is consensus that data is important for advancing the global ecosystem. Data shapes how we observe, reason, and act. Tableau and other global organizations working on skills of the future commonly cite data analytics as a new basic skill for the 21st century. Data use in the private and public sector is transforming the work of businesses and organizations all over the world. Data is a distinctive driver of economic growth and AI is projected to contribute USD 15 trillion to the world's economy by 2030. Yet significant potential remains for data

³ 'How machine learning is transforming healthcare, finance and agriculture'; Mahbulul Alam; 2021

⁴ 'How big data is transforming industries in big ways' 3 Pillar Global, 2021



to transform the social impact sector, including by boosting productivity, providing greater insight into the context in which SIOs work, and supporting more effective measurement of results.

Overall, data for social impact (DSI) is a nascent field that seeks to address this untapped potential. DSI refers to the use of data to more equitably and effectively benefit people, communities, organizations, and the environment. Social and environmental flourishing is the prime objective, rather than deriving profit.

However, the ability to maximize the potential of data varies greatly across contexts. This variability extends to the adoption of a data focus between SIOs and the private sector. While studies agree that data skills are increasingly relevant across sectors, most studies are focused on private sector context in developed countries. Very few studies seek to establish and quantify the potential of data in a social impact context, especially in developing countries. Understanding the potential for DSI also requires an understanding of the ecosystem within which it operates. A number of factors contribute to the ecosystem, including underlying industry and labor market

structure, the dynamics of supply and demand, and characteristics of an early-stage emerging field, as well as policy and financing.⁵

A lack of common standards and shared language makes it difficult to understand the state of data for social impact. The speed at which the data field is evolving makes defining a typology of data professionals challenging. Also lacking is a commonly agreed-upon framework for assessing and measuring data maturity, as well as a clear picture of sector-wide demand. While more data is needed, organizations typically fit into one of three categories:

1. Innovators and trailblazers in the application of data for organizational impact have and are advancing their own data-driven strategies and have built teams and solutions that anchor on the power of data. These innovative SIOs have relatively advanced levels of data analytics capabilities but may face operational challenges and uncertainty about how to scale. Often, they use historical data to create predictive or prescriptive models to assess potential outcomes. In addition, they may be leveraging artificial intelligence (AI) or machine learning systems to discern patterns from unstructured data.
2. Emerging and evolving organizations regularly use historical data to explain past performance and discern

trends that have bearing on current decision-making.

3. Early stage of data maturity are organizations that are not yet aware of the full potential of data to amplify their impact. They may engage in data collection and analysis on an as-needed basis, and not as part of regular operations.

Centering on inclusion, diversity, equity, and accessibility (IDEA) in the DSI field has the potential to amplify impact and reduce the harm that is currently caused by exclusion. We define the values of IDEA as creating the conditions for belonging so all individuals are respected and valued and can participate fully in decision making. This ensures engagement of a broad and representative community of stakeholders, partners, and allies and promotes fairness, justice, and transparency in opportunity and distribution of resources. When we create opportunities and welcoming spaces for engagement, including by providing assistive technologies and devices, we see an increase in participatory decision making and shared outcomes.

Strengthening IDEA in the DSI space leads to outsized benefits for society. When the “people behind the screen” understand—or, better yet, have lived experience with—the challenges faced by the beneficiaries of their work, they are likely to design more effective analytical solutions that lead to greater impact. Focusing on IDEA values is also the right thing to do.

⁵ For additional reflections on characteristics of an emerging field, see: How do you build a field? Lessons from Public Health, Rockefeller Foundation. 2015

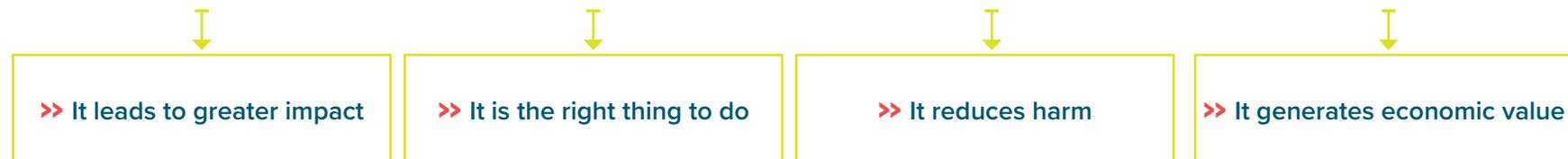
Reducing the exclusion of underrepresented groups in DSI jobs is imperative, especially given the nature of these jobs, such as those that are higher paying with a significant opportunity for further growth versus other careers.

Doing so can also reduce harm, as excluding the needs and nuances of target populations from the design and implementation of data solutions can reinforce harmful biases on a grand scale. The lack of representation and diversity of data talent has been shown to drive ignorant—or worse, malicious—algorithms and reinforce bias at a much larger scale than what was previously possible. These biases often stem from the individuals designing the algorithms and the biases of the data that are fed into models. A now-discontinued

recruitment software used by Amazon, for example, consistently favored male candidates over women.

Improving IDEA in the DSI space also presents an opportunity to generate greater economic impact. For example, improving ethnic and gender diversity in the U.S. technology workforce could create as much as USD 570 billion in new value for the tech industry and could add 1.2 to 1.6% to national GDP.⁶ This is not surprising, as teams and workforces with greater gender diversity continuously outperform male dominated workforces.⁷ On an even larger scale, closing the gender gap in the workforce—by addressing time poverty, digital inclusion, financial access, and education—could add USD 28 trillion to the global GDP.⁸

FIG 2 » SUMMARY OF THE ARGUMENTS FOR CENTERING IDEA⁹



Realizing the potential of data in the social sector requires investment across four main areas: talent, organizational change, sector evolution, and ecosystem growth (Figure 3).

While recognizing the complex interactions of these different factors, this study focuses on the supply of data talent.

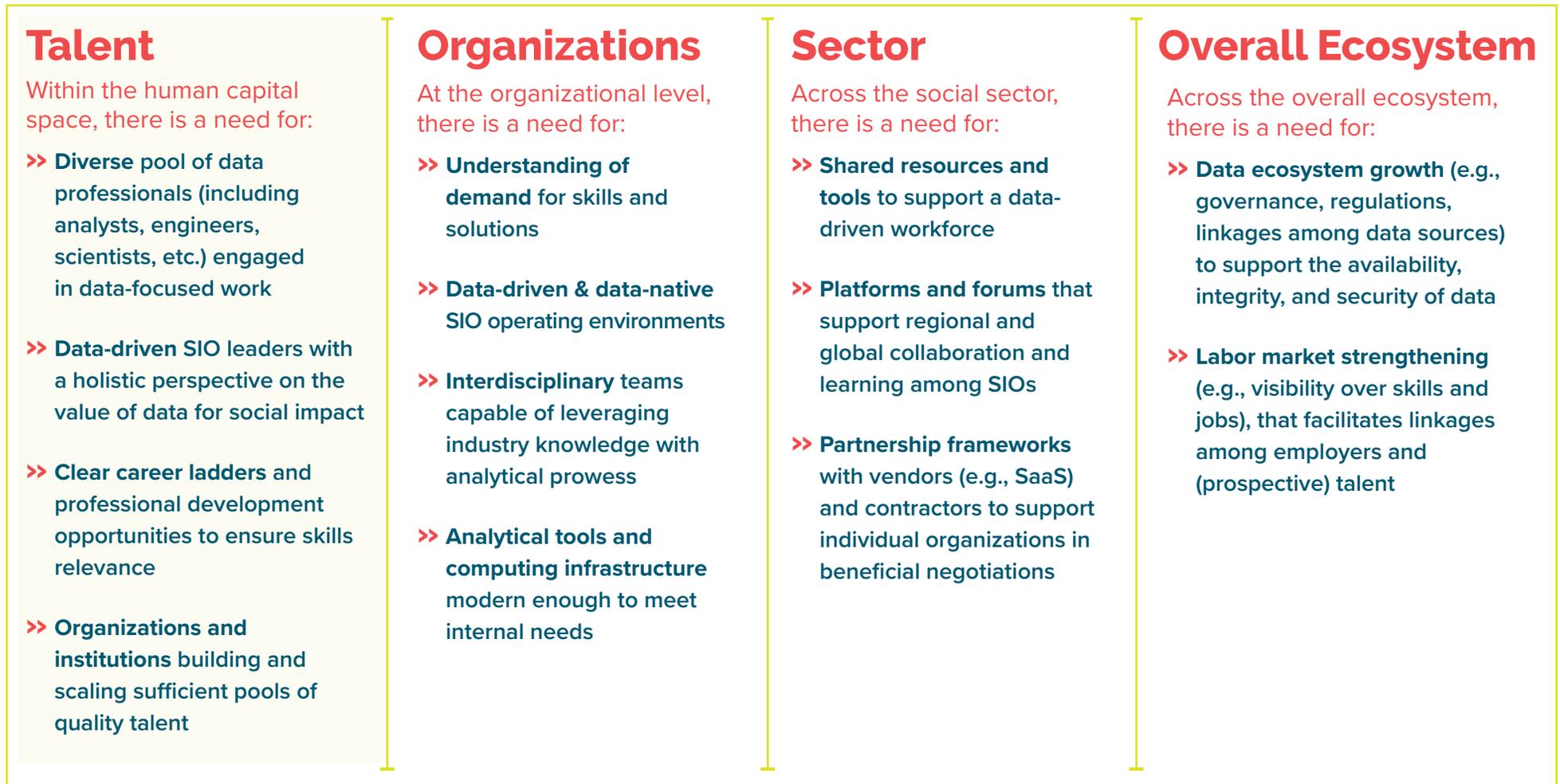
6 Dalberg/Intel – Decoding Diversity 2016 (US Tech workforce).

7 Bloomberg Gender Equality Index.

8 Centre on Foreign Relations.

9 Intel and Dalberg, Decoding diversity: the financial and economic returns of diversity in tech2016; Every Learner Everywhere, What are the risks of algorithmic bias in higher education?, 2021.

FIG 3 » **KEY COMPONENTS OF A DATA FOR SOCIAL IMPACT ECOSYSTEM**



↑ **OUR FOCUS FOR THIS STUDY**

2 THE TALENT LANDSCAPE



Data skills include a broad range of capabilities—from basic data collection and data cleaning to analytics and data strategy. Recognizing the diversity of these skills and their application across a data value chain, we lay out a high-level taxonomy of data skills (Figure 4) under the rubrics of basic data skills, intermediate data skills, and advanced data skills.

Based on discussions with experts in the DSI field and other relevant fields, this report focuses on the roles of intermediate and advanced data skills as well as leadership as a critical driver of growth across SIOs. The focus on more advanced skills recognizes the importance of meeting an immediate demand for skills while also creating space for the growth of more junior talent over time. In addition, because of the unique role SIO leaders play in pushing organizations to use data more powerfully and effectively, and in driving demand for data talent, they hold direct responsibility and accountability for building and growing data teams. Finally, SIO leaders have the agency not only to build skills within their organizations, but also to push for systemic change in the sector.



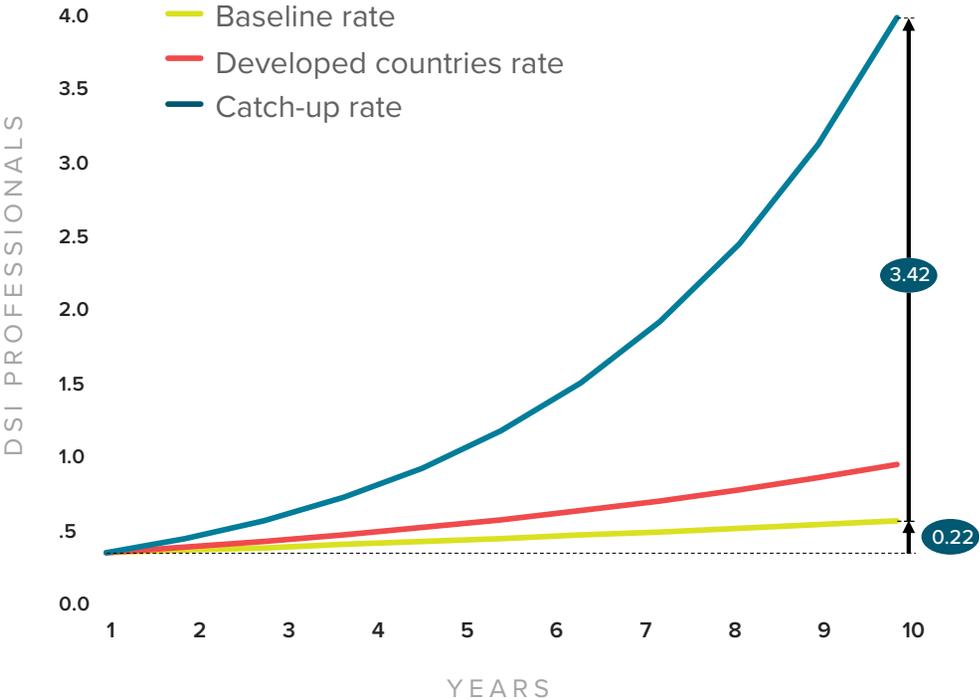
FIG 4 » **TAXONOMY OF DATA SKILLS** (Non-exhaustive and highly dependent on roles and organizations)

| | Examples of technical skills | Examples of translational skills | Examples of Roles |
|--------------------------|---|---|--|
| Basic data skills | <ul style="list-style-type: none"> >> Basic use of statistical software (e.g., Excel) >> Data collection and entry >> Basic data analysis and manipulation >> Basic data visualization | <ul style="list-style-type: none"> >> Communication (verbal/written) >> Ability to interpret and understand key takeaways from data >> Collaboration and teamwork >> Critical thinking, problem-solving >> Data-driven decision-making | <ul style="list-style-type: none"> >> All non-data specialist professionals who use data in decision-making >> Data entry roles >> Data collection officer |
| Intermediate data skills | <ul style="list-style-type: none"> >> Ability to perform complex functions (e.g., create and maintain complex spreadsheets on Excel; using Stata) >> Some programming languages (R, SQL, Python) >> Data quality management (DQM) >> More complex data visualization >> Some knowledge of machine learning >> Statistical knowledge (e.g., sampling techniques, ability to design research tools) | <ul style="list-style-type: none"> >> Ability to understand the nuances and takeaways of data analysis >> Ability to communicate results in the organization and to the public >> Data presentation skills >> Critical and structured thinking >> Continuous learning >> Ability to work in teams | <ul style="list-style-type: none"> >> Research analyst >> Data analyst >> Data associate >> M&E associate/officer >> Database administrator >> Data coordinator >> Cloud operations associate |
| Advanced data skills | <ul style="list-style-type: none"> >> AI and machine learning >> Deep learning >> Advanced data analytics and modeling >> Predictive analytics >> Advanced data visualization >> Advanced cloud computing and engineering skills >> Deep theoretical knowledge of statistics >> Ability to plan and/or manage the entire data lifecycle | <ul style="list-style-type: none"> >> Data strategy (developing and leading the implementation of strategies) >> Ability to work with and lead data teams >> Data stewardship >> Data ethics >> Storytelling skills >> Critical and structured thinking | <ul style="list-style-type: none"> >> Data scientist >> Machine learning engineer >> AI engineer >> Cloud computing engineer >> Data engineer >> Data architect >> Head of research/analytics |

There is a significant opportunity to create and support a talent pool of over 3.5 million jobs in developing countries¹⁰ over the next 10 years. If the DSI field keeps growing at the same rate as all jobs in the formal sector, there will be roughly 220,000

DSI jobs to fill in developing contexts over the next 10 years. However, if the sector is stimulated and current ecosystem challenges are addressed, we can create and fill over fifteen times as many DSI jobs over this same period.

FIG 5» PROJECTED GROWTH IN DSI PROFESSIONALS



- >> If the DSI field keeps growing at the same rate as all jobs in the formal sector, there will be ~220,000 DSI jobs to fill in developing countries over the next 10 years
- >> However, if the sector is stimulated, this number could reach 3.5 Million DSI jobs over this same period
- >> It is important to recognize that in order to reach this potential (where the percentage of the workforce in the social sector in developing countries that occupies data roles, equals that of developed countries such as the UK), there is a need to stimulate demand for DSI talent

¹⁰ This exercise was specifically focused on developing countries (rather than developing contexts more broadly) in order to simplify our calculations and facilitate access to the required data points.

These analyses and estimates have several limitations. It is important to note that in sizing the data professionals' talent opportunity, we are considering *potential*, i.e., the possibility of new jobs to be added to the current trajectory, rather than a current employment gap. In order to reach this potential—whereby the percentage of the data workforce in the social sector in developing contexts equals that of developed contexts—it will be necessary to stimulate the *demand* for DSI talent.

DSI talent in developing contexts faces several challenges stemming from its position at the intersection of three areas: the data field, developing contexts, and the social impact sector. To begin with, the data field faces several complex challenges that persist across different geographies and sectors—most notably, an overall supply shortage of talent and a workforce characterized by a lack of many forms of diversity. For example, just 26% of data and AI professionals globally are women, and in the US just 3% of data and analytics professionals are Black Americans (compared to 12% of the US population overall).¹¹

Organizations that operate in developing contexts often face several challenges related to working within nascent data ecosystems and immature labor systems. Most developing countries' labor market data are limited and incomplete, and

data are not aggregated across countries. Moreover, there is no common taxonomy of data professionals across countries. However, we believe that this analysis is critical to perform with the evidence available to us.

SIOs in developing contexts also experience the challenges of working within the social sector. Compared to private sector companies, SIOs are often at a disadvantage in competing for talent, regardless of the types of talent and the geographies they operate in, including systemic wage disparities across the sector, financial instability, and limited awareness of or access to professional development opportunities.¹² In addition, SIOs in developing contexts also face sector-specific difficulties in hiring and retaining talent.

Seeding and growing DSI will be influenced by a range of trends, including the global scarcity of digital and data skills;¹³ the unique dynamics of the social sector and developing labor market contexts; and a recognition that building an inclusive, diverse, equitable, and accessible workforce is paramount. Global demand for data skills is fiercely competitive and rapidly changing—in fact, the World Economic Forum (WEF) Future of Jobs report found that data skills were the most in-demand skills globally in 2020.

11 WEF Global Gender Gap Report 2020; The Harnham Data & Analytics Diversity Report, 2020

12 Refer to Figure 8 for an example of wage disparities for data talent between the social and private sector.

13 McKinsey, Tech talent tectonics: Ten new realities for finding, keeping, and developing talent. April 2022.

The unique dynamics of the social sector will also heavily influence the DSI field's trajectory. Wage differentiation, variable data maturity of organizations, limited access to data and datasets, low levels of data usability, and lack of tech teams to build and grow the ecosystem are all contributing factors. At the same time, the growing demand for greater inclusion, diversity, equity, and accessibility of skills will need to be addressed at many different points along the journey of DSI talent.

Underpinning each of these trends are critical labor market drivers and dynamics that provide the foundation for DSI talent ecosystem.

We began our analysis by exploring the various approaches that organizations might take to identify, attract, retain, and build data skills and talent. We ultimately identified and explored four professional development pathways (Figure 7) through which talent can evolve in order to increase the number of data professionals in the social sector. The following sections of the report analyze the challenges embedded within each pathway that currently constrain access to and retention and utilization of DSI talent in developing contexts, as well as the distinctive examples where organizations have overcome these challenges. We further explore a range of ways in which stakeholders might potentially address these challenges in the future in order to accelerate the growth and improve the quality of the DSI talent pool.



FIG 6 » DATA FOR SOCIAL IMPACT TALENT: LABOR MARKET DYNAMIC

Supply challenges

Inadequate/insufficient supply of talent

Quantity

- » Insufficient number of data professionals overall, even less so when it comes to those seeking social impact pathways

Quality

- » Gaps between skills developed and the lived experience of applying those skills in the workplace

Relevance

- » Mismatch between the types of skills developed and those demanded by employers

Representation

- » Inadequate levels of diversity, inclusion, access, and equity when it comes to opportunities

Demand challenges

Mostly latent demand for data professionals

Awareness

- » Limited understanding of the potential of hiring diverse data talent and/or lack of willingness to invest in talent

Absorptive capacity

- » Limited ability to hire data professionals (human resources, financing, infrastructure)

Utilization and retention

- » Limited ability of SIOs to utilize and retain data professionals when they are hired

Limited number
of DSI
professionals
in developing
contexts

Ecosystem challenges

Nascent nature of the DSI field (i.e., limited use cases), nascent labor market systems (i.e., limited visibility over supply and demand, limited job linkages), and limited maturity of data ecosystems (i.e., data governance, overall data culture, etc.)

FIG 7 » PATHWAYS FOR ADVANCING DSI TALENT

Expanding supply of DSI professionals through existing and new talent

New talent

- » Potential DSI talent entering the workforce for the first time with no prior data work experience

Existing social sector talent

- » Current professionals in the social sector who could be upskilled in data

Transitional talent

- » Existing data professionals outside the social sector who could be upskilled or attracted to work in SIOs

Expanding the availability, retention, and utilization of DSI talent within SIOs

Increasing awareness and absorptive capacity of SIOs

Leadership

- » SIO leaders and senior executives who could acquire new capabilities and support to make their organizations more data-driven and to attract, retain, and utilize DSI professionals

3 FOUR PATHWAYS TO EXPAND DSI TALENT



Supporting the growth of DSI talent requires understanding the ways in which certain interconnected systems need to work more effectively.

Educational institutions need to prepare young people to enter data fields and a vibrant ecosystem of innovative training providers and business models needs to offer access to those skills. SIOs also need a more broadly supportive ecosystem in order to grow, and robust data needs to be collected for analysis in order to identify and describe evolving trends. Underpinning it all, sufficient and sustainable capital needs to flow in order to support access to and expansion of promising models.

Recognizing the dynamic interaction of these various systems, our approach focused on first exploring the different ways in which organizations access talent. This included identifying how data strategies emerge and, subsequently, what pathways organizations might explore for accessing and building talent. Assessing the landscape of DSI talent starts with posing the question:

How does an aspiring data professional emerge?

We have identified and analyzed the four key pathways through which talent can evolve in order to expand the availability, retention, and utilization of DSI talent in developing contexts:

1. New talent
2. Existing social sector talent
3. Transitional talent
4. Leadership

We conducted a literature review of existing studies, a review of a database of nearly 200 existing initiatives that target data and DSI talent and interviewed more than 30 experts and ecosystem actors across all four talent pathways. Our analyses also considered potential variations across geographies, SIOs' data maturity levels, IDEA factors, and sectors and issue areas where relevant. For each pathway, we mapped the current needs, identified systemic challenges, reviewed existing approaches, and began to identify potential solutions.

PATHWAY 1: NEW TALENT

The WEF 2020 Future of Jobs report highlighted data analysts and scientists as the jobs with the highest level of increasing demand.¹⁴ Even as a subset of the full data talent market, the importance of these jobs to the global economy places considerable pressure on the providers of data skills.

Historically, data professionals have graduated from universities with STEM degrees, however, a number of online trainings, boot camps, and short course programs have begun to emerge in recent years. Very few explicitly apply a DSI lens, yet the lessons from what works, what does not, and how the provision of data skills is evolving can provide insight into how we might further expand the pool of new DSI talent entering the workforce.

Critical questions remain: How do we create greater exposure for students demonstrating an interest in DSI? How do we expand the pool of talent by total volume? How do we advance diversity and inclusion across the talent pool?

Definition

We define “new DSI talent” as individuals entering the workforce from traditional or non-traditional educational institutions and programs for the first time. This section focuses on new talent that occupy DSI roles that require intermediate or advanced data skills. For example, talent coming out of training institutions (e.g., universities, boot camps, apprenticeships) or self-taught talent who are entering the labor market to hold positions such as data analyst, data scientist, measurement, evaluation, and learning (MEL) officer, etc., for the benefit of SIOs.

Challenges

Leaving demand constant, the growth of new talent is limited by four critical factors: the low supply of data professionals (quantity), insufficient skills in the labor market (quality), inadequate experience with DSI (relevance), and lack of inclusion, diversity, equity and accessibility (representation).

¹⁴ World Economic Forum, Future of Jobs Report 2020, October 2020.

FIG 8 » OVERVIEW OF KEY FACTORS THAT INHIBIT THE GROWTH OF NEW DSI PROFESSIONALS (Non-exhaustive)

| | | Evidence | Context variations | |
|-----------------------|--|--|--|--|
| Quantity | <ul style="list-style-type: none"> » Insufficient number of graduates with data skills | <ul style="list-style-type: none"> » Limited supply of training opportunities (# of spots) » Limited interest in DS/STEM » Barriers to entry (fees, time, devices, internet, education) | <ul style="list-style-type: none"> — Strathmore University has 400 applicants every year for the MSc. DSA vs. only 60 seats in the program — In Colombia, STEM represents 24% of enrollment, while business majors alone represent 42% — The cost of data science bootcamp programs in Nigeria range between 50-58% of annual wages | <ul style="list-style-type: none"> — Less developed contexts have lower training opportunities — This challenge is less prominent in certain countries (e.g., India) |
| | <ul style="list-style-type: none"> » Graduates with data skills do not come to the social sector | <ul style="list-style-type: none"> » Limited exposure to pathways in the DSI space » Relative attractiveness of S.S. vs. other opportunities | <ul style="list-style-type: none"> — Avg annual salary for data scientists in the social sector in India is ~16.5K USD vs 27K for those in logistics | <ul style="list-style-type: none"> — In some countries, very large SIOs can be the highest paying employers |
| Quality | <ul style="list-style-type: none"> » Limited inclusion of translational skills | <ul style="list-style-type: none"> » Awareness of providers of importance of T.S. » Ability of training providers to deliver translational skills | <ul style="list-style-type: none"> — Curriculum update procedures can be long (7 years in some place), ensuring outdated university curricula — A DIAL study suggests T.S are lacking because they are best learnt in complex projects, which training providers often do not have the capacity to provide | |
| | <ul style="list-style-type: none"> » Limited practical experience/on-the-job training | <ul style="list-style-type: none"> » Awareness/willingness of providers to deliver OTJ » Providers' ability (e.g., access to datasets/ orgs) | <ul style="list-style-type: none"> — Interview with the ACEDS in Rwanda suggest limited job linkages and availability of practical datasets | <ul style="list-style-type: none"> — Emerging markets are beginning to aggregate available data sets |
| Relevance | <ul style="list-style-type: none"> » Limited focus on DSI in developing countries | <ul style="list-style-type: none"> » Inadequate curricula design/alignment with job market | <ul style="list-style-type: none"> — The database shows that DSI programs are a very small minority of DS programs, despite the significant need | <ul style="list-style-type: none"> — More nascent ecosystems have less aligned/ updated curricula |
| | <ul style="list-style-type: none"> » Inadequate job linkages | <ul style="list-style-type: none"> » Inadequate career support for graduates | <ul style="list-style-type: none"> — Interviews suggest that most data bootcamps do not provide meaningful linkages to jobs (e.g., 1 day fair) | |
| Representation | <ul style="list-style-type: none"> » Limited inclusion, diversity, equity and accessibility across the DSI talent lifecycle | <ul style="list-style-type: none"> » Greater barriers to entry » Fewer professional development opportunities after entry | <ul style="list-style-type: none"> — E.g., tertiary enrollment for people with disabilities (UNESCO: Egypt 43% vs. 89%; Indonesia 53% vs. 98%) — A PwC study of 32,500 workers in 19 countries found unequal access to career/ training opportunities on the basis of gender, ethnicity, class, etc. | <ul style="list-style-type: none"> — Specific under-represented groups vary across countries and regions (e.g., minority religious groups, caste, race, etc.) |

Quantity

There are not enough graduates with data skills to meet demand. There are still relatively few students entering data and DSI degree programs due to the limited number of training opportunities, differential access to available training programs, and perceptions of value across most developing contexts.

Global studies show that the demand for data analytics and science skills is outpacing supply. A study by the African Development Bank showed that less than 25% of tertiary graduates pursued STEM degrees (vs. 40+ in other emerging market contexts). In Kenya, for example, a brief review of degree options revealed only two bachelor's degrees in data science¹⁵ and no degrees that focus on DSI, either at the graduate or undergraduate levels. On the other hand, a preliminary search in the UK finds almost 30 bachelor's degrees in data science, many of which are applied science degrees and almost 80 master's degrees in data science, several of which are applied.

Graduates with data skills tend not to enter the social sector. Students and other trainees in data fields have minimal exposure to career pathways in the social sector. These pathways are often less visible or seemingly less attractive than private sector opportunities due to factors that include lower compensation, lack of a “recruiting engine” to capture

new graduates, and fewer professional development opportunities. The average annual salary for data scientists in the social sector in India, for example, is USD 16,500, compared with USD 27,000 for those in logistics.

“Our instructors/coaches are highly trained to ‘coach’ instead of teach and are able to provide the right level of practical support for the trainees. This is not the case with most bootcamps.”

— Matias Hoyl, New Ventures Director,
Laboratoria

Quality and relevance

Interviews with training institutions and a selection of social impact organizations suggest that, due to deficiencies in curricula and instruction, data training programs do not provide professionals with the relevant skills to meet the demands of a social sector workplace. In particular, the programs are perceived to fall short on dimensions of critical thinking, translational skills, and practical experience, thanks in part to limited faculty and lack of access to relevant datasets.

¹⁵ The bachelor of science in data science and analytics at the United States International University – Africa and the bachelor of science in data science and analytics program at Jomo Kenyatta University of Agriculture and Technology. African Data Science Training Organization.

Representation

The DSI field presents significant barriers to entry that disproportionately affect disadvantaged and underrepresented groups. Most data training programs are not affordable relative to average incomes or realistic for working individuals.

Graduate programs are the most expensive. For example, CMU-Africa costs USD 16,000 per year. The least expensive, such as MOOCs, are free of charge, yet still require internet access and computer equipment that may be unaffordable for aspiring trainees. While boot camp tuition is typically lower than that of tertiary educational institutions, it is still unaffordable for the average trainee—the average cost of a boot camp is USD 4,430 while the median annual income in Kenya is USD 3,315. These challenges are most significant for disadvantaged groups, including lower-income groups (fees) and women (time poverty).¹⁶

The data professionals that do work in developing contexts—and really, in all contexts—tend not to be representative of the diversity of their societies.

In 2019, UNESCO data showed that 107 out of 114 countries studied had fewer women graduating with STEM degrees than men.¹⁷ In the data field, women only represent 26% of the data and AI workforce globally. Similarly, racial minorities are underrepresented. For instance, Black talent only represents 3% of the data & analytics space in the US, and only 1% of leadership positions.¹⁸ Moreover, reports focused on the impact of COVID-19 on workforce dynamics suggest that the pandemic has further reduced opportunities and equity for women, in particular, due to increased responsibilities of family and home.¹⁹

Current landscape

While an explicit focus on DSI is hard to find, the training opportunities for new data talent are growing in number—ranging from MOOCs to short courses to tertiary education degrees—and offer learning that is relevant for building the DSI field.

¹⁶ The average ratio of unpaid care work done by women to men is 2.9X in SSA, 6.5X in South Asia, and 4.5X in the MENA. Dalberg.

¹⁷ There are fewer female than male STEM graduates in 107 of 114 economies, World Bank, May 2019.

¹⁸ WEF Global Gender Gap Report 2020; The Harnham Data & Analytics Diversity Report.2020.

¹⁹ 'The shadow pandemic: How the COVID-19 crisis is exacerbating gender inequality', United Nations Foundation, 2020.

FIG 9 » LANDSCAPE OF TRAINING PROVIDERS FOCUSED ON DATA SKILLS (Non-exhaustive)

| | | |
|---------------|---|--|
| University | Data science undergraduate degrees | |
| | Data science graduate degrees | |
| Short courses | Paid data science bootcamps | |
| | Pay-at-employment data courses | |
| | Summer programs for university students | |
| | Apprenticeships | |
| | MOOCs | |

What has worked well

Despite the extremely limited focus on social impact, our review of 100+ training initiatives for new talent highlighted several approaches that appear to be working well.

- *Integration of job readiness with skills training.* Many online and tech training programs have been criticized for a lack of alignment with industry demands. However, a subset of data skills training business models is beginning to link training to on-the-job experience, including apprenticeship programs. At Google, for example, apprentices gain exposure to a

variety of areas and teams at the company and receive external training to help develop their professional digital skills. Google views apprenticeships both as a means of improving apprentices' job readiness and of fostering a sense of community. In 2022, Google is launching a data analytics apprenticeship in India.

- *Emphasis on placements, not just training.* Interviews suggest that programs that offer internship placements or practical capstone projects to complement in-class teaching demonstrate greater propensity to support professionals in



securing work. For example, Laboratoria engages its extensive network of partner companies throughout the training period in delivering talks, providing feedback on student-run projects, judging the final hackathons, and participating in recruiting events. The trust this builds with employers over time and the opportunity it grants them to interact with students on multiple occasions are key drivers of Laboratoria's high placement rates in high-quality jobs.

- *Robust and customized career support and mentorship.* Programs that go beyond a one-day career fair or a job board, and instead offer customized support to identify the right career path, reach out to potential employers, and prepare for interviews. For example, Andela's Technical Leadership Program, in collaboration with the Rwanda Ministry of ICT and Innovation, thoroughly integrates career guidance and mentorship into the different stages of their nine-month program.

Remaining gaps

We identified a number of gaps in our scan of training programs for new talent. These include:

- *Very limited exposure*, if any, to the applications of data for social impact. Engagement with universities and networks focused on STEM revealed that learning about the application of data for social impact is nearly non-existent outside of the United States. Interviews with experts suggest that this is due to a range of factors, including lack of ready-

“In universities in Mexico and many other Latin American countries, collaboration among departments is notoriously difficult. While in other countries (e.g., US, UK), data science degrees are often a collaboration between departments (e.g., math/ stats, CS, and/ or social sciences depending on the focus), they are usually led by one of these departments in Mexico. As a result, the programs currently available are biased towards one discipline or the other. Data science degrees that come from CS departments are focused on the technology rather than on the problems to be solved. The ones from social science departments usually take an econometrics approach to data science (which is usually simplistic and hence potentially harmful).”

— Jesus Ramos, VP, SoCieDat

to-teach social sector case studies, and limitations of social datasets that can easily be transformed into learning platforms for students.

- *Insufficient focus on placements and job linkages.* Training programs provide limited career exploration support, and a low percentage of trainees are retained for employment post-internship.

- *Limited scale.* The demand for data skills training still far exceeds the scale of training that high-quality programs can provide. For example, Strathmore’s Master of Science in Data Science and Analytics (MSc DSA) can admit just 60 of its 400 applicants per year. Programs have been slow to scale, while the ability of trainees to access and pay for courses further limits the reach of training programs.
- *Low quality or missing skills.* According to the World Economic Forum’s Jobs Report for 2020, skills gaps continue to be a pressing issue, particularly as the demands of jobs change. WEF cites soft skills such as problem solving, critical thinking, self-management, and analysis as key gaps. Another study, by the Digital Impact Alliance, showed that only 35% of software training programs in Sub-Saharan Africa include these “soft skills”²⁰ in their curricula.²¹ While this is not a perfect corollary for data skills training, the findings potentially point to greater systemic challenges across the industry.
- *IDEA.* Access to training programs is not equitable due to a range of reasons, including those that affect access to STEM education more broadly. Girls perform equally if not better than boys at primary levels; however, between just 3 and 7%

“The long-term process of acquiring these skills begins by enhancing tertiary education. Because of the wide array of competencies that data scientists are expected to possess, university and graduate degree programs may have to be altered, particularly in lower-income countries.”

— Data for Better Lives, World Bank, 2021

of girls who attend higher education study STEM-related courses. Geography is another powerful limitation on access: three out of four DSI courses in our scan are in developed contexts, while over 90% of data apprenticeships we identified are in developed contexts.

Emerging opportunities

Based on the analysis, opportunities to strengthen new talent pathways begin to emerge:

- *Support the growth and scale* of training programs that combine theoretical and practical learning while providing greater emphasis on placements.

²⁰ This study defines ‘soft skills’ as required skills that fall outside technical and managerial skills (e.g., Time management, communication, and teamwork).

²¹ Digital Impact Alliance, Building Next Generation of Developers in Sub-Saharan Africa. 2019. By soft skills we refer to common skills or core skills, which are skills applicable to all professions. These include critical thinking, problem solving, public speaking, professional writing, teamwork, digital literacy, leadership, professional attitude, work ethic, career management and intercultural fluency.

- » *Identify specific programs* with effective training and high placement rates (e.g., boot camps for intermediate talent, tertiary degrees for advanced talent) and provide funding or technical assistance to focus on DSI, expand within existing markets, or pilot in new markets.
- » *Integrate DSI courses and datasets* into existing university curricula and provide mentorship. Work with universities and SIOs (starting with more mature SIOs) to develop DSI datasets and case studies and integrate them into the curricula of STEM and general data degrees. Work with universities that provide social science degrees to improve students' exposure to data skills and integrate DSI databases and case studies into data courses. Work with universities to develop mentoring programs for students interested in the DSI space.
- » *Consider investment in experimentation* that will drive new DSI training models such as apprenticeships, DSI university curriculum and degrees in developing contexts, and new programs that combine virtual and in-person teaching with internships.
- *Expand equitable access* to training opportunities through new financing models and certifications (e.g., conditional scholarships, income-sharing agreements) and awareness raising within the field (including with an IDEA focus).
- *Deploy innovative financing tools* to expand access to and efficacy of training programs. Provide conditional scholarships, such as forgivable loans in exchange for five years in the social sector, or income-sharing models to expand training opportunities for lower-income youth. Set up innovative results-based financing tools that provide funding for a range of implementing partners if they commit to training and placing a certain number of youths in DSI roles in developing contexts.

“Results-based financing mechanisms can unlock training opportunities to potential trainees as they can receive funding to pay for training and living expenses and pay back when they get a well-paying job. For example, a national pilot program by a Boston non-profit is training 500 low-income people for in-demand jobs without requiring them to pay the tuition fees upfront until they get a well-paying job. The training is delivered by General Assembly and is funded by career impact bonds drawn from a \$40 million fund provided by philanthropic investors.”

— Anonymous Talent Organization

PATHWAY 2: EXISTING SOCIAL SECTOR TALENT

The economic benefits of upskilling or reskilling employees have prompted organizations across sectors to rapidly adopt this approach to building their workforces. Organizations are motivated to upskill employees in order to increase staff retention and preserve institutional knowledge by providing a competitive professional development opportunity, cut costs through increased workplace productivity, and support long-term growth by training a competitive pipeline of skilled employees.

Similarly, reskilling allows organizations to ensure a pipeline of qualified talent when there are existing gaps in labor market supply and reduce the costs of acquiring skilled talent. Studies estimate hiring new talent costs USD 4,000, compared to USD 1,500 to retrain.²² Finally, reskilling allows organizations to retain staff with organizational or sectoral expertise who would perform at higher levels once reskilled, compared to external talent with similar technical skills. Individuals, meanwhile, see upskilling as a means to obtain continuous professional development throughout their careers, improve their daily work experience, and provide stronger job security by keeping up with industry trends.

Organizations are rapidly doubling down on their internal training initiatives. According to the World Economic Forum's 2020 Jobs Report, since 2018 there has been a fivefold increase in employer provision of online learning opportunities for their workers globally, and 70% of employers expect to offer reskilling and upskilling to a majority of their employees by 2025. In addition, 66% of employers surveyed expect to get a return on their investment in upskilling and reskilling within one year. At the same time, the number of workers seeking new opportunities for learning online has increased by a factor of four.²³

Definition

We define “existing social talent” as professionals already working in SIOs who have the potential to be upskilled or reskilled to take on data roles.

SIOs lag in the implementation of upskilling and reskilling initiatives due to financial constraints, opportunity costs in productivity, and limited knowledge of needs. Stakeholder interviews demonstrate that there is limited use of upskilling in data skills from SIOs. When there is demand, this is still focused on basic skills related to impact evaluation or MEL.²⁴ According to a study by Data Orchard, just over a third of interviewees say their organization supports staff in developing their data and analytics knowledge and skills.²⁵

22 'How to Evaluate Existing Talent for Upskilling and Retraining Purposes', Censia, 2020.

23 World Economic Forum, Upskilling for Shared Prosperity, 2021; World Economic Forum, Future of Jobs, 2020.

24 Interview with J-PAL.

25 State of the Sector Data Maturity in The Not-For-Profit Sector, Data Orchard, July 2021.

Challenges

SIOs at different levels of data maturity have different upskilling needs and face different challenges that constrain their ability to implement initiatives. We observe four key types of challenges: perception of a low return on investment, limited human resources and fear of poaching, insufficient financial resources, and lack of access to relevant training resources.

- *Return on investment.* SIOs tend not to be sufficiently aware of the benefits of upskilling or reskilling programs and the need for data skills more broadly. At all stages, SIOs may also perceive a risk of poaching or leakages after upskilling—that is, return on investment is perceived to be low because upskilled employees will quickly take more lucrative opportunities in the private sector, due to the general supply shortage of data talent across industries.
- *Limited human resources.* Many SIOs are stretched thin with their day-to-day operations, and struggle to find ways to continue to operate while staff receive part-time or full-time training. In one study, 58% of nonprofit workers in the UK cited the lack of time as one of the most important factors in choosing not to participate in training²⁶—a rationale echoed in our stakeholder engagement in emerging markets.

“Demand for data use from non-profits is typically around M&E and impact evaluation. In most cases there is not yet demand for more advanced data uses for decision making such as to design innovations or targeting, except in a few promising use cases where we see creative use of mobile apps, satellite data, and others.”

— Aparna Krishnan, Project Director
IDEA Lab, J-PAL South Asia

“What we need to keep in mind with regards to upskilling and reskilling, is the human resource realities that organizations face. There are real problems regarding who will do certain jobs while employees are out of office getting trained.”

— Han Sheng Chia, Executive VP Innovation
GiveDirectly

²⁶ ‘Small charities lack funds to train staff’, Charity Digital.

- *Insufficient financial resources.* Nearly 75% of UK nonprofits surveyed allocate less than 2% of their budgets to training employees and volunteers.²⁷ SIOs are typically limited by donors' priorities, which are often focused on programmatic funding and near-term, more visible results.
- *Lack of training.* Many SIOs also lack the ability to assess, design, and deliver effective programs. This may be particularly true for SIOs that are still in the nascent stage of data maturity; however, interviews with experts suggested that the challenge of defining specific data skill needs is not limited to the less data mature organizations, nor to the social sector.

Finally, even when organizations' awareness, human resources, and financial resources are not obstacles, there is a lack of available training resources tailored to the social sector.

Current landscape

Upskilling initiatives can take several forms, including in-house, outsourcing, and sponsorship models. In-house initiatives draw on the ability and capacity of an employee or team to identify skill gaps and to design and deliver an appropriate curriculum with the appropriate training model. Outsourced initiatives involve engaging external contractors or partners or identifying online courses and workshops. Organizations can also choose to support employee upskilling or reskilling through a

²⁷ State of the Sector Data Maturity in The Not-For-Profit Sector, Data Orchard, July 2021.

“The biggest mistake that we’ve been making in this work is offering data-specific training. We need to be offering programmatic courses that incorporate data. For example, human services or homelessness services training, and data is a component of this training. This will show organizations how they can get closer to their goals, instead of offering training for the sake of training. Despite being well-intentioned, we need to listen more closely to what non-profits actually need, instead of taking a top-down approach.”

— Neal Myrick, Vice President,
Transformative Philanthropy, Salesforce

“We have successfully worked with employers across sectors to design bespoke upskilling programs. The more tailored the upskilling program is, the more effective it is and the more interest and participation we get from the trainees.”

— Dr John Olukuru, Head of Data Science
and Analytics Strathmore University

sponsorship model of directly subsidizing or covering the entire costs of ongoing education and training.

While most upskilling and reskilling initiatives that we identified originated outside the social sector, the applicability of these programs could be promising for the advancement of DSI

“We were able to take advantage of the presence of data science fellows in the organization, to set up in-house upskilling programs. The fellows delivered workshops and training sessions to a select number of our staff, and now we have two staff members who have become proficient in building dashboards.”

– Santana Khurana, Director MEL
Educate Girls

talent and skills. The table below outlines nine potential models that organizations can use to upskill or reskill their staff.

“In the interest of embedding capacity within organizations, we now work with teams instead of individuals. The driving logic is twofold: a) the skills should benefit the organization even if an individual moves on from it & b) the individual’s role at the organization should be in alignment with the skills being taught for proper assimilation of said skills.”

– Anusha Meher Bhargava, Co-founder
& Chief Insights Officer, Tech4Good
Community

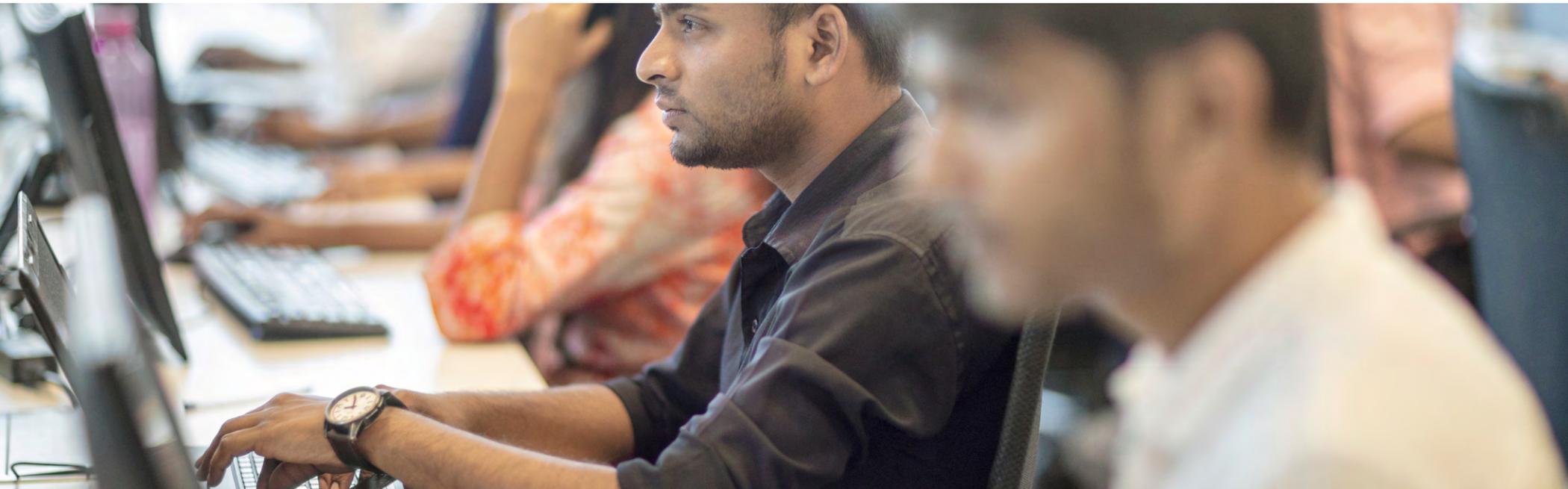


FIG 10 » LANDSCAPE OF TRAINING PROVIDERS FOCUSED ON DATA SKILLS

| | | |
|---------------------|-------------------------------|---|
| In-house | Job rotation | »» The practice of moving employees between jobs in an organization to transfer and distribute skills, knowledge, and competencies |
| | Job enlargement | »» The practice of including additional activities to an existing role to increase experience with different tasks |
| | Peer mentoring | »» A program in which experienced or skilled employees support their peers to expand, refine, and build new skills |
| | Internal workshops | »» Internally designed training workshops (refreshers or exposure to basic concepts) that are delivered by existing employees that are already proficient in the skills |
| | Internal short courses | »» Internally designed training courses that are delivered in-person or virtually by existing employees who have an advanced level in the needed skills |
| Outsourcing | External workshops | »» Training workshops that are designed and delivered by external facilitators, according to the needs of the organization |
| | External short courses | »» Training courses that are designed and delivered by external training providers, (e.g., universities, bootcamps, etc.) according to the needs of the organization |
| | Online courses | »» Online educational courses that are used by the organization's HR departments to upskill or reskill their staff (e.g., Kubicle) |
| Sponsorships | | »» Directly covering the costs or subsidizing training at existing programs (e.g., sponsored Masters programs, subsidized online courses, etc.) |

What has worked well

- *Upskilling programs* have found success by starting with diagnostics to identify specific training needs, such as running a skills assessment in order to identify specific objectives before making decisions on training.
- *Collaborations with local institutions* to develop and deliver trainings have also produced promising results by mixing in-person and virtual teaching and practical application within the employer organizations. And as one training provider explained, “The more tailored the upskilling program is [to a specific sector or organization], the more effective it is and the more interest and participation we get from the trainees.”²⁸
- When full-time or temporary data talent is available, *in-house upskilling* in basic and intermediate skills can be successful.

Remaining gaps

We identified a number of gaps in our scan of training programs for existing talent. These include:

- *Unresolved challenges of considering, designing, and implementing upskilling.* Most SIOs still lack the awareness of needs and benefits, the technical ability to diagnose and assess skills needs, the ability to finance training programs, etc. Few initiatives provide comprehensive or large-scale support for SIOs to overcome these challenges.

- *Leakages post-training.* SIOs risk losing upskilled employees (and the investment in their training) to better career opportunities outside of the field. When single employees receive training, as opposed to teams, the loss is particularly acute.
- *IDEA.* Upskilling programs rarely prioritize elements of inclusion, diversity, equity, and accessibility in the recruitment or training processes, or as part of a retention strategy. The lack of intentionality can have a range of unforeseen impacts and lead to missed opportunities. For example, many upskilling trainings are often delivered after work, which may be particularly burdensome for time impoverished workers—in most cases, women.²⁹
- *Human resources challenges.* Training programs often give limited consideration to HR challenges from the SIO’s perspective—most notably, how to fill gaps when employees are being upskilled or reskilled.

²⁸ Interview with Strathmore University

²⁹ Guild Education

Emerging opportunities

Across these findings, several opportunities begin to emerge:

- *Expand access to upskilling opportunities* through new financing models, such as conditional financing for trainees who can improve their long-term career prospects in exchange for staying at their SIOs for a set number of years.
- *Develop new partnerships* with leading academic institutions across developing contexts in order to design and deliver targeted upskilling programs for SIOs.
- *Strengthen SIOs' ability to deliver upskilling programs* through direct funding and technical assistance as well as freely available diagnostic tools.
- *Raise awareness* among leaders and management on the necessity and benefits of upskilling or reskilling, how to run a diagnostic assessment of training needs, and the intentionality of prioritizing factors that drive inclusion, diversity, equity, and access.
- *Develop and disseminate trainings* as public goods that can be used either by SIOs that have in-house capacity to deliver training or by training providers.



PATHWAY 3: TRANSITIONAL TALENT

Definition

We define transitional talent as data professionals who can potentially move from the private or public sector into the social sector. In this report, we are specifically focusing on existing data professionals with intermediate and advanced data skills, such as data analysts, scientists in telcos, banks, tech companies, etc.

Challenges

While the challenges that existing data professionals face when considering a move into the social sector vary greatly (depending on sectors, stage of data maturity of the SIO they're moving into, geographic location, IDEA dimensions, etc.), we have identified two key types of challenges: limited attraction to or interest in opportunities in the sector, and inadequate readiness to work and excel in the sector.

Opportunities in the social sector may hold minimal attraction or interest to existing data professionals. First, as discussed earlier, compensation is relatively low compared to the private sector. In addition, professional development opportunities at SIOs may be more limited compared with those available at larger companies, and smaller data teams mean fewer opportunities for peer learning. A lack of interest in the sector could also stem from low visibility.

Existing data professionals may feel unprepared to navigate sector dynamics or work in ecosystems with different levels of data maturity to which they are accustomed. Data professionals may find their limited knowledge of social sector dynamics and inadequate expertise in relevant sectors to be challenging. A further challenge may be the expectation that they work with fewer datasets, less structure, and greater role flexibility.

“We have limited success with volunteering programs, because most corporate CSR programs look very cool until you see the real offering – which is usually ‘3 hours of an employee’s time’. Nothing happens in three hours when contextual knowledge is so vital to the impact you can have.”

— Dianna Langley
COO, NetHope

Current landscape

Based on a review of more than 30 initiatives that aim to attract data professionals to the social impact sector, distinct models emerged. These include hands-on fellowships, short courses, volunteer opportunities, hack-a-thons, rotational leadership programs, outsourcing services, advisory services, and a portfolio of toolkits, platforms, and best practices.

FIG 11 » **OVERVIEW OF LEADERSHIP INITIATIVES LINKED TO DATA TALENT FOR SOCIAL IMPACT** (Non-exhaustive)

| | |
|--|--|
| Data for Good hands-on fellowships |      |
| Data for Good short courses |    |
| Volunteering opportunities at social impact organizations |    |
| Full-time rotational programs for future leaders |  |
| Data competitions/hackathon programs |    |
| Specialized organizations that provide outsourcing services |         |
| Specialized organizations that provide advisory services |     |
| Toolkits, platforms, and best practices |    |

What has worked well

- *Long-term fellowships* for one-year or longer seem to have greater success. Also promising are placements that are not attached to full-time jobs—with the expectation that data professionals will return to their jobs—or that offer stipends or salaries that are relatively in line with their current salaries.
- *External exposure* to the SIO context. One of the organizations interviewed observed that bringing data professionals into the organization in order to upskill employees led to greater exposure for the data professional and simultaneously supported that person's interest in the organization itself.
- *Visionary data-driven leaders*. Distinct examples of attracting high-quality talent are generally SIOs that have visionary and capable leaders and strong reputations in the field, as well as funders that are committed to investing in data capabilities. SIOs that have managed to develop relatively distinguished reputations, well-known use cases, or the ability to attract certain types of funding have been successful in attracting existing talent.

Remaining gaps

Despite interesting examples of initiatives that attract data professionals to the social impact sector, gaps remain.

These include:

- *Disproportionate focus* on short engagements that don't appear to deliver tangible results. Interviewees saw programs at three months or shorter as too brief to have a meaningful impact.
- *Glamour project priority*. Funders and program implementers tend to focus on projects that seem externally exciting, as opposed to the actual needs of the organization.
- *Understanding time and effort*. SIOs often do not think through the significant time and effort needed to manage this data talent or the skills that will be needed beyond the support period.
- *IDEA*. Because fellowships for data talent are generally less lucrative than jobs, those from more advantaged backgrounds are typically more able to avail themselves of fellowship opportunities.

“Technical talent who are in the private sector often have the perception that there isn't sufficient professional growth in the social sector, which limits our access to talent. For example, when they ask about the size of the team or the type of work we're doing, they often believe that it is not sufficiently challenging and that they will not evolve professionally or technically.”

— Shравan Kumar Yadav, Data Architect,
Girl Effect

- *Long-term career guidance.* Few programs provide long-term guidance, such as mentorship or career planning, for data talent throughout their career path.
- *Uncompetitive.* In terms of salaries, benefits, and career progression, most SIOs are still unable to compete with other sectors for either fellows or full-time data talent.

Emerging opportunities

We see many opportunities for training programs to address these gaps:

- *Expand successful programs* or create new ones that can meaningfully attract data talent to the social sector.
- *Engage funders and implementers* to improve the delivery of existing programs by, for example, integrating support for SIOs on staffing and HR management during and after placements.
- *Expose people early* to DSI in data courses, which can influence their choices later in their career as they consider a move from the private sector.
- *Build mentorship programs* (with a strong IDEA lens) to provide career guidance and help individuals navigate career switches in and out of the DSI space.
- *Give technical assistance* to SIOs to improve and develop HR strategies—whom to attract, how to attract them, ways to build tech-friendly culture and working environments, etc.
- *Support SIO fundraising* to build data teams and attract data talent—engage SIO funders to raise awareness about the importance of investing in talent and provide better storytelling back to funders about illustrative wins.

PATHWAY 4: LEADERSHIP

Building a new field requires leaders who are willing to take risks and pave a different path. These leaders are essential for shaping DSI. They will need to envision and define data-driven strategies that unlock impact; establish resources and infrastructure to adequately support teams; attract, hire, and retain talent capable of realizing this vision; and influence stakeholders in order to secure resources, establish partnerships, and deliver on goals.

Definition

We define DSI leaders as senior executives or well-positioned individuals who have the role or agency to design or supervise the execution of an organization's data strategy (e.g., CEOs, CTOs, CIOs, Chief Scientists, Head of Analytics, etc.), and who will influence an organization's data usage, practices, and culture.

However, DSI leaders are at very different stages of engagement when it comes to the value of data. Some are in the early stage of exploring and seeking to understand the opportunities and applications of data. Others may understand the potential value but require resources to invest in new systems, solutions, or experimentation. Others may struggle to convince boards or funders to experiment with new approaches or consider new investment in talent. And finally, even innovators who have been able to secure resources and support may struggle with the question of when and whether to outsource or build data

“The big push to build an army of data scientists for jobs in the public sector, private sector, and civil society must be complemented with efforts to create enabling institutional and leadership environments that place a high premium on the use of data and evidence.”

— Data for Better Lives, World Bank, 2021

talent teams in-house. Across these diverse stages, the leadership challenges are diffuse. And, while there are distinctive DSI leaders setting examples for others, these are not the norm, and most leaders are still facing a range of challenges.

Challenges

Through interviews with leaders, experts, funders, and other stakeholders who have found ways to overcome critical challenges, we have a better understanding of the tensions and frictions they face.

SIO leaders have few visible examples to follow and a lack of diverse, relevant role models. SIO leaders tend to have limited awareness of specific DSI use cases in relevant sectors. Many leaders are unfamiliar with the important roles data skills and professionals can play beyond MEL, and even those who recognize its value struggle to justify the trade-offs of allocating existing resources. According to a study by Data

Orchard, only 1 in 20 organizations have leadership that understands how to use data to improve organizational performance. Another survey, by EveryAction, found that while 90% of nonprofits in the US collect data, roughly 50% aren't fully aware of the ways data can impact their work.³⁰

SIO leaders also have limited skills and capabilities to develop and lead data visions and strategies. Most have inadequate basic data skills and need more support to develop data visions and strategies for the organization. In fact, a survey by TechSoup shows that while 83% of NGOs see the benefit in using and investing in technology (including data capabilities), only 40% have managed to define a strategy and a timeline for achieving digital readiness.³¹ In addition, a study by Data Orchard found that only a third of nonprofits globally say they have appropriate numbers of staff managing and developing their data capabilities. SIO leaders also have a limited ability to drive change management. A study by Qlik shows that leaders face difficulties with tackling resistance from the workforce, as 45% of employees surveyed heavily rely on “gut feelings” to make decisions.

Limited resources also present a challenge. SIO leaders lack access to sufficient or appropriate datasets, tools, and technical knowledge, as well as support to define and build teams. Leaders have limited understanding of ways to engage

“The fellowship only admits one person from each organization because we are aware that this is a significant investment, for both the program and the participants.”

— Leadership fellowship program provider

“In our current program for women, we take four half days (i.e., two full days) of their time every month, and this is already too much for them. Leaders are very busy and find it very difficult to attend all the sessions.”

— Shehzia Lilani, Country Director,
Amani Institute

and influence key decision makers (including boards and funders), and so find minimal success in mobilizing resources to invest in data capabilities. An IBM study found that up to 74% of nonprofit leaders cited budget constraints, due to competing internal investments and the restrictive nature of grants, as one of their top barriers to advancing analytic capabilities. And a survey by TechSoup has found that 82% of NGOs in the Global South cite a lack of funding as their

³⁰ ‘The state of data in the nonprofit sector’, EveryAction.

³¹ ‘Data Handling and Digital Readiness in Civil Society’, TechSoup, 2020

biggest barrier to adopting digital tools for data efficacy.³² Resources are difficult to find for operations and for building the capabilities of the organization itself.

SIO leaders pointed to the difficulty of translating strategy into the talent base. Even with a clear, data-driven strategy in place, leaders struggled to articulate specific skills and talent needs, understand how quickly analytic skills may automate, and determine when to build tools and talent in-house versus assign data responsibilities to outsourced teams.

Finally, SIO leaders generally lack access to networks of other data-driven SIO leaders to draw on for advice and support. As one expert stated, “The ability to be part of a network and share ideas and experiences is very powerful for nonprofit leaders.” The lack of this kind of network focused on DSI deprives SIO leaders of the opportunity to learn from the insights of peers who are familiar with the rewards and challenges of integrating data into the programming and operations of social impact organizations.

“Most non-profit leaders are already aware of the need to use data science. The key question here is to have sustained change management processes in place inside the organization to shift the way it works.”

— Lauren Woodman, CEO, DataKind

“There are many funders that also do not fully understand data, and don’t know how to fund data capabilities and what the potential impacts of this funding will be.”

— Neal Myrick, Vice President,
Transformative Philanthropy, Salesforce

“Most funding for non-profits is focused on programmatic budgets (i.e., goes directly to the end beneficiary), which starves organizations of resources to build themselves, including investing in leadership capacity and data capabilities.”

— Isha Sharma, Associate Director
India Leaders for Social Sector (ILSS)

³² ‘Data Handling and Digital Readiness in Civil Society’, TechSoup, 2020.

“We need traditional funding to shift. Otherwise, all the digital initiatives must come out of unrestricted funding, and that starves it. The only way we can figure this out is if it additionally sustained through restricted funding mechanisms.”

— Dianna Langley, COO, NetHope

“We need more funders to understand the value of unrestricted funding. There are a few that already get this, but we have an opportunity to educate the rest.”

— Christine Lopez, Head of Strategy,
Social Impact, Splunk

“Interviews with representatives from NGOs in Argentina, Kenya, and Nepal revealed that lack of funding can constrain the collection of citizen-generated data. Similarly, although the cost of sensors has steadily fallen over the last few years, the costs of equipment, deployment, and transmission, as well as the lack of off-the-shelf tools for environments facing resource constraints, are still major barriers to the generation and use of machine-generated data.”

— Data for Better Lives, World Bank, 2021



Current landscape

The current landscape of solutions and resources to support these field-builders is limited. A review of more than 30 leadership programs in developing contexts revealed that only four could be considered to have a focus on DSI. But the overlap between programs focused on social entrepreneurship, technology innovation, and issue-based

leadership in the social sector can provide us with some insight. The goals of most leadership programs offer common direction and thus potential for learning. Specifically, the majority of leadership programs we reviewed included a focus on one of three distinct outcomes: i) community building, ii) leadership training, or iii) initiative building.

FIG 12 » **MAPPING OF LEADERSHIP PROGRAMS RELATED TO DSI OBJECTIVES**



Emerging opportunities

Through interviews and analysis of these programs, opportunities for supporting leadership programs begin to emerge:

- 1.** *Have a clear goal.* Whether to drive awareness, build skills, advance ideas, or influence key stakeholders. An evaluation of Mozilla Foundation’s mentorship programs highlighted the importance of being clear on the program’s intentions right from the outset.
- 2.** *Be intentional about inclusion, diversity, equity, and access.* While most interviewees state a clear understanding of the value of IDEA, very few programs intentionally embed it as a grounding principle of design. There is an opportunity to both define clearer paths to IDEA talent and support interventions to realize more diverse outcomes. Build a distinct community of diverse DSI professionals, either tackling similar challenges or at similar stages of data-driven innovation within their organizations.
- 3.** *Link to practical solutions and complementary efforts.* Interviewees emphasized the value of practical learning that allowed them to identify their own challenges and problem solve in order to advance initiatives, experiments, and solutions.
- 4.** *Target teams over individuals.* Most programs fund only individual participation while leaders struggle with the isolation of “tech ideas.” Interviewees highlighted the importance of considering programs and initiatives that support the interconnected links between data professionals, their teams, their organizations, and the sector.
- 5.** *Ensure sufficient time and space* to engage and benefit from programs. Multiple interviewees cited challenges linked to finding time to balance professional responsibilities with leadership programs.
- 6.** *Actively learn and evaluate* in order to refine and grow. The nascency of leadership programs and difficulty measuring impact requires active and continuous focus on learning and evolving as the programs progress.
- 7.** *Recognize the value of alumni.* A few programs—in particular, Mozilla Fellows and AFLI Tutu Network—are actively investing in ways to formalize the strength of alumni networks in order to harness the passion and commitment of the community beyond the current program participants and faculty.

Building from these insights, ways to support the growth of DSI leaders around the world begin to come into focus. Build from early momentum. While very few DSI-specific leadership programs exist today, the bright spots identified in this report could offer space for learning, growth, and expansion more deeply into DSI leadership support.

A critical aspect of building broader-based support is working to create allies across the ecosystem.

Engagement or intervention with key influencers can help establish a shared understanding of the value of data-driven solutions, the complexity of resource allocation, and how to identify and prove a business case for investing in data talent and systems.

EMERGING OPPORTUNITIES ACROSS THE DSI ECOSYSTEM

While each pathway has distinctive characteristics affecting the balance of supply and demand, cross-cutting challenges and opportunities stand out across the DSI ecosystem.

The demand for DSI talent among SIOs in developing contexts needs to be better understood and measured in order to more effectively match demand to available skills and talent. Limited availability of data on how DSI ecosystems are evolving continues to be a barrier to the growth of the DSI labor market. In addition, more dynamic data on the DSI labor market will enhance the resilience and responsiveness of stakeholders across the ecosystem—for example, employers, education

providers, and work seekers themselves. This need for granular information is even more acute for demographic, socioeconomic, and other factors linked to IDEA. Without these data, it is difficult to get a clear, segmented picture of the state of IDEA in DSI. Furthermore, a lack of standard performance measures and performance data make it difficult to assess promising models.

We see an opportunity to advance IDEA values as foundational assets in enhancing the access to, value of, and impact of DSI talent initiatives. While the underlying drivers of exclusion are many, very few pathways revealed examples of initiatives, programs, or models that intentionally focused on building and shaping an IDEA ecosystem. At the same time, a range of organizations are actively highlighting both the value and challenge of building truly IDEA environments. Through targeted partnership, there is an opportunity to identify approaches, share learning, and actively test new solutions that can expand the DSI data talent pool.

There is also an opportunity for funding to drive more inclusive, equitable access to DSI training and encourage DSI placements rather than short-term outputs. Across nearly all pathways, respondents highlighted the opportunity created by better aligning funding and financing with longer-term outcomes. This included consideration of structures such as pay for performance, as well as experimental funding for new approaches to attracting and retaining talent and seeding IDEA participation in current or future training programs.

Lastly, the “connective tissue” could be strengthened to better align complementary efforts across labor market resilience, digital transformation, national development agendas, and social impact initiatives. While these areas of focus do not fully map to DSI objectives, their relevance suggests that greater engagement, alignment, and coordination among actors could be beneficial—to avoid duplication, acknowledge and value public goods, and unlock future growth. For example, many initiatives and investments focused on the digital transformation of national economies include targeted work with educational institutions in order to advance skills development. Aligning resources, training initiatives, and research could prove more effective than operating in silos.

“Our global open enrollment post-graduate program in social innovation management is usually predominantly composed of women and has very few male participants. However, when we first implemented a program for mid-senior managers in India, we were surprised to have a cohort of 3 women and 21 men. Once you go from junior towards mid to senior management, the DEI statistics shift drastically, and it’s difficult to attract a diverse leadership cohort when most organization leaders are not diverse. We were eventually able to get to an almost 50-50 split in some of our cohorts with intentional commitment to DEI, but it’s still an ongoing challenge for us.”

— Shehzia Lilani
Country Director, Amani Institute

4 RECOMMENDATIONS TO ADVANCE THE FIELD



As we look across this emerging field, we see significant opportunities to accelerate the growth of a truly diverse and inclusive DSI talent base. We offer seven overarching recommendations to make the most of these opportunities:

1 EXPERIMENT EARLY AND EVALUATE OFTEN

Experiment early and evaluate often. The current landscape for purpose-driven data professionals is nascent in terms of its overall size and organization but shows momentum and growth. The nature of a nascent field requires actions that will crowd in others, organize efforts, draw attention to what works, and facilitate frequent pivots.

Very few existing solutions specifically target social impact leaders or organizations. Of approximately 125 training initiatives for new talent we reviewed, fewer than 20 could be considered DSI focused. Given this starting point, rather than seeking to invest or build for scale, start by crowding in more ideas and new actors ready to build or modify solutions to create more DSI focus when it comes to talent.

At the same time, widen the potential talent base by providing more exposure for more young people and including DSI curriculum in both social science and STEM programs.

Overall, iterate and evaluate often. We do not yet know what works, so we need to build solutions that allow for active adjustment and realignment in order to ensure that we are driving toward intended outcomes.

2 PRIORITIZE FOCUSING ON AND INVESTING IN INCLUSION, DIVERSITY, EQUITY, AND ACCESSIBILITY

Prioritize focusing on and investing in inclusion, diversity, equity, and accessibility when considering access to education and training, links between training and placement, and absorptive capacity of maturing data ecosystems.

Although diversity, inclusion, equity, and accessibility are widely acknowledged as important, they are not yet woven into program design or outcomes, either for DSI-specific initiatives or data initiatives more broadly.

Consider partnering early with organizations and institutions that can support a stronger IDEA lens on talent identification, recruitment, and retention, and enhance research focused on understanding both the value and potential harms of IDEA in the DSI landscape.

3 RECOGNIZE THE INTERDISCIPLINARY NATURE OF DATA FOR SOCIAL IMPACT

Recognize the interdisciplinary nature of data for social impact where the depth of technological understanding and expertise is matched with the discipline and understanding of social sciences.

Consider broadening the DSI training and education curricula to include modules from other disciplines, like social sciences and humanities, focusing on key areas like, for example, data ethics, working with vulnerable communities, understanding cultural context of data, being aware of power dynamics in data collection and access, and issues such as data colonialism.

4 MOVE FROM INDIVIDUALS TO TEAMS TO ECOSYSTEMS

Move from individuals to teams to ecosystems. Recognize the role an individual leader plays within an organization, a sector, or a broader ecosystem, and align efforts and investments accordingly.

Beyond the paucity of programs specifically focused on DSI professionals, our analysis of existing programs revealed the shortcomings of interventions that train a single individual within an organization, and flagged a need for greater investment in supporting teams or projects. A more holistic approach supports the growth of DSI leaders and allows them to advance specific goals or tackle key challenges.

5 INVEST IN APPLIED LEARNING AND STRONGER LINKS TO PROFESSIONAL PLACEMENTS AND ADVANCEMENT

Invest in applied learning and stronger links to professional placements and advancement. We see shifting from a focus on “the number of people trained” as the outcome to “the number of people playing an active role addressing social issues or working within organizations” as an essential evolution in improving employment and organizational advancement.

As non-traditional training models such as MOOCs, online learning, and short courses have proliferated, we have seen both a lack of scale and highly variable results. Program outcomes are often disconnected from longer-term outcomes and many programs demonstrate a bias toward technical training rather than integrated translational skills and work readiness.

In this context, consider better aligning funding with intended outcomes. Training for the sake of training is not enough. Financing mechanisms that link training to sustainable employment, however, could nudge data training models toward having real impact in the social sector.

6 COORDINATE COMPLEMENTARY EFFORTS

Coordinate complementary efforts. Looking across various efforts linked to enhancing data ecosystems, digital transformation, and data-driven strategies for social impact-oriented organizations, we see significant potential for greater coordination in order to advance DSI as a field.

Momentum around topics such as digital transformation and building data ecosystems is generating meaningful investment and attention. While these investments often include a focus on building skills or strengthening institutions to advance skills training, they are rarely connected to DSI efforts and, according to interviewees, are often overlapping, uncoordinated, and even competitive.

Focused effort to understand where complementary capital, investment, or attention to the value of data and digital skills focused on social impact may yield unforeseen resources, partnerships, and momentum to advance DSI talent goals.

“Many organizations have entered this space in the last 10 years, and several are operating in a fragmented manner. As a result, many of these initiatives are not aligned to countries’ medium-to-long term priorities and related efforts by organizations working to support these priorities. A result has been sub-optimal outcomes for digital literacy and data literacy capacity across stakeholder groups in many countries. A key lesson from this is the clear need for greater coordination and alignment to build comprehensive packages of support for data and statistics priorities, and move away from ad-hoc and fragmented support.”

— Craig Hammer, Program Manager
with the Development Data Group,
World Bank

“We all have to work in concert, otherwise the system just snaps back to where it was.”

— Dianna Langley
COO, NetHope

7 CONTINUOUSLY INVEST IN MORE AND BETTER VISIBILITY INTO HOW THE ECOSYSTEM IS FUNCTIONING

Continuously invest in more and better visibility into how the ecosystem is functioning. We need better data to understand the connections between talent and other aspects of the broader DSI ecosystem. The ability of training programs to adjust their business models is hindered by a lack of accurate market demand data for skills, especially in the social impact sector. While there is a growing appreciation of the value of data for social impact across public, private, and social sectors, the understanding of demand remains too broad to guide accurate decision making.

Better labor market visibility would go a long way toward providing a clearer picture of what is working and what is not in the field of DSI talent—particularly in developing contexts, and with regard to the efficacy on newer, non-traditional training programs.

As a starting point, we see a sector-wide need to build a common language for understanding the skills, job types, and titles that define data roles broadly, and potentially DSI more specifically. In addition, there is an opportunity to evolve and standardize credentials and skills across a range of providers, re-imagining credentialing and skilling entirely.



ANNEXES



DEFINITIONS

- **Data ecosystem:** A context in which several actors interact with a collection of data infrastructure, analytics, and applications that are evolving in real time.
- **Data lifecycle:** A journey recognizing end-to-end data usage by any organization, including generating, collecting, processing, storing, managing, analyzing, visualizing, and interpreting data.
- **Data maturity:** Data maturity is a way to measure the extent and sophistication of data usage within an organization. It recognizes that organizations have different levels of experience, particularly with defining data purpose, understanding data value, and establishing strategies and management practices, for data.
- **Data professional:** A person that contributes to any part of the data lifecycle as a primary or significant part of their professional activity, including generating, collecting, processing, storing, managing, analyzing, visualizing, or interpreting data. For the purposes of this report, we are specifically focusing on data professionals who have and use intermediate and advanced skills as a primary part of their professional activity.
- **Data science:** Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data and apply knowledge and actionable insights from data across a broad range of application domains.³³
- **Data for social impact:** The use of data by social impact organizations to benefit organizations, people, their communities, and the environment more efficiently and effectively.
- **Developing contexts:** Countries, regions, sub-regions or other “contexts” where the infrastructure, social, or economic context is still emerging. This term recognizes that countries often have a range of social, economic, technological, and structural realities that vary significantly. The term was selected as a counter to “developing countries” in recognition of the high levels of variation within economies. The terms low-, lower-middle-, upper-middle-, and high-income countries are used in this study to refer to countries with low-income or middle-income economies, per the definition used by the World Bank.³⁴

³³ Amnesty USA.

³⁴ <https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries>

- **Existing social sector talent:** Current employees of nonprofit organizations or other mission-driven institutions, organizations, or companies.
- **Intersectionality:** A theoretical framework that posits that multiple social categories intersect at the micro level of individual experience to reflect multiple interlocking systems of privilege and oppression at the macro, social-structural level (e.g., racism, sexism, heterosexism).
- **Measurement, evaluation, and learning (MEL):** The purpose of measurement, evaluation, and learning practices is to apply knowledge gained from evidence and analysis to improve development outcomes and ensure accountability for the resources used to achieve them.³⁵
- **New talent:** Individuals entering the workforce for the first time who will be occupying new roles as data professionals.
- **Data talent for social impact/purpose-driven data professionals:** Data professionals who work either as paid employees or volunteers at an organization, institution, or company that is oriented toward social impact.
- **Reskilling:** Learning new skills with the aim of diversifying one's work profile or transitioning into a role with different responsibilities.
- **Social impact organization (SIO):** An organization (for- or nonprofit) working consciously, systematically, and sustainably to address social, environmental, economic, health, or related challenges to drive positive and desired change. Typically, an SIO works to serve marginalized groups. For the purposes of the report, SIO will refer only to organizations that are registered as nonprofits.
- **SIO leaders:** Organization leaders and senior executives who are tasked with designing or supervising the execution of an organization's data strategy and who have the agency to bring about change within the organization's data usage, practices, and culture.
- **Transitional talent:** Existing data professionals who are not currently employed by nonprofits in developing countries.
- **Upskilling:** Learning new skills with the aim of optimizing one's performance in one's current role or updating skills in a changing context.

OVERVIEW OF THE METHODOLOGY

The Data Talent for Social Impact landscape report is the result of a research effort that was launched in October 2021 by data.org and the Patrick J. McGovern Foundation (PJMF), in collaboration with Dalberg. This report has recognized and built upon the growing body of evidence in the data field and in the nascent DSI space, and engaged with several stakeholders from across the ecosystem, as well as adjacent industries.

More specifically, this study consisted of the following activities over a duration of six months:

- *A market sizing exercise* to estimate the potential number of DSI professionals in developing countries over the next 10 years.³⁶ Due to the nascent nature of the DSI field, as well as disconnected and underdeveloped labor markets, there is little evidence and data on the size of the current DSI workforce and its future potential for growth. Therefore, in order to illustrate the magnitude of the opportunity we have at hand, we conducted a market sizing exercise that estimates the current size of the DSI talent pool and provides three possible growth scenarios over the next decade.
- *A literature review* of 90+ studies and research reports to understand the existing data on, analysis of, characteristics of, knowledge of, and discourse about the DSI space. The team anchored on several key questions that are important for this study. For each of these questions, we scanned the available studies and reports to identify (i) areas of consensus, (ii) areas where there are variations across contexts, (iii) trends that have been observed but have insufficient sample sizes, and (iv) questions that have yet to be answered by existing studies. The takeaways from this review were important inputs for the state of the DSI talent landscape, as well as our recommendations for the future of the field.
- *The creation of a database* of almost 200 training initiatives across the four talent pathways. This database was compiled using (i) desk research from public repositories, (ii) data.org and PJMF's ongoing and past partnerships and projects, such as data.org's "Charting the 'Data for Good' Landscape", and (iii) Dalberg Advisors' wide network of 30 offices worldwide and more than 2,000 projects executed in the space over the past 20 years. The goal of this analysis was to identify emerging gaps in the current data and DSI training offered, key trends that have been observed over the past few years, as well as pockets of momentum and innovation in the training space.

³⁶ This exercise was specifically focused on developing countries (rather than developing contexts more broadly) in order to simplify our calculations and facilitate access to the required data points.

- *Stakeholder interviews* with several ecosystem actors that complemented the sizing exercise, the literature review, and the database of existing initiatives, and that validated our emerging hypotheses. We interviewed approximately 30 stakeholders across several categories, including training providers, funders, and leaders of social impact organizations, as well as experts in the field of DSI or in adjacent fields.

LITERATURE REVIEW

The database of 93 sources included 68 external reports, studies, and papers, and 25 articles, essays, and databases. The graphic below showcases some of the organizations behind these publications.

Five key questions guided the literature review:

1. What do we know about the relevance of the data field, especially in the social sector in emerging countries?
2. What do we know about the data skills labor market and supply and demand of professionals globally and across sectors?
3. What do we know about the role of SIO leadership in advancing the use of data for social impact? What about the challenges and opportunities they face?

4. What are the underlying challenges and opportunities for new data talent entering the labor market?
5. What do we know about the trends, challenges, and opportunities in upskilling existing social talent in data skills, and upskilling or attracting existing data talent to the social sector?

For each of the questions, we reviewed the existing knowledge base to identify the following four categories of information:

1. *Areas of consensus*: the findings and observations on which experts and research reports are aligned.
2. *Variations across contexts*: the findings that have emerged, but that are still debated across reports and studies, or that only apply to certain contexts.
3. *Trends with insufficient sample sizes*: observations that have emerged, but that lack depth, or do not have sufficient sample sizes to support generalization.
4. *Unanswered questions*: missing information, data points, and evidence, as well as questions that have not yet been explored or answered through existing research.

We organized the takeaways from the literature review by talent pathway and used them to uncover and analyze the challenges and potential opportunities in the DSI field.

FIG 13 » **LITERATURE REVIEW** (Non-exhaustive)

55

External reports and papers

Spanning the future of the data field, the relevance and demand for data skills, momentum in upskilling and reskilling initiatives, and snapshots of data use in SIOs

13

Internal Dalberg studies

Including prior research on digital skills demand in Africa and Asia, including several country landscapes

25

Articles, essays, and databases

Including industry insights from experts, market research, and public datasets shared by multilateral institutions



ANALYSIS OF EXISTING DATA TALENT FOR SOCIAL IMPACT INITIATIVES

In order to evaluate the current landscape of training opportunities, we built a database of close to 200 initiatives, including (i) 31 leadership programs for social impact leaders, (ii) 105 training programs for new data talent, (iii) 28 internal and external upskilling and reskilling programs in data skills, and (iv) 29 initiatives for transitional data talent. These initiatives also looked at the broader data field in order to

identify trends, gaps, and opportunities that are relevant for this study.

This database allowed us to uncover several trends, such as the rapid increase in non-traditional or alternative data training opportunities in recent years. It also allowed us to identify some of the difficulties and potential opportunities for data talent and pockets of momentum or innovation that we further explored through stakeholder interviews.

FIG 14 » ANALYSIS OF DATA TALENT FOR SOCIAL IMPACT INITIATIVES (Non-exhaustive)

31
Leadership programs for social impact organizations' leaders

105
Training initiatives for new talent

28
Up/reskilling initiatives for social talent

29
Up/reskilling initiatives for data talent



MARKET SIZING METHODOLOGY

The focus of the market sizing was to determine the “potential opportunity for DSI talent over the next 10 years.” Intentionally, this does not aim to assess a static “gap” between demand and supply, but rather, to determine the size of the opportunity if both demand and supply, as well as the overall ecosystem, are stimulated.

We would like to note that this exercise has several limitations due to the lack of data and evidence in this space. However, our objective is to offer a starting point that will allow for others to enhance, build, and deepen the research.

- *Year 0 number:* There is no existing data on the total size of the DSI workforce due to several challenges, including limited labor market data and divergent taxonomies for data skills and jobs. However, we have found 244,000 profiles of DSI professionals in developing countries on LinkedIn.³⁷ We used an imperfect estimate of 350,000 professionals to account for professionals who are not on LinkedIn.
- *Baseline rate:* This scenario assumes that the number of DSI professionals in developing contexts will continue to grow at the same rate as the overall formal employment base. Because the formal workforce growth rate is not readily available, we have calculated it for a few developing

contexts in places like Kenya, South Africa, and India, using UN data from 2016-2020 to arrive at a suitable estimate. The average of these rates was 5.3%, so we used 5% as a relatively conservative estimate for the baseline growth rate. While this methodology is flawed, it is meant to provide a placeholder rate that can substitute the growth rate of the formal labor forces of developing countries.

- *Developed countries’ rate:* This scenario assumes that the number of DSI professionals in developing countries will start to grow at the same rate as the one in developed countries. There are no existing data on the growth rate of data professionals across developed countries, and the statistics from different countries follow different taxonomies and definitions of data professionals. Therefore, we used the projected annual growth rate of data professionals in the EU from the European Data Market Monitoring Tool (EU Commission) as an estimate (10.5%).³⁸
- *Catch-up rate:* This scenario assumes that in 10 years, the rate of data professionals as a percentage of the formal social sector workforce is equal to that of data professionals in the overall workforce in developed countries. In other terms, this scenario assumes that the social sector in developing countries grows at a rapid enough pace to “catch up” with the private sector in developed countries. To

³⁷ Dalberg analysis: A scan on LinkedIn of the roles that include data in their title, and/or match the roles included in the taxonomy table; and that are in non-profit organizations in developing countries.

³⁸ Note: This was the high growth scenario in the study.

calculate the necessary growth rate for this scenario, we used the Employment Share of Data Professionals (ESDP) in the UK³⁹ (the nation with the highest ESDP in Europe) of 5.2% as the “target rate” in 10 years. We then used UN data to estimate the social sector workforce in developing countries and used a 5% rate to estimate the formal workforce in the social sector.⁴⁰ Next, we assumed that this total number (the social sector workforce across developing countries) is growing at the baseline rate (5%) for 10 years. Finally, we applied ESDP (5.2%) in Year 10 and worked backwards to calculate the yearly growth rate (27.5%).

INCLUSION, DIVERSITY, EQUITY, ACCESSIBILITY (IDEA)

We have anchored this report on a framework for inclusion, diversity, equity, and accessibility.

Definition of IDEA

- *Inclusion*: An outcome to ensure that diverse individuals feel and are welcomed and able to participate fully in the decision-making processes and development opportunities within an organization or group. Inclusion outcomes are

met when organizations, institutions, groups, and programs are truly inviting to all.⁴¹

- *Diversity*: The presence of differences that may include race, gender, religion, sexual orientation, ethnicity, nationality, socioeconomic status, language, ability, age, religious commitment, or political perspective.⁴²
- *Equity*: Promoting justice, impartiality, and fairness within the procedures, processes, and distribution of resources by institutions or systems. Tackling equity issues requires an understanding of the root causes of outcome disparities within our society.⁴³
- *Accessibility*: Dismantling barriers to full participation, including physical and virtual spaces, opportunities, culture, language, information, and relationships.⁴⁴

Intersectionality

Intersectionality is a theoretical framework that posits that multiple social categories (e.g., race, ethnicity, gender, sexual orientation, socioeconomic status) intersect at the micro level of individual experience to reflect multiple interlocking systems

³⁹ The European Data Market Monitoring Tool.

⁴⁰ A study conducted by Johns Hopkins University across 13 countries showed that the non-profit sector employs 7.4% of the total workforce on average. In the absence of alternative robust data, we decided to take a more conservative estimate of 5% because (i) these countries were predominantly developed and hence not representative of our target geographies, and (ii) employment in the non-profit sector also includes workers outside our definition of the social sector (e.g., employees in religious institutions).

⁴¹ dei.extension.org

⁴² dei.extension.org

⁴³ dei.extension.org

⁴⁴ Amnesty USA.

of privilege and oppression at the macro, social-structural level (e.g., racism, sexism, heterosexism).

Key sources that showcase the potential opportunities and benefits of centering IDEA in data talent

We have reviewed several sources that make the case for the importance of IDEA across several dimensions, including gender, ethnic and cultural backgrounds, disability, etc., for the performance of individual workplaces as well as the health of the technology sector and the entire economy. Below is a brief overview of some of these sources:

- “Decoding Diversity: The Financial and Economic Returns to Diversity in Tech”:⁴⁵ This study shows that improving ethnic and gender diversity in the workforce could create USD 470 to USD 570B in new value for the tech industry and could add 1.2 to 1.6% to national GDP.
- “Growing Economies through Gender Parity”:⁴⁶ Closing the gender gap in the workforce—through addressing time poverty, digital inclusion, financial access and education—could add USD 28 trillion to global GDP. 28 trillion to the global gross domestic product (GDP)

- “Diversity Wins: How Inclusion Matters”:⁴⁷ The report reaffirms the business case for diversity in the workplace. For instance, using data from 15 countries and more than 1,000 large companies, the report shows that companies in the top quartile for gender diversity on executive teams were 25% more likely to have above-average profitability than companies in the fourth quartile. Similarly, the report makes the case for ethnic and cultural diversity and shows that top-quartile companies outperformed those in the fourth quartile by 36% in profitability.

Key sources that showcase the threats of exclusion and the importance of centering IDEA to avoid some of its downfalls

- “Examining Racial and Gender Bias in Facial Analysis Software”:⁴⁸ Scientist, activist, and founder of the Algorithmic Justice League, Joy Buolamwini found that facial recognition error rates in commercial AI services vary significantly across dimensions such as race and gender. In fact, the rates go from 0.8% for light-skinned males all the way to 34.7% for dark-skinned females. These errors have significant harmful consequences, as reliance on facial recognition is rising (e.g., in law enforcement surveillance⁴⁹). This is a consequence of providing AI algorithms with data

45 ‘Decoding Diversity: The financial and economic returns to diversity in Tech’, Intel & Dalberg, 2016.

46 ‘Growing Economies Through Gender Parity’, Council on Foreign relations.

47 ‘Diversity wins: How inclusion matters’, McKinsey and Company, 2020.

48 ‘Examining racial and gender bias in facial analysis software’ Joy Buolamwini, Google.

49 ‘Racial Discrimination in Face Recognition Technology’, Alex Najibi, Harvard University’s School of Engineering and Applied Sciences.

for one group of people (e.g., Caucasians),⁵⁰ which is more likely to happen if the data teams in charge of the algorithms are not diverse.

- **Weapons of Math Destruction:**⁵¹ Cathy O’Neil’s book highlights how mathematical models or algorithms that claim to quantify important traits (e.g., teacher quality, recidivism risk, creditworthiness) but have harmful outcomes and often reinforce inequality (e.g., by encoding racism or other biases into algorithms).
- **Timnit Gebru:** Gebru is a widely respected leader in AI ethics research, who gained prominence after co-authoring a groundbreaking paper⁵² that raises concerns about algorithmic bias in machine learning and the latent perils that AI presents for marginalized communities. When Gebru was forced out of Google after refusing to retract the paper, the conversation about the tech industry’s inherent diversity

problem resurfaced.⁵³ Timnit Gebru’s leadership in this space (including her role as co-founder of Black in AI) allows us to understand the ways in which the lack of diversity in the data field leads to real and significant threats for underrepresented and marginalized groups.

- **“Notes on Algorithmic Violence”:**⁵⁴ In this article, Mimi Onuoha highlights several examples of the term “algorithmic violence,” which she defines as “the violence that an algorithm or automated decision-making system inflicts by preventing people from meeting their basic needs.”
- **“What Do We Do About the Biases in AI?”:**⁵⁵ After highlighting some of the ways that biases can make their way into AI systems and cause harmful results, the authors conclude that “a more diverse AI community would be better equipped to anticipate, review, and spot bias and engage communities affected.”

50 ‘In AI, diversity is a business imperative’, Carlos Melendez, Forbes Technology Council.

51 ‘Review: Weapons of Math Destruction’, Scientific American, 2016.

52 ‘On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?’, Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchell.

53 ‘Taking On Tech: Dr. Timnit Gebru Exposes The Underbelly Of Performative Diversity In The Tech Industry’, Ashlee Marie Preston, Forbes.

54 ‘Notes on Algorithmic Violence’, Mimi Onuoha, Github.

55 ‘What Do We Do About the Biases in AI?’, James Manyika, Jake Silberg, and Brittany Presten, Harvard Business Review, 2019.

LIST OF INTERVIEWS

| Program/Organization | Geography | Contact name | Contact position |
|--|------------------------|----------------------|---|
| Acumen Fellows Program | Africa | Gladwel Anzaye | Program associate |
| African Center of Excellence in Data Science (ACEDS) | Rwanda | Charles Nzaramyimana | Head of Postgraduate Studies |
| African Institute of Mathematical Sciences (AIMS) | Africa (multi-country) | Bubacarr Bah | Chair of Data Science |
| Amani Institute | Global | Shehzia Lilani | Country Director |
| Andela | Africa (multi-country) | Sabrina Roshan | Head of Expansion Strategy |
| Dasra Social Impact Leadership Program | India | Neera Nundy | Managing Partner |
| DataKind | Global | Lauren Woodman | CEO |
| Educate Girls | India | Santana Khurana | Director, Measurement, Learning and Evaluation |
| eMobilis | Kenya | Ken Gikunda | Co-founder and Managing Director |
| Global Expert (former Microsoft) | Global | Frank McCosker | Previous General Manager, Affordable Access and Smart Financing 4Afrika (Microsoft) |
| Girl Effect | Global | Shravan Yadav | Data Architect |
| GiveDirectly | Global | Han Sheng Chia | Former Vice President, Innovation |

| Program/Organization | Geography | Contact name | Contact position |
|---|--|--------------------------------------|--|
| Harambee | South Africa | Yatin Nana | Ex-product Manager |
| International Finance Corporation (IFC) | US | Loretta Foran | Senior Operations Officer |
| ILSS Leadership Program | India | Isha Sharma Anuradha Prasad | Senior Program Manager, Founder & CEO |
| J-PAL | Global | Evan Williams | Senior Research, Education, and Training Associate |
| Laboratoria | Latin America | Matías Hoyl | New Ventures Director |
| Molengeek | Belgium | Ibrahim Ouassari | Founder |
| Moringa School | Kenya | Yatin Nana | Expert |
| Mozilla Foundation | US | Hanan Elmasu | Director, Fellowships & Awards |
| Namibia University of Science & Technology | Namibia | Jose Quenum | Professor in the Faculty of Computing and Informatics |
| NetHope | US | Lance Pierce | CEO |
| One Acre Fund | Multiple African countries, and India | James Mwangi | Board Member |
| Rainforest Alliance | Global | Madhusudan Iyer; Catherine Rivier | Information Technology Director; Senior Data Manager |
| Salesforce | | Neal Myreck | Vice President, Transformative Philanthropy |

| Program/Organization | Geography | Contact name | Contact position |
|-----------------------|---------------|--------------------------|---|
| Skoll World Forum | Global | James Mwangi | Board Member |
| SoCieDat | Latin America | Jesus Ramos | Vice President |
| Splunk Ventures | US | Stephanie Reitz | Senior Program Manager |
| Strathmore University | Kenya | John Olukuru | Head of Data Science and Analytics Strathmore University |
| Tableau Foundation | Global | Lauren Begg | Project Manager Operations |
| Talent Rewire | US | Nicole Trimble | Former Managing Director |
| Tech Change | US | Nick Martin | Founder & CEO |
| Tech4Good Community | India | Anusha Meher Bhargava | Co-founder & Chief Insights Officer |
| World Bank | Global | Craig Hammer | Program Manager, Development Data Group |
| Ziob Consulting | Global | Lutz Ziob | Expert/founder |

BIBLIOGRAPHY

| Title | Author | Year |
|---|---|------|
| 21st Century Skills: Transversal Skills Development in Latin America and The Caribbean | Intra-American Development Bank (IDB) | 2019 |
| Adapting To Changing Skill Needs in Southeast Asia | OECD | 2021 |
| AI Talent Insights from The European Market | LinkedIn | 2019 |
| Advancing The Field of Data Science for Inclusive Growth | data.org | 2021 |
| An Action Plan to Solve the Industry's Talent Gap | World Economic Forum, Boston Consulting Group | 2018 |
| Applying AI For Social Good: Discussion Paper | McKinsey Global Institute | 2018 |
| Big Data Research for Social Science and Social Impact | Lytras, Miltiadis., Vixvizi, Anna | 2019 |
| Building Africa's Artificial Intelligence Talent: 1 Million Artificial Intelligence Talents In 10 Years | Data Science Nigeria | 2020 |
| Building Digital Workforce Capacity and Skills for Data-Intensive Science | OECD Global Science Forum | 2020 |
| Data-Driven Decision Making for Not-For-Profit Organizations | CPA Journal | 2019 |
| Data Handling and Digital Readiness in Civil Society | TechSoup Global Network | 2020 |
| Data for Better Lives | World Bank | 2021 |
| Data Science in The New Economy | World Economic Forum | 2019 |
| Data Skills Training Program: Case Studies | Federal CDO Council | 2021 |

| Title | Author | Year |
|--|---|------|
| Digital Skills Insights 2020 | International Telecommunication Union (ITU) | 2020 |
| Digital Development Training Landscape and Needs Assessment: Phase 1 & 2 | Principles For Digital Development | 2019 |
| Dynamics Of Data Science Skills: How Can All Sectors Benefit from Data Science Talent? | The Royal Society | 2019 |
| E-Conomy Africa 2020 | Google, IFC | 2021 |
| Global Innovation Index 2020 | INSEAD, Cornell SC Johnson School Of Business And World Intellectual Property Organization (WIPO) | 2020 |
| Global Skills Report 2021 | Coursera | 2021 |
| How To Attract and Retain Top Talent at Not-For-Profit Organizations | ICC Partners | 2020 |
| Impact And Implementation of Data Science for Social Good Projects | Goerg, Jennifer | 2019 |
| Investigating The Data Science Skill Gap: An Empirical Analysis | Krogstie, John, Mikalef, Patrick, | 2019 |
| Investment Appraisal of Upskilling of Employees in The Non-Profit Sector | INDECON International Research Economists | 2020 |
| Kenya Strategy for The Development of Statistics | Government Of Kenya | 2019 |
| Labor Market Analysis and Curriculum Gap Assessment Using Big Data in Kenya | Headai Finland On Behalf of The World Bank | 2019 |
| Lead With Data: How to Drive Data Literacy in The Enterprise | Qlik | 2018 |
| Leap Before You Lag: Nonprofits with Deeper Data Capabilities See Stronger Impact, Transparency, And Decisions | IBM Institute for Business Value | 2017 |

| Title | Author | Year |
|---|---|------|
| National Policy for Digital Transformation and Artificial Intelligence: Columbia | Government Of Columbia | 2019 |
| National Strategy for Artificial Intelligence: India | Government Of India | 2021 |
| Non-Profit Trends Report | Salesforce | 2020 |
| Repurposing NGO Data for Better Research Outcomes: A Scoping Review of The Use and Secondary Analysis of NGO Data in Health Policy and Systems Research | Masefield, et al. | 2020 |
| Skills For The Future: Asia Pacific | Brand Future Skills | 2020 |
| Stanford Institute for Human-Centered Artificial Intelligence Annual Report | Stanford Institute for Human-Centered Artificial Intelligence | 2020 |
| Social Sector Data Maturity Framework | DataKind/Data Orchard | 2017 |
| State Of the Coding Bootcamp Market Report | Career Karma | 2020 |
| State Of the Sector Data Maturity in The Not-For-Profit Sector | Data Orchard | 2020 |
| Tech Savvy: Skills for the Digital Economy: Report On Experimental Qualitative Research | Cenfri | 2019 |
| The Data for Good Growth Map: Decision Points for Designing a University-Based Data for Good Program | Dailey Et Al., University of Washington, Escience Institute, Data Science for Social Good | 2021 |
| The Data Literacy Index | Qlik | 2018 |
| The Digital Nonprofit Ability Assessment | The Center for The Digital Nonprofit | 2018 |
| The Future of Work in Africa Harnessing the Potential of Digital Technologies for All | Africa Development Forum | 2019 |

| Title | Author | Year |
|--|--|------|
| The Great Data Literacy Gap: Demand for Data Skills Exceeds Supply | Forrester Consulting on Behalf of Tableau | 2021 |
| The Human Impact of Data Literacy | Qlik, Accenture and The Data Literacy Project | 2020 |
| The Path to Becoming a Data-Driven Public Sector | OECD | 2019 |
| The Skills Balancing Act in Sub-Saharan Africa Development Forum Investing in Skills for Productivity, Inclusivity, And Adaptability | Agence Française De Développement And The World Bank | 2019 |
| The State of Data in The Nonprofit Sector | Everyaction And Non-Profit Hub | |
| The State of Public Sector Digital Transformation | The World Bank | 2021 |
| The Upskilling Imperative - Building A Future-Ready Workforce For the AI Age | Deloitte | 2020 |
| Upskilling For Shared Prosperity | World Economic Forum | 2021 |
| Youth Jobs, Skill, And Educational Mismatches in Africa | African Development Bank (AfDB) | 2019 |
| How To Attract and Retain Top Talent at Not-For-Profit Organizations | ICC Partners | 2020 |
| The European Data Market Monitoring Tool | Data Landscape EU | 2020 |
| The Data Scientist Shortage In 2020 | Quant Hub Data Skills Platform | 2020 |
| Gartner Top 10 Trends in Data and Analytics For 2020 | Gartner | 2020 |
| 10 Charts That Will Change Your Perspective of Big Data's Growth | Forbes | 2018 |
| Global Data Shows Surge in Remote Work | LinkedIn | 2020 |

| Title | Author | Year |
|---|---|------|
| Global NGOs and Charitable Organizations Market Report 2020: Market Is Expected to Decline From \$255,705.9 Million In 2019 To \$253,336.3 Million In 2020 - Forecast To 2030, 2021 | Global Newswire | 2021 |
| Investment In Upskilling Could Boost Global GDP By \$6.5 Trillion By 2030 | World Economic Forum | 2021 |
| Machines Will Do More Tasks Than Humans By 2025 But Robot Revolution Will Still Create 58 Million Net New Jobs In Next Five Years | World Economic Forum | 2018 |
| Gartner Says More Than 40 Percent of Data Science Tasks Will Be Automated By 2020 | Gartner | 2020 |
| Coding Bootcamp Market Size and Forecast | Verified Market Research | 2021 |
| Online Data Science Training Programs Market 2018-2022 | Technavio | 2018 |
| Why The World Needs More Women Data Scientists | Center For Global Development | 2021 |
| The Data Scientist Shortage In 2020 | Quant Hub Data Skills Platform | 2020 |
| The Charitable Sector | Independent Sector | 2019 |
| The Nonprofit Sector in Brief 2019 | Urban Institute – National Center for Charitable Statistics | 2020 |
| Quantifying The UK Data Skills Gap | Gov.UK | 2021 |
| AIMS: Strategy Review and Validation | Dalberg On Behalf Of AIMS | 2021 |
| Advanced Software Development Skills in Sub-Saharan Africa | Dalberg On Behalf of The Digital Impact Alliance | 2018 |
| Data Science for Social Impact: Value Proposition, Partnerships, And Investments | Dalberg On Behalf of data.org | 2019 |
| Developing Advanced Software Development Skills in Sub-Saharan Africa: Program Design | Dalberg On Behalf of The Digital Impact Alliance | 2019 |

| Title | Author | Year |
|---|---|------|
| Facebook Digital Skills Africa | Facebook And Dalberg | 2020 |
| Genu Franchise Accelerator Strategy | Dalberg On Behalf of Genu | 2021 |
| Moringa School – Advancing Outcomes Research & Strategy | Dalberg On Behalf of Moringa School | 2020 |
| National Skills Development and Employment Promotion Strategy | Dalberg On Behalf of The Government of Rwanda | 2019 |
| Private Sector Engagement Strategy | Dalberg On Behalf of Ajira Digital Clubs | 2020 |
| USAID Digital Stakeholder Partnership Assessment | Dalberg On Behalf Of USAID | 2021 |
| Decoding Diversity: The Financial and Economic Returns of Diversity in Tech | Dalberg And Intel | 2016 |

data.org



Patrick J McGovern
FOUNDATION

Dalberg