

So, you want to hire a data person?

- Pointers on finding, building, and retaining data expertise in social accountability organizations
- Elizabeth Dodds



— This brief was prepared by Elizabeth Dodds with support of Michael Jarvis.

— **Acknowledgements:** TAI would like to thank all those interviewed for their generosity and time in sharing their stories, insights and ideas. We would also like to acknowledge TAI members and partners for their valuable feedback and perspectives that helped to improve the final product.

— **For more information contact:**
Transparency and Accountability Initiative
OpenGov Hub, 1110 Vermont Ave NW #500
Washington, DC 20005

— **www.transparency-initiative.org**

Copyright © 2020 Transparency and Accountability Initiative
This work is licensed under the Attribution-NonCommercial-No Derivates 4.0 International (CC BY-NC-ND 4.0). Sharing in its original format and with proper attribution is highly encouraged.

— **Citation:** Dodds, Elizabeth. (2020). So You Want to Hire a Data Person?: Pointers on Finding, Building, and Retaining Data Expertise in Social Accountability Organizations. Washington, DC: Transparency and Accountability Initiative.

Report designed by TAI. Cover photo courtesy of [Luminate](#) and their grantee [Data & Society](#).

Table of Contents

Should you hire a data scientist?	2
Understanding various data roles: the differences between analysts, engineers and scientists	5
What data role(s) make sense given existing data/tools/systems/resources?	5
Once hired, how can data experts be set up for success?	8
What could this look like in practice?	10
Ideas for funders	11

Context

The growing volume of available data has created enormous opportunity for public, private and non-profit actors to push for accountability, create better products and services, and improve decision-making and overall impact. To take full advantage of the opportunities afforded by open data and big data analytics, for-profit and non-profit organizations alike are seeking to ramp up their in-house data expertise, which has contributed to a frenzy of hiring data science roles in recent years. Funders are fielding a growing number of requests from grantees for support in hiring data experts, sometimes without a clear definition or justification for what the role might entail. In the social accountability field, the challenge is understanding the types of data expertise that would be most beneficial and how best to harness the talents of data experts once on board.

In line with the Transparency and Accountability Initiative (TAI)'s goal of supporting funders to improve the targeting and collective impact of their grant-making, TAI is offering guidance to funders and grantees on what to consider when seeking to build in-house data expertise. The following note shares insights drawn from social accountability organizations that have hired or are seeking to hire data experts, as well as from the experts themselves, on what has worked well and not so well, and what could have been done better. It outlines the kinds of data expertise available, how to establish and meet internal data needs (whether through in-house or outsourced data experts or a combination of the two), and once hired, how to set data experts up for success.

Should you hire a data scientist?

If you are a social accountability organization, probably not – although this depends on the state of your organization’s data, capacity, resources and needs, and what is meant by the term “data scientist.” Sometimes organizations assume they need a data scientist, when in practice they may need someone more in line with a data analyst. Now organizations also have access to a wide array of tools and platforms that can provide the functions of a data scientist, but more efficiently and at a lower cost.¹ While hiring a data scientist is likely not necessary for many of these organizations, **building data expertise** remains a critical undertaking in order for organizations to better leverage data in improving decision-making and the products and services they offer. (See table below for a breakdown of specific roles).

In-house or outsourced expertise?

One approach is **outsourcing data expertise to consultants, fellows, volunteers or through partnerships**, particularly for one-off, exploratory work or to build specific automated tools or platforms. Even in cases where data experts are hired in-house, organizations can benefit from external support to fill capacity gaps. When it comes to advanced data science techniques and technologies, organizations may gain more from partnering with companies or experts that have already built up expertise in these areas – for example, partnering with the non-profit arm of an AI organization. While this can be helpful to bring in creative ideas and approaches, a few interviewees noted that building internal staff capacity on data should remain a priority.

One interviewee from a social accountability organization partners with a data science organization to do more advanced analytical work. In this case, partnering lends further credibility to her data work and could be helpful in convincing senior leadership of the value of data science (and potentially, of the need to hire in-house). While partnerships don’t always help to build internal capacity, in this case it did: the interviewee became a visiting analyst for the partner organization, working directly with data scientists and learning from them. In her words, this worked because she could “speak the same language” as the data scientists and therefore act as a bridge between them and the content experts in her organization.

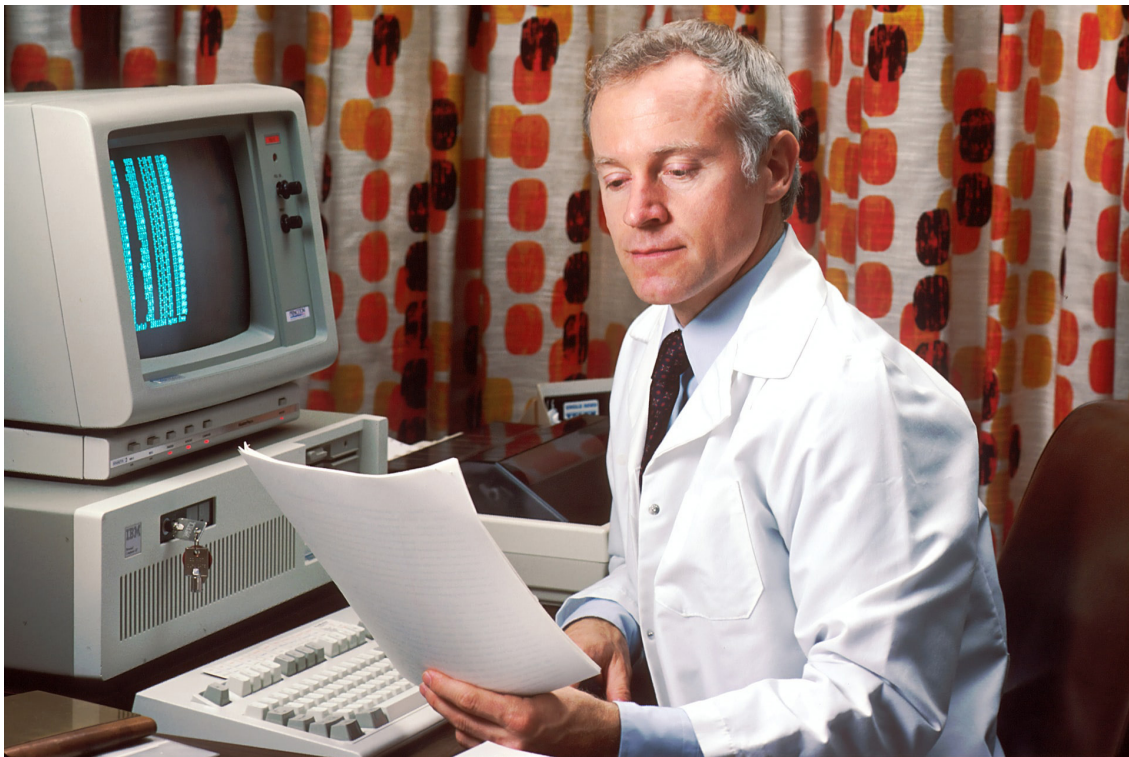
There is also the option of working with an organization such as **DataKind**, which brings volunteer data scientists, developers and social change organizations together through short- and longer-term engagements. These range from weekend-long data dives into a particular problem, to customized support on non-profit projects. Project support can be provided to create a model, system, tool or approach that the organization can use on an ongoing basis without the need of a high level data expert (for example, a machine learning model to better target interventions). Support is generally provided pro bono. Organizations may benefit from external support when they have a clear question or problem and a dataset to address it, which can easily be handed over to a consulting data scientist or other type of expert.

1. Goldman, J. (2018, July 18). Whatever you do, don't hire a data scientist (here's why). Inc.com. <https://www.inc.com/jeremy-goldman/why-brands-agencies-brands-shouldnt-hire-data-scientists.html>

If hiring in-house, what data roles might be available?

In an ideal world, organizations would be able to hire a team of data experts, including data wranglers, analysts, engineers as well as senior-level experts and data scientists, if needed. Practically speaking, however, funding in non-profits may only cover one or two data positions (or none at all). So who should organizations hire and how can they meet their data needs with limited funding?

Data positions in non-profit organizations may not be clear-cut and could encompass a number of different roles requiring a certain degree of flexibility. Nevertheless, it is helpful to understand distinctions between various data roles and capabilities to know what might align best with the organization's needs and to be able to prioritize and define the position to fill. The table below outlines some common data roles and the activities and competencies that are most closely associated with them.



Courtesy of the National Cancer Institute on Unsplash.

Since the 1980s, the skillset and stereotype of the data scientist has changed a great deal

Understanding various data roles: the differences between analysts, engineers and scientists

Professional Role	Activities	Knowledge and core competencies
Data wrangler (non-technical term): Collecting better data and building datasets ready to be leveraged for downstream analytics	Data "munging", including finding and extracting data, scraping, structuring, cleaning, validating, transforming/ parsing raw data from one format to another, joining/aggregating, and/or standardizing to generate structured datasets. This may involve traditional methods using excel, hand-coding solutions, or using data wrangling products from other companies such as Trifacta and Datawatch Monarch	Expertise in Excel and programming languages such as Python, SQL, R, among others, as well as statistical software such as SAS Some level of contextual awareness to be able to interpret and transform the data in the most relevant and accessible way
Data engineer: Creating big data pipelines and software solutions around big data	Back-end: Streamlining how data is coming in and connecting to each other: developing routine or automated approaches and systems for data collection, cleaning, processing, organization, storage, updating, maintenance, security, backup, and either routine or automated integration of data between systems (being able to cross-reference databases); deploying machine learning and statistical models Front-end: Building databases/warehouses/repositories, interactive data visualization tools	Advanced programming skills (generally in Java, Scala or Python), specialization or emphasis in distributed systems, ability to create and integrate APIs, experience building data pipelines Understanding of various technologies and frameworks, knowing which ones to use when, and how to combine them to create solutions
Data analyst: Interpreting the data	Analyzing, visualizing and/or interpreting the data (including descriptive, diagnostic, predictive and/or prescriptive analytical techniques). This could range from basic data analysis using spreadsheets and basic visualizations to more advanced forecasting and predictive and prescriptive analytics	Expertise in Excel and in programming languages such as SQL, as well as comfort with other languages such as Python or R; data management and visualization skills. Advanced math or statistics degrees may be required depending on level of analysis.
Data scientist: Extracting useful insights from the data and applying mathematical models to guide future decisions and predict outcomes	Exploring and understanding the data; Designing and building machine learning algorithms and deriving insights from various data sources using statistical methods and machine learning models to identify patterns and underlying relationships between variables and make predictions; Undertaking strategic planning for data analytics	Advanced math and statistics background; experience creating machine learning algorithms; some programming ability (knowledge of Python, R, SQL, and others too such as Java, Scala, MATLAB, etc.) to be able to apply mathematical models in their analytics; May require a more advanced understanding of deep learning and AI.
Senior-level advisor/lead: Informing or influencing decision-making using data analytics	Interpreting/translating, visualizing and presenting analytical findings to inform and/or influence non-technical audiences, including internal management, partners, wider field, etc. Could also involve overseeing and advising data experts.	Combination of high-level subject and technical knowledge (understanding of underlying data techniques); Ability to "verbally and visually communicate complex results and observations in a way that the [organization or audience] can understand and act on them" ²

2. Vallaey, M. (2019, April 16). Why a data scientist is not a data engineer. Big Data Matters Blog. <http://info.bigindustries.be/analogue-cloud/why-a-data-scientist-is-not-a-data-engineer>

Given how new these professions are, they lack standardization, with some amount of fluidity and overlap between them. This can lead to misplaced expectations of what a given role entails. Most data scientists will have some experience with data wrangling, programming and big data engineering, but may be less efficient and less satisfied in a job where they are required to spend a lot of time doing this. Data wrangling in particular (i.e. preparing the data for further analysis) tends to take up a large chunk of time spent on a project. Some people may have a unique combination of skills across these positions – for example, a “machine learning engineer” might be highly proficient in both data engineering and data science – but they are relatively less common and likely to be highly sought after.

What data role(s) make sense given existing data/tools/systems/resources?

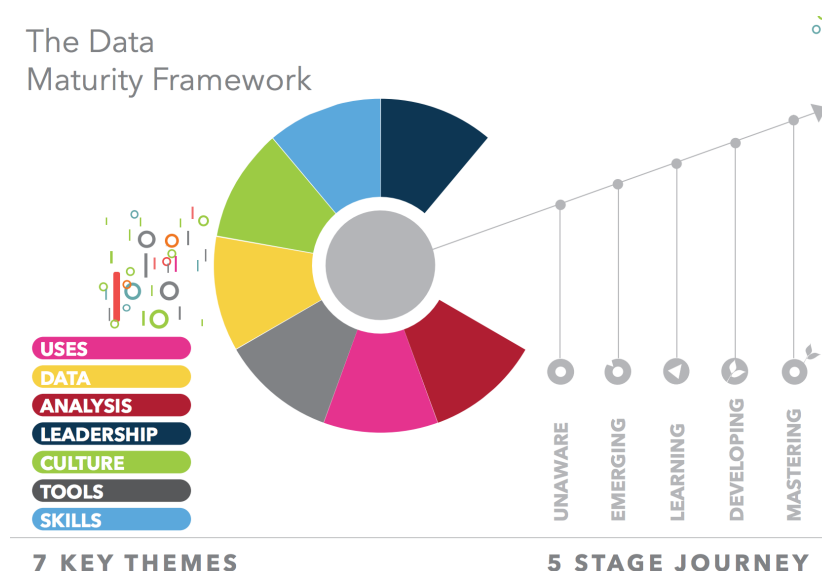
Non-profit organizations often struggle with defining their needs in terms of data expertise, and how to go about recruiting such positions and preparing for their integration in the organization’s strategy and operations.

A fundamental first step is **defining what the organization could answer or solve with data**. While this seems obvious, according to an interviewee, organizations often lack “a nuanced understanding of {the function and value of data} and how that fits in the organization.” Many senior-level positions in these organizations do not have a data background, and therefore may be unable to envision how data could be used to solve problems. Some examples of data-driven goals include:

- Internally: using data to meet compliance standards, inform strategy or planning, improve performance or targeting of services/interventions, and/or build internal knowledge and expertise;
- Externally: using data to create data repositories and analysis as a public good for the field, build an evidence base for credibility or influence, contribute to public debate, influence policy decisions, and/or improve communications.

Achieving these goals may or may not be possible, depending on the nature of the organization’s data, systems, and tools.

First, for some organizations, it may be helpful to enlist external consultants or partner organizations with data expertise to help build an understanding of the power of data and to broaden the thinking of data use beyond descriptive statistics or basic accountability tasks. Consultants or partners could also be leveraged to help the organization take stock of “**how advanced (the) organization is in the way it manages and uses data.**”³ Datakind UK and Data Orchard, which both support social sector organizations to use data better, teamed up to design a **data maturity framework** for the non-profit sector. This outlines levels of maturity when it comes to working with data, ranging from “unaware to mastering.” The assessment is based on the data an organization has or needs as well as data uses, analysis, resources, culture, tools and skills. Generally, the less data mature an organization is, the less likely it is that they would benefit from data science tools and techniques.



Source: Data Orchard. (2017). Data Maturity Framework for the Non-profit Sector.

One part of assessing data maturity is understanding the **kind and quality of data you currently have** and what questions can be asked or answered with it, and any data you may need to acquire. For example,

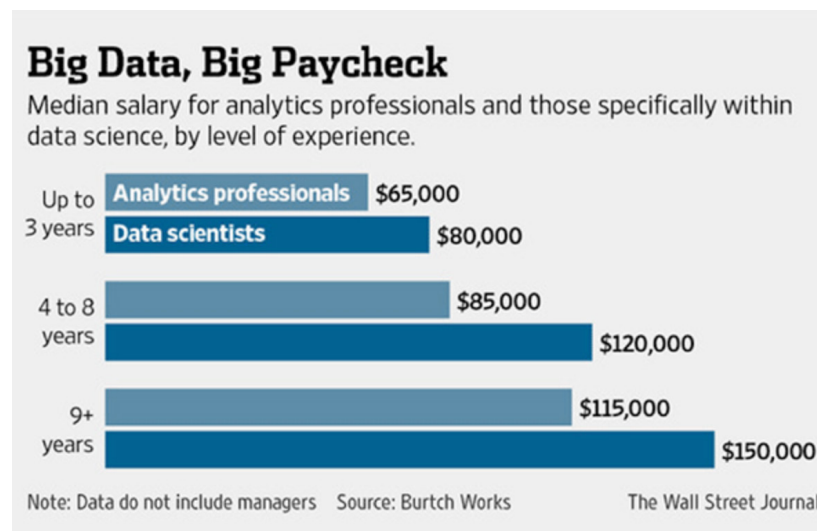
- **Is your data actually “big data”?** Applying big data technologies and skills to performing small data tasks would be overkill and would likely overcomplicate the analysis. Even in cases when the data is sufficient for data science techniques, **advanced analytics might not even be a priority.**
- **How much upstream work is required before the data is ready to be analyzed?** Among non-profits, data quality is often a challenge and the data available are generally not ready for analysis, let alone more complex modeling. Generally speaking, non-profit organizations tend to have smaller and messier datasets than corporations, and may therefore be better served hiring data wranglers and analysts than big data engineers and data scientists.

3. Symons, T. and Prest, E. (2017, Jan. 26). The data evolution: New tools to help organizations get more from their data. Nesta. <https://www.nesta.org.uk/blog/the-data-evolution-new-tools-to-help-organisations-get-more-from-their-data/>

The kind and quality of data and what you want to do with it should also be used to determine the **systems and tools** that are needed to support the data expert.

Interviewees broadly agreed that **the data skills most social accountability/non-profit organizations need will not fit the profile of a data scientist**. There may not be enough data, or enough quality data to jump right into analytics, and most of the data work likely doesn't hinge on advanced machine learning techniques. As one person noted, data scientists coming in to the organization "would be quite shocked, expecting clean, large datasets; in this case {it's} not a good fit" and "a data scientist will not have anything to do." Meanwhile in many NGOs, "the person you need is the person to help collect better data" and "who can build platforms to collect and structure" the data in preparation for analytics down the line – in other words, data wranglers, analysts and/or engineers. Another interviewee noted that "programming skills have also come on hugely once we've figured out what the organization needed."

There are also critical cost considerations. **Data scientists are expensive**, even in the non-profit field. According to Glassdoor, the average annual salary of a data scientist in the U.S. is \$113,000 (\$99,000 in the non-profit field), with senior data scientists making an average of \$134,000 a year. Data analysts can expect to make just over half of that (an average of \$62,000/year, or \$54,000/year in a non-profit organization). This jumps up significantly with higher levels of expertise, as demonstrated by the (somewhat dated) graphic⁴ below. However, it is important to note that this will vary significantly based on country context and availability of expertise.



As interviewees noted, the **data expertise needed by non-profits is not only technical**. As one data scientist indicated, the approach to hiring data experts in non-profits should be different from the corporate world (where there is generally more clarity on the problem and approaches to solving it). In many non-profits, data roles will likely need to be more creative and collaborative, involving talking to people throughout the organization to understand

4. From: Waller, N. (2014, April 9). Get Familiar With Big Data Now—or Face 'Permanent Pink Slip'. The Wallstreet Journal. <https://www.wsj.com/articles/get-familiar-with-big-data-now-or-face-permanent-pink-slip-1396999456>

what they do and to define what the problems are, and then pitching solutions on how to address them. In addition to data and software capabilities, certain “softer” skills were highlighted as being just as (if not more) critical to getting value from data – such as creative thinking and problem-solving, ability to work collaboratively, having a flexible/growth mind-set, being able to communicate persuasively to a non-technical audience, as well as investigative skills. Having a firm grasp of broader issues such as **data ethics and data privacy** can also be hugely important to inform the organization’s wider approach to data.

There was a general consensus among interviewees that **communication skills should be prioritized** so that the data expert can effectively explain what it is they’re doing and convince and influence others on this basis. To ensure data work leads to actual change or lasting impact, it is important for data people to have the confidence and credibility to be able to convince senior leadership of their findings. It was noted that this can be a challenge in non-profits where the data roles might be more junior and communication is less prioritized over technical ability. For these roles, it is also important for the data person to “feel empowered to say no” and to be able to communicate why certain approaches or goals related to data are unrealistic.

Once organizations have defined the skills they need and the position they are seeking to hire, an added challenge is **recruiting the right people**. According to a data scientist, organizations may not be “speaking the right language or asking for the right skills.” For example, the organization might need five or six skills and “end up hiring somebody who can do skills three and four, but {it turns out they} might really need skills 1 and 2 but don’t realize until it’s too late.” Part of the challenge may be in setting expectations of what the job will entail, the data maturity of the organization, and how the employee would spend the majority of their time. While there may be a tendency to entice people to join the organization by offering the prospect of exciting and meaningful data science work, if the organization lacks data infrastructure and there is no data pipeline in place, this should be clearly communicated during the recruiting process. This lack of expectation setting may be one reason many data experts leave their jobs so frequently. Again, the use of external consultants or partners with data expertise can be helpful here not only in translating data needs into clear and meaningful language for data experts, but also in leveraging their existing networks of experts in order to recruit the right people.

Once hired, how can data experts be set up for success?

It may be helpful to **have some initial flexibility** in how data roles are defined and integrated into the organization, to allow them to evolve based on a growing understanding of data needs and challenges. However, some non-profits may not be prepared to onboard such positions. As one interviewee noted, “standard procurement or the hiring process isn’t designed for new positions that are less rigid and more flexible” and that might not fit the standard definitions for employees. This may require new approaches or creative thinking to bypass these restrictions.

Although some flexibility might be needed, it is important to know how data expertise will be utilized within the organization and to **define goals and expectations for the role**. As one data expert pointed out, some organizations will hire “the data person and have no idea

what that means.” This stems from a “lack of exposure to what {data experts} can do in these organizations and the impact they can have.” They emphasized that “if you’re going to hire a data person, define what they {will} do. Data is used across the whole organization. {There are} going to be a lot of competing demands on that person’s time.” Since the role is often not very well understood, the data person may end up functioning as the “IT team,” or this can lead to situations where data experts are either underutilized or spread too thin across the organization.

It is also important to **manage expectations of how the data role will function throughout the organization**. With limited resources, organizations may only hire one data person and expect them to be responsible for any and all data-related work. As noted above, however, it can be difficult to find someone who is willing and able to perform a wide range of data tasks. Another challenge highlighted by interviewees is that the data role is often isolated and not expected to function collaboratively. In their words, the thinking among other staff often is: “I’ve got all this data, can you do something really cool with it?” Instead, there should be two-way communication between data and content experts to better understand their questions and the challenges to be addressed, and if and how data might be used to address them.

For the data expert or team to function most effectively, this would “involve wider change that organizations are not {usually} ready for,” according to one interviewee. This includes **building buy-in and a broader understanding of the role and value of data throughout the organization**, and evolving the organization’s strategy and approach to how data is incorporated into overall decision-making. As several interviewees noted, to do this effectively requires “building up the general data literacy {of the organization}” to be able to “collaboratively interrogate a dataset to ask a question you want to answer.” Understanding which data capabilities to grow across the organization will require some prioritization of certain data functions based on how integral they may be to the day-to-day work of the organization. It was mentioned that one function of the data expert could be to help guide this internal thinking and capacity building, so that the data expert is not expected to shoulder all data-related tasks.

Given that non-profits may not have the resources to hire more senior data positions, it was mentioned by those interviewed that more junior data staff may be on their own within the organization with little guidance and few opportunities to learn more advanced data skills. One data scientist emphasized the value of **providing junior data staff with mentorship opportunities** (internally or externally) to ensure a degree of quality assurance in the organization’s data work, provide added credibility, help translate the data findings and persuade senior leadership. Senior-level advisors can also help organizations bring the right people on board, by helping to define and translate organizational data needs and connecting them with the right data experts.

What could this look like in practice?

In one of the social accountability organizations we spoke with, the process of hiring and integrating a data expert into the organization involved a significant amount of learning on the part of the organization as well as the data person. Some lessons from this include:

- **Flexibility and adaptability are essential.** Initially, the organization did not have a position dedicated to digital or data work. When the data lead was first hired, this position covered a lot of ground for the organization: as a data analyst, this involved acquiring data sources, building databases, writing scrapers, and undertaking statistical analysis to inform the research of others. Initially, the role was not well understood and the problems that it would address were not well defined. As a result, the organization wanted “wizardry and magic” with data, and lacked a nuanced understanding of how the data available could actually be used to drive decisions and outcomes. Over time, however, this became clearer, particularly once the organization put in place a data strategy.
- **Analytical work involving data science techniques (such as AI and machine learning) might best be outsourced.** For more exploratory work using data science, the organization partnered with an AI company and with DataKind. It was noted, however, that generally the problems and data they work with do not require machine learning techniques.
- **Prioritize ability to learn quickly and independently, creative thinking, collaboration and communication.** The ability of the data expert to work collaboratively with others in the organization to help define the problems to address using data, and to communicate to non-technical team members what is possible with the data available was particularly helpful. This also required an eagerness to learn about the work content as well as to acquire new skills (in this case, such as programming) through on-the-job learning.
- **Continue to expand the data literacy of the organization and the capabilities of the data person/team.** As the data work program was expanded and the role of the data expert was better defined and understood over time, another data role was created to provide technical support (i.e. scraping datasets) and allow for some division of labor for the data expert to dedicate more time to other tasks. There is also now a focus on building the internal data curricula of the organization by bringing in external experts to teach new skills and techniques, for example, intelligence investigation techniques.

Ideas for funders

Taking a step back, funders should consider how they might better address the need for data expertise field wide. Beyond approving requests from each of their grantees for data scientists or other data roles, how could funding be used to build up a cohort of data experts with the right mix of skills suited to the non-profit field and to the problems social accountability organizations aim to solve?

Funders might consider the following ways of **filling the gap in senior-level data expertise across non-profits**:

- Leveraging dataskills organizations, such as DataKind, to provide a suite of services or consulting support across a number of social accountability organizations.
- Financing the creation of a small team of part-time or rotating senior-level fellows with enough content expertise to be able to understand the problems the organizations are working to address, while also having expertise in all of the data techniques needed by the organization. This team could be jointly funded and shared across grantees to support them in defining and integrating data expertise within their organizations, as well as serving as mentors and advocates for the organization's junior data staff. The team could provide support at various stages: recruiting and hiring the right positions, the initial set-up of the data stream of work, providing guidance and quality assurance, helping the data person or team learn more advanced techniques (as needed), and helping to communicate and advocate for the data work to ensure senior-level buy-in.
- Enabling productive partnerships between well-regarded academics and evidence-based advocacy organizations to agree on data and statistical interpretation and lend credibility to social accountability organization's findings.



TRANSPARENCY &
ACCOUNTABILITY
INITIATIVE

Transparency and Accountability Initiative is a collaborative of leading funders of transparency, accountability and participation worldwide. It envisions a world where citizens are informed and empowered; governments are open and responsive; and collective action advances the public good. Toward this end, TAI aims to increase the collective impact of transparency and accountability interventions by strengthening grantmaking practice, learning and collaboration among its members.

1110 Vermont Ave NW #500,
Washington, DC 20005

www.transparency-initiative.org

@TAInitiative