SIMLab

Good Data Collaborative Consultation Report

September 2017
Laura Walker McDonald and Kelly Church

Introduction and Executive Summary

Digital tools have empowered nonprofits and civil society actors to collect, store, and process increasingly more—and increasingly more sensitive—data in the course of their ordinary service delivery. Alongside new opportunities, as data becomes more intertwined with an organization's core programmatic activities, there are new considerations: new obligations to protect and responsibly handle client and beneficiary data, and new risks, as the interaction between data and societal challenges creates new potential harms for vulnerable groups.

At the same time, national and local regulations are rapidly advancing, as governments wake up to the implications of sensitive data collection, transmission, and storage. These new regulations introduce risk to both organizations and individual practitioners, and introduces a body of law that many civil society organizations are simply unprepared to deal with, and few monitor. Institutional donors, too, are part of a complex ecosystem of sometimes conflicting pressures and demands which complicates which and how much data is captured, and what happens to it.

SIMLab, along with the Center for Democracy and Technology (CDT), Future of Privacy Forum (FPF) and The Engine Room (TER) came together as the Good Data Collaborative, and were funded by Stanford's Center on Philanthropy and Civil Society to develop resources to help social change organizations to manage data more responsibly. The project seeks to identify the gaps and challenges in responsible data (RD) resources at all levels of social change, through three strands of work. First, a literature review bringing together the core principles of RD, as expressed in many existing guidance and frameworks, into one document. Second, User Experience (UX) research on the existing ResponsibleData.io site run by TER. Third, our consultation, and an effort to bring together what we've learned to guide the development of prototype tools later this year.

This project is focussed on the needs and challenges of organizations for whom digital data is a by-product of their work. Data is not at the core of their mission or operations, but is produced as exhaust from projects - for example, household-level income data generated in the process of providing cash assistance in Niger, or information about sexual orientation gathered in providing counselling to young people in Atlanta, Georgia, USA. There are ample resources dedicated to the challenges of data handling in so-called 'big data' projects, by experts in data analysis and collection. We are interested here in a group of practitioners whose major skillsets and interests lie elsewhere, but who nonetheless find themselves responsible for what happens as they gather, handle, retain and then (hopefully) delete this data.

Our findings suggest that RD is a complex issue, and rarely handled effectively even in organizations who recognize the need to improve their compliance with RD principles. For many of our respondents, 'responsible data' is a new concept, without organizational compliance mechanisms or even broad understanding. Our interviewees see data practices as largely left to individual actors to implement, monitor and enforce; people do not know where to go for help, and even where they do understand the basic principles of RD, they express feeling overwhelmed by the complexity of implementing the required practices in their organization. In many cases, they responded to the uncertainty and discomfort by putting the

responsibility for becoming compliant on other colleagues. Many expressed concern about not properly understanding the law covering data management, and many do not publicly admit to their uncertainty. Donors and platform providers are equally challenged to provide guidance and investment in a challenging and potentially expensive area with such strong links to legal liability and capacity-building. There may be a brewing 'culture war' been data-centric and rights-based approaches to technology in social change projects which will needs open discussion. In summary, SIMLab's findings point to an ecosystem-level challenge, one that will not be solved by toolkits alone, but more likely will require institutionally-supported behavior change, individual internalization of these practices as good practice, sharing lessons learned and cooperative working between implementers and donors to change the way we think about operational data and the infrastructure needed to manage it responsibly - in short, will need time and commitment.

This paper summarizes our methodology, findings and some brief recommendations. Our primary intention was to inform the work of our Collaborative, and in particular the development of the prototype tools. We hope that the report will also provoke open and constructive discussion among those using and supporting technology in social change projects. This will be a necessary first step towards surmounting some of the challenges we uncover with responsibly handling data in practice.

Methodology

In order to develop resources for social change organizations who end up generating and managing other people's data in their work, we needed to first understand where people are, what resources they have and what is missing.

The aim of the consultation was to hear from a diverse group of people - from those who think about data everyday to those who occasionally come across data within the course of their work. This consultation sought to draw on the experiences of a small group of practitioners, policy-makers, philanthropists and platform providers to get a sense of their understandings of RD, the challenges they face, and the resources they turn to for guidance.

During the research period, SIMLab also ran an open-access survey using an online tool called Typeform. The survey was created to supplement the findings in time-intensive interviews and thus asked similar questions, but were reframed to be better suited for short answers or multiple choice. At time of writing 41 people have completed the online survey. Analysis of the responses forms part of our findings in this paper. The survey is included here as Annex 2.

We also draw on the User Experience (UX) research conducted by The Engine Room as part of this project.

During the consultation we were especially interested in the kinds of data that were being handled, and general practices around it; conversations about RD and who was driving them; organizational policies and legal regulation that the interviewees were aware of; conventional or habit-based data practices; efforts to share and learn new concepts within organizations and drivers for good and bad practices throughout the entire ecosystem.

For the purposes of our research, we adopted the definition of 'Responsible Data' on the ResponsibleData.io site:

> The duty to ensure people's rights to consent, privacy, security and ownership around the information processes of collection, analysis, storage, presentation and reuse of data, while respecting the values of transparency and openness. (_Responsible Data Forum, working definition, September 2014_)

This report is presented with due recognition of its incompleteness - the insights within it cannot be assumed to apply to all social change organizations, and should not be used to draw precise conclusions.

Selection of interviewees and survey respondents

We identified four main actors who shape the RD ecosystem:

1. practitioners, who collect or manage data at field level, or who manage field level staff from regional or head offices;

2. tool providers, who create platforms that collect, store, manage, analyze or transmit data;

3. policy experts and researchers, who think and write about data practices, and work with others to advise on RD practices, work with donors to inform policy, or simply consider implications and opportunities;

4. donors, who fund organizations which implement programs in social change work that gather and manage data. Here, we refer to institutional donors such as foundations and government departments, rather than individual donors giving direct to non-profits.

SIMLab and consortium partners developed a list of potential interviewees covering all of these stakeholder groups, aiming for diversity in organization size, sector, geography, and organization type, although we would not suggest that our interviewees are truly representative of the universe of possible RD practitioners. For example, although we sought geographic diversity, we did not manage to speak with anyone from North Africa, South and Central America, the Caribbean, Eastern Europe or Oceania. Although we were intentional in seeking to go beyond the 'usual suspects' in interviewee selection, it should be noted that everyone we spoke with is only one or two degrees removed from the Collaborative and our networks.

In inviting organizations to interview, we provided a description of our project and the types of questions we were interested in and left the interviewee decision up to the organization. After speaking with one or two donor representatives and failing to find many interviewees working specifically on RD issues, we deliberately sought out a few more respondents in this group, who might be able to shed light on philanthropic and institutional donor work in this area.

The Typeform survey is open-access and has been shared on numerous mail-groups, listservs, various slack channels and over Twitter and Facebook. As with our interviews, while we hoped to get a broad range of respondents, we suspect there is some self-selection bias in the survey results, for example, a number of respondents who found us through the Responsible Data

listserv. At time of writing the survey has been viewed by 182 unique visitors and completed in full 41 times.

Interviews

Given time and resource constraints, our research here was necessarily limited to not more than thirty semi-structured interviews. Interviews were conducted by Laura Walker McDonald (SIMLab CEO) and Kelly Church (SIMLab Project Director) over Skype or via phone call. The interviews ranged from 30-60 minutes and were all conducted in English. Several interviews were recorded and notes made later, but for most we simply took notes live. The interviews were made up of eight broad questions, or nine for practitioners.The Interview Guide is included here as Annex 1. We spoke with:

- 10 practitioners

- 6 policy and researchers

- 3 tool providers

- 6 donors

Each interview began by seeking informed consent: by reading a brief statement about the purpose of the interview, the project, the organizations involved, how the interview notes would be used, and where the transcription would be stored and when it would be destroyed. Each of the interviewees offered verbal consent and then the interview began.

All of the questions were open-ended and some questions were interpreted differently by different people. When necessary we offered clarification on the questions but were conscious to not steer interviewees in any direction. The interviewer allowed interviewees to skip questions they felt were not relevant to them or their work, or which they did not feel capable of speaking to - at times, the interviewer made this decision, particularly with non-practitioner interviewees. Some interviews were left incomplete due to time and because of the order of the interviews, the two last interview questions received fewer responses than the others.

Analysis

The analysis process consisted of reading through the notes of interviews and finding common themes, and divergent perspectives. We also compared and contrasted responses within roles as well as across roles.

Themes arising from the consultations

In the pages that follow, we report themes we observed arising from our consultation. Our reflections can only speak to the individuals we heard from, and may not be true of the entire class of people they represent. For the sake of brevity, we limit the number of caveats in the

text itself. For example, where we report that 'donors felt' a particular concern, we are referring only to those individuals with whom we spoke.

How do you feel about responsible data?

We opened our interviews with this subjective question, intended to elicit an emotional response. The responses were telling.

Across the board, our respondents agree that data presents a huge opportunity for social change work, improving service delivery and impact monitoring, among other things. But there is a tension - that data also presents a large risk, and practitioners are left to decide whether the risk is worth the benefit. This calculation is part of the way that RD is understood.

For some respondents, RD was a totally new idea, which made them feel as if they had neglected important practices in their roles. A small group (5-10% of our respondents in both interviews and surveys) simply don't feel RD is part of their job, or something they should be thinking about. For others, RD practices are a compliance issue they see bearing down on them, adding additional work. Some describe RD as simply something that needs to happen, in the same way as project management, or monitoring and evaluation, and note that parts of the practice are simple - tracking datasets in a spreadsheet, for example. One practitioner suggested that 'we might one day have an RD department as we do an HR department, and people would just have innate knowledge of what you can and can't do', but that this seemed far off.

But the majority of our respondents nonetheless expressed feeling deeply concerned and overwhelmed by RD. One respondent described it as 'not exciting', 'scary' and 'opaque'. Many had only a hazy idea of what RD means, identifying key elements and concepts, such as informed consent, but without a framework through which to describe it, or clear examples of how to implement it in practice. One practitioner described bringing up RD at a recent event and being received with silence, saying that "everyone is a little afraid they they are not compliant or have overlooked something." Many people mentioned the unknowable legal implications of their work, saying that "it's like you need to be a legal expert to know" how to follow the law in their host and home countries, and other jurisdictions they might need to consider. It appears that most people keep quiet about their lack of knowledge and compliance in this area, concerned that it might somehow be used against them legally or professionally. But even the most knowledgeable describe impossible trade-offs and 'overwhelming' density of information and complexity.

Describing current practice

Practitioners were asked to describe the type of data they gathered, and how and where it was stored and transmitted. Most described fairly simple datasets manipulated in Excel and sent around by email, and hosted on cloud-based services like Google Drive and DropBox. A few used a privately hosted server.

Many of the policy-level interviewees felt that RD was better understood, and more frequently discussed, than five years ago when the open data conversation was just beginning. Research

like ours was seen as positive, having potential to raise awareness and 'get traction'. Although RD is still not 'sexy' and still carries with it a veneer of conservatism and negativity compared to the positivity of the open data and big data movements, we are hitting a 'quality control moment' where RD has joined the ranks of basic best practice issues that data practitioners know they must proactively address. Many are developing guidelines, and although practice is still in its infancy, 'it's natural and normal that we're not going to get from one to ten in a day.'

But it was clear that, in practice, people don't know of these guidelines, or find them hard to implement or follow. Practitioners tend to 'do first, and ask questions later'. One advisor to non-profits noted that 'people come to [sites like] responsibledata.io because they are in the middle of a project and then realize they're doing something potentially really bad, or go to a talk. They realize they are holding all this sensitive data, and then begin to think more critically.' Another, previously a practitioner, admitted that she had 'literally googled how to protect stuff within the humanitarian space... sitting in a tiny [field] office watching people passing around spreadsheets containing PII [personally identifiable data].'

Many know, or fear, that real harm is being done by poor data management practices, particularly at field level. One practitioner described data collectors in one project having been put at risk by a data breach, but many hinted at known harms or potential harms.

Several people mentioned that 'something bad' happening is what will drive real change in this area, that a 'worst case scenario' to point to would be helpful for them to encourage colleagues to follow RD principles. One respondent said that exposées in a major newspaper or site would be very influential.

Systems

Lack of guidance and the inability of organizations to follow available guidance was a pervasive theme in the consultations. Practitioners feel they are missing a great deal of the information and policies that they need, and whatever information they have been provided is not enough, might be outdated, followed (and known about) by only a few, or in some cases is counter to RD principles. One practitioner noted that her organization has not done enough to make things easy on practitioners; that there is not enough guidance or compliance, nor any person who is in charge of RD internal policy, training or introduction. Another in an organization without a policy described the onus being on the person handing the data at that time to manage it 'appropriately' but that the focus is on "getting results back to people quickly, not safely."

Among staff from organizations with 100 employees or fewer, only 33% report having a responsible data policy, while 79% of staff from organizations with 101 employees or more staff report positively. Smaller organizations therefore may be less likely to have a responsible data policy. However, although 37% of our survey respondents are based at company headquarters, and 42% come from large organizations with more than 501 staff, 37% of all survey respondents stated they do not have a responsible data policy at their organization, and 13% are not sure if they do or don't. In some cases, even staff from organizations that had put out best practice guidance for the field reported not having, or not being aware of, a responsible data policy. Lack of knowledge seems to lead to inaction: because people don't know everything about RD, they don't do any of it.

It's worth noting that very few people reported complying fully with their organization's RD policies. None of the people who did self-report full compliance were field-level staff collecting or in charge of handling data themselves. Of the 50% of survey respondents who reported having a data policy in their organization, only 26% report 'absolutely' complying with their organization's data policy. Some explained that they circumvent policy because it is out of date, incomplete, or not practical in the given context; others pointed to time constraints, or said that they can't comply because systems and resources are stored on the internet or intranet, when practitioners not always able to physically access the internet or even computers.

In emergencies, lack of time and a 'humanitarian imperative' override RD concerns. One practitioner stated that some vendors may not follow their advertised policies during emergencies, and that her organization simply doesn't have time to investigate.

In the absence of a policy which expressly encourages deletion or minimalism, the tendency seems to be to retain as much data as possible for as long as possible. Several organizations we spoke with explained their policies are to keep data on file for as long as possible or to 'never delete.' When asking one organization whether they have a policy of when to delete data, the practitioner stated "that's a new concept...it's important to keep information, you can't know that you won't have another partner that wants it, so you keep it." One practitioner who sometimes publishes results based on simple datasets described anonymising them, but said there was no policy or time limit around retention of data.

Working with partners

It seemed that many respondents did not typically consider what happens to data once it left their hands, for example, when hosted by platform providers or handled by partners: "we follow policies, not sure though what our partners do. I guess they could do anything they wanted." One practitioner asked; "how do you understand what a provider is really doing with your data?"

In most conversations with practitioners there was some uncertainty about how others in the organization dealt with data, so that even if the practitioner said he or she was compliant there was often a question about whether village or field-level offices were, and whether staff from those offices were even aware of the policy and its implications. One practitioner noted that they do have a policy at the field office for destroying data after 7 years but was "unsure of what that looks like, not sure there's any guidance on it," and suggested "maybe the finance team would know."

In projects where different agencies and service providers fulfill different roles, this challenge is compounded. Increasingly, service providers are stepping in to handle discrete portions of a project where storage, analysis, sharing and transmission of data is replicated many times over between different staff members, consultants and organizations. In order for the data to be protected, every person touching the data needs to have a shared understanding of RD and the same policies in place. For instance, if one organization's policy is to not publish names, but photos are allowed, and another organization published names but not photos and the two data are compared side by side, suddenly the full picture is available to anyone who wants it.

What drives data management practice?

Ethics versus compliance and the law

There is a clear carrot and stick dynamic in the way that RD practice moves up the priority list to a place where it becomes institutionalized.

Many respondents grounded their definitions of RD in human rights - people's rights in their own data; respect for them, and for people's ability to consent to its use for specifically the purpose expressed at the time of asking. They talk about it as an extension of their organization's commitment to human agency and human rights, and a logical thing for them to do. Others with a more humanitarian lens, saw it as a protection issue, related to 'do no harm' and power: concerns that we, as implementers, are putting people at risk through bad practice and poor risk modelling. But these same practitioners described feeling helpless, under-resourced, and unsure of how to fit RD into their existing practices. Because people do not immediately feel negative consequences for bad RD practice, respondents noted, it was easy to let it continue - although there might be longer-term negative consequences that have not yet been realized.

Yet, much of this practice is covered by the law, if not in the host country, then in the home country of many NGOs, or the countries in which their platforms and data are hosted or their donors are located. Most respondents were not aware of the law covering RD.

For example, many European NGOs are currently working to understand how the impending entry into force of the European General Data Protection Regulation (GDPR) in May 2018 will affect them. In the UK, a group of NGOs are working with the Information Commissioner's Office (ICO) to understand the implications of the new law, and how to become compliant with it. It seems likely that the GDPR will affect organizations elsewhere whose projects touch Europe, through funding, transmission or storage of data, or partnership with European entities. But no non-European respondent mentioned the GDPR.

Confusion over legal liability is also clear when we compare accounts from practitioners and platform providers - the organizations who build and maintain the technology tools that many tech-enabled social change projects rely on. Practitioners seem to largely trust service providers and assume that they will take proper care of the data and make appropriate decisions about issues like hosting and security. While many of the service providers we interviewed described following requirements set by their customers, one noted that, in the absence of directives from them, they did not consider themselves liable for compliance with the law or for harms that might arise from misuse or mishandling of the data. They rely on their terms of service to shield them from liability, and consider practitioners responsible for the consequences of their design decisions. This raises the prospect of platform providers and their clients each believing that the other is following some unstated responsible data best

practice, or ensuring compliance with the law - when in fact, neither is. Liability for breaches here may be shared, and there is little guidance available.

One interviewee pointed out that there are very few specialist lawyers in the field, who can operate at the intersection of international and domestic law, technology and social change, to provide appropriate and meaningful advice;

> *'most organizations don't invest in programmatic lawyers. Traditional law firms tend to be risk averse and expensive; general counsels often focus on an organization's transactional needs. There are very few lawyers with expertise in advising development and social good organizations at a programmatic level, or who understand the social impact of information architecture at a design level.'*

They argue that using terminology like 'responsible data' rather than pointing to concrete legal requirements allows implementers to adopt a softer and more forgiving stance than is allowable under the law, in many cases allowing organizations to give themselves a pass for failing to fully implement when in fact, this failure could be viewed pretty starkly.

One thing is clear - the majority of interviewees make clear that they see ethics as are a primary driver of their obligation to practice RD, and were careful to distinguish this from compliance with legal requirements. However, it's clear that, with very few exceptions, only those NGOs who are conscious of regulation like GDPR are working towards full institutional implementation of RD with any alacrity.

It seems that both carrot and stick - both ethical considerations and external compliance mechanisms - are required to get this issue on the agenda of senior leadership, and without that, RD is not institutionalized.

Donors, compliance and data positivity

Donors are seen as having the most power to drive good practice. One platform provider told us that a client was 'religious' about following USAID's data privacy assessment, once it was released. Another referred to donors 'slapping RD in all their agreements' as a positive thing that would drive improved practice. Where donors ask about RD practice, it helps to 'capture the ED's attention' in a context where organizational to-do lists are endless. Respondents felt that requiring information on RD practice as part of due diligence would have a huge impact, as all levels of an implementing organization would need to be involved in a conversation and subsequent action around the practices. They pointed to other areas of development such as gender, accountability and feedback mechanisms as evidence of the success of top-down compliance.

But, it is clear from our interviews with them that donors are not always aware of the impact that their attention or inattention in this area can have on practice. Some donor respondents

told us that no grantee had raised the issue with them or asked for help. Several did not themselves have expert RD capacity and felt unsure of how to provide guidance. Several donor respondents conflated RD with digital security issues, in one case referring us to an expert on an unrelated state-level cybersecurity field.

And while some donors, such as USAID, are investing in RD and related codes of conduct such as the Principles for Digital Development, there are tensions with expectations of broad and detailed data collection and sharing under grantee quality and accountability mechanisms. Respondents described donors asking for 'verifiable' data as part of their value for money and quality monitoring, which by its very nature can be PII, or carry with it a risk to data subjects. One former implementer told us that the pressure to provide granular data for accountability, decision-making or just 'openness' was overwhelming, 'but it's hard to make sure it won't get leaked, hacked or casually shared. And there are pressures to collect more and more, monetize data, and use data to make programming sustainable.'

Another commented that "if you don't collect [data] you aren't filling your mandate...data becomes the jewel you don't want to give up, but you don't know what to do with it [or] how to use it." While RD practice advises data minimalism and carefully planned and purposeful data collection, it seems that this is not yet culturally ingrained in all stakeholders in the data ecosystem.

Some institutional donors are taking their investments in a different direction, expressing enthusiasm about the opportunity they see in monetizing data about clients, service users or aid recipients as part of poverty reduction work, as long as the data relationships featured 'trust and transparency'. They highlighted services aimed at low-income people that rely on personally identifiable data, such as mobile data usage, to generate benefits such as credit histories and access to low-interest microloans.

Money and time

The clearest factor impinging on respondents' ability to do RD well was human and financial resource. Frequently we were told that while RD is understood as important, there is no time to practice it, or do so well. It's not clear how to match your RD practice to your resources - what compromises might be available to those needing to strike a balance.

Complex guidelines that make RD feel difficult may make people less likely to do anything: "If hurdles are super high no-one will do it…. [it's] hard to decide what's good enough." For this reason, many asked for 'good enough' guides to RD which are simple, practical and would 'get people to 80%'. This is particularly the case in remote working locations - where, ironically, practitioners are most likely to come into contact with the most sensitive data. This disparity is a point of tension between policy experts and practitioners, the former being more inclined to make sure the correct processes are being followed fully, every time.

Respondents did feel that donors could make a difference here, through 'a commitment... to support a transformative effort around the infrastructure of NGOs.' Donor respondents pointed

to efforts by some to add digital security issues to pre-grant inquiry and capacity-building projects for core grantees.

Recommendations

The focus of this consultation was, at least in part, to understand how RD tools and resources can best contribute to improved practice in this field. Many recommendations and comments on this came out of the consultation, and they are summarised below.

However, two higher-level issues seem to us to be more critical and a longer-term project to address. Without these questions being addressed, even the best possible suite of tools in many languages will not result in sustainable, universal improved RD practice by default.

1. Resource non-profit investment in critical systems and people

The interviews reflect how critical a gap time and human resource are in maintaining the poor standard of RD knowledge and understanding. Investment in core NGO capacity and infrastructure is crucial. Specific funding for improvement here, as part of grant funding, would be a huge step forward.

This is particularly important in the field, where investment at head office and policy level is very difficult to see reflected, outside of pockets of good practice driven by individuals.

Implementers seeking to be influential in the space could consider making RD a prerequisite for all their partners and donors, which might encourage investment by leadership, although at present, given the reported state of compliance in all social change organizations we spoke to, this would need additional investment in guidance and core funding as part of such grants, and probably time for organizations to initiate change.

2. Research and open, safe spaces for dialogue

We need more, and more open discussion of RD in all fora, and more research and guidance on specific topics, outlined below.

a) How big is the problem?

For donors and philanthropies, simple anonymous surveys of their grantees could shed light on whether and how far this is an issue for their portfolio. Donors may themselves need to review their internal RD practice, and should consider how they provide support and require good practice from their grantees in this area. Some donor policies around RD are beginning to be implemented, and the fruit of this should also be examined.

Policy experts may need to better understand the true spectrum of knowledge and practice of RD, in order to be able to better target their advice and support.

b) Legal issues

Broadly speaking, the legal landscape needs to be mapped - at first generally and globally, looking at international schemes and common principles and sites of regulation within national and international law; how platform terms of service and hosting invite multi-jurisdictional issues; highlighting the need for legal advice prior to setting up new platforms and systems; and beginning to identify ways to get help and support.

Further research is necessary to understand the position of, and aim guidance and support at, other actors in the system - specifically, donors and platform providers. Tool providers should be more transparent with where and how they store data, how they are approaching liability and RD issues, and where they are laying responsibility for RD on their users.

c) Data-centric versus rights-based approaches to use of technology in social change projects

Our interviews pointed to a widening gap between proponents of RD and the enthusiasm and funding flows towards data-centric, utilitarian projects which prioritize access to and monetization of data over RD principles. These opposing views might arise even within the same institution, in some cases. The data-centric vision of tech for social change, though, appears well-represented among a group of donors with relatively large spending power, lighter emphasis on evidence-based granting, and swift decision-making - although SIMLab also sees it in the 'innovation' spaces around technology, particularly in humanitarian aid.

These powerful organizations can heavily influence practice beyond their grantees. One conversation painted a picture of a pendulum swinging too far towards data as a good in itself, with a more context-specific and grounded analysis of risk and benefit seen as negative, expensive and too conservative, in the face of the limitless potential of data-driven work. This translates into lack of funding and interest in implementation of RD, and perhaps more damagingly, lots of funding for data-centric approaches that do not respect the rights of data owners. In this analysis, there will be no way to promote or encourage a cautious, human-centric approach to data management until 'the pendulum swings the other way.'

This raises urgent and, to SIMLab, disturbing questions about power and the role of data and ethics in the tech for social good space. These rapidly polarizing views of and approaches are setting up inconsistencies between and across platforms and projects that would make it impossible for data subjects to navigate the marketplace of such projects and make informed decisions about who should get their data and how it would be managed. We need to explore and openly discuss the tension between data positivism and the responsible data movement.

3. Guides and tools

Practitioners need a better-signposted journey to individual and institutional compliance.

Language and compassion

At entry level, simplified definitions or principles could be centred around affected people and practitioners relationships with them, rather than the legalistic language of 'informed consent' and 'retention'. SIMLab spoke with one person who suggested that compliance should start with compassion; creating an understanding of why certain rules and policies are important. "We try to 'put our staff into beneficiaries' shoes' when we talk about about data responsibility."

Create entry level guidance

All guidance should include entry-level or '101' advice, building to links to more advanced resources and practice, recognizing that not all practitioners are at the same stage with RD, and differing capacities across large organizations or among partners. Stand-out phrases from our consultation included requests for:

> *'1-5 tips before you go to the field'; '5 things you can do in today to protect beneficiaries'; 'What are three things I need to do to get to 80%?'; 'I need a 'good enough' guide.'*

One respondent suggested:

> *'think about how most people understand security - locks, for example, as a spectrum from single lock, deadbolt, security system, wall, cameras - what is the equivalent hierarchy or breakdown so I can choose between the bare minimum that's legal, or the 'Rolls Royce' thing for people using sensitive data?'*

It's important to note the challenges that this level of simplicity poses for the policy people trying to draft them, who are loath to set a low barrier to entry when poor RD practice has the potential to cause both harm, and legal liability.

Make resources interactive and compelling

To create behavior change, resources will need to be interactive, discussion based, thought-provoking and relevant to both field staff and HQ staff. Talking about the principles of RD can be more powerful than coming up with a precise check-list of things to do, because principles allow for discussion and thought and integration into an organization's existing approach and world-view.

Some specific suggested tool formats included:
- Interactive workshop, featuring a combination of reading to do ahead of time, hands-on learning, and then more material and guidance available as implementation begins
- Ongoing coaching through tips and reminders, delivered via email or SMS

- Case studies from those who have made organizational level changes with practical ideas others can follow.
- Tools for different moments in the project cycle, or that reflect the need to continually reassess RD during the project
- Resources according to threat model, so that I can implement practical, discrete protections to counter the threat.
- Lists of 'responsible features' to look for in platforms or tools, and 'things to consider' when adopting new platforms - not a repository of 'responsible' platforms, as tools and platforms change over time
- Examples of good and bad practice would enable practitioners to recognize themselves in the guidance

Other suggestions included:
- Guidance should be positive in tone and avoid 'nagging' or intimidating users
- Benchmarking could reflect different levels of progress toward an agreed standard of RD behaviour similar to the Core Humanitarian [Standard](#) on quality and accountability in humanitarian aid. This would enable audit and clear steps toward improvement.
- Guidance should include the process of institutionalizing RD, with sample staffing structures and compliance structures
- Wherever possible organizations should be supported to recognize sites of integration for RD with existing systems, e.g. human resource management, research, and monitoring and evaluation, and to make RD part of how they do business.
- All resources should be translated into multiple languages and be provided in various formats including machine readable text, downloadable to be easily accessed offline, and with small image sizes for smaller download files

Respondents mentioned the following existing tools as useful:

- Oxfam Responsible Data Tool Kit
- The ELAN Data Starter Kit
- ICRC guidance on Standards in Protection Work
- Guidance from the UK Information Commissioner's office
- UNHCR data protection policy
- Good Practices Group
- Some donor guidance (USAID)

But it is clear from our consultation that, by themselves, guides and tools will only achieve so much. Other, more critical areas of work must proceed alongside this, or the responsibility for RD will always remain with overworked, under-resourced practitioners who simply lack the capacity to implement it.

Annex 1: Interview Guide

We will be asking you some questions about how you and your organization feel about responsible data, how you handle it and how you think we could all manage data more responsibly. I/we will be making notes as we go/recording our interview. The notes/recording will only be used for our research, and we will try our best to look after them so that no-one else can access them. We will not share them with our research partners or quote you directly without asking you first. We will destroy the notes/recording in two years. If you want us to destroy them sooner, or give you a copy, you can ask us.

Do you agree to those rules? Are you ok to continue the interview?

1. What does the term 'responsible data' mean to you?

   Our definition of Responsible data for the purposes of this project is:

   The duty to ensure people's rights to consent, privacy, security and ownership around the information processes of collection, analysis, storage, presentation and reuse of data, while respecting the values of transparency and openness. (Responsible Data Forum, working definition, September 2014)

2. How do you feel about it?

3. When do you think or talk about responsible data? [*what is your role?]*

4. [Practitioner-specific] About the data you gather

   a. What kind of data does your organization handle?
   b. Where is it stored (format, location) at different times?
   c. How and when is it transmitted?
   d. Which stakeholders, if any, are driving conversations about responsible data in your organization?
   e. What, if any, legal requirements bound your data use?
   f. What processes/procedures do you have in place throughout the data life cycle (creation, management, deletion)?
   g. How are these managed, enforced? Who or what ensures compliance?
   h. How does theory differ from practice in terms of how your organization handles responsible data?

5. What are the pain points of responsible data? What are you worried about?

   a. For you?
   b. For your organization?
   c. For the ecosystem?

6. [Practitioner-specific] What would happen (or what would you do) with your data if your organization closed down tomorrow?

7. What do you think drives good and bad practice? (nudges/design for good practice)

What has helped you do better responsible data? Has there been a turning point for you? What were key moments? (Examples: A challenge (such as a data breach) A well-designed resource (what about it made it great?) A speaker or conference?)

8. What tools do you or your organization use and recommend for RD?

    a. In what form are these resources?
    b. What's helpful about the resources? What's challenging?
    c. When do you consult these resources?

9. More generally (not just responsible data):

    a. which types of tools or resources influence you?
        What professional resources do you read every week?
        What do you read on planes/trips?
        What do you recommend to colleagues?
    b. When you want your colleagues to learn or understand something how do you do that? What's the most compelling content?
        i.   For people who work where you work
        ii. For people who work at head office
        iii. For people who work 'in the field'

Annex 2: Survey questions

SIMLab is working on a project funded by Stanford's Philanthropy and Civil Society Team to develop better resources for social change organizations who end up generating and managing other people's data in their work - things like personal details, or information about their health or habits. First, though, we want to know what people are doing now.

You are not required to take this survey, nor are you required to identify yourself through your email address. We will not use any of your responses without asking for your consent.

1. What is your job title?

    a. executive or leadership

    b. project management

    c. technical advisor (not technology)

    d. technologist or developer

    e. Other

2. How many people work for your organziation?

    a. 1-10

    b. 11-25

    c. 26-50

    d. 51-100

    e. 101-500

    f. 501+

3. In which of your organization's offices are you based? Select the option that most closely matches your situation.

    a. main or only location

    b. headquarters office

    c. regional office

    d. national office

    e. local branch

      f.   remote

      g.   Other

4. Over what medium do you prefer to learn new things relating to your work?

      a.   reading (short articles)

      b.   reading (in depth report or research)

      c.   reading (tools or guide)

      d.   listening (e.g. podcast)

      e.   watching (e.g. video)

      f.   looking at visuals (e.g. infographics, flowcharts)

      g.   Other

5. Does your organization collect, analyze or in any way manage personally identifiable information (PII)?

      a.   Yes

      b.   No

      c.   Not sure

      d.   What is PII?

6. Does your organization collect, view, store, transmit, analyze or generate any of the following data?

      a.   Financial

      b.   Health

      c.   Location

      d.   We don't ever see any data

      e.   Other

7. How do you store data?

      a.   web-based file syncing (e.g. Dropbox or Google Drive)

      b.   web-based secure data storage (e.g. SmartVault or TSD)

      c.   file on computer hard drive

      d.   handwritten paper records

      e.   external hard drive or USB stick

f. we don't ever store data

g. Unknown

h. Other

8. How is data most often transmitted (sent from one place to another)?

    a. Email

    b. external hard drive or USB stick

    c. cloud-based file sync

    d. organizational intranet (local private server)

    e. Secure site

    f. Local area network (LAN)

    g. Unknown

    h. Other

9. Is there a data policy at your organization? (e.g. processes/procedures throughout the data life cycle: creation, management, deletion).

    a. Yes

    b. No

    c. Not sure

10. Do you comply with your responsible data policy?

    a. yes, absolutely

    b. Yes, mostly

    c. Sometimes

    d. Not really

    e. Not at all

    f. We do not have a responsible data policy

11. Why?

12. Are you aware of any legal restrictions or regulations related to how you manage data? Please explain

13. What tools or resources do you use to educate you and assist you with data management ?

14. Who do you feel should be responsible for driving good data practice?

      a. My manager

      b. Executive leadership at my organization

      c. Donors

      d. Governments

      e. Me

      f. Other

15. What is something you could do to encourage responsible data management?

16. Please complete this sentence: I feel BLANK about responsible data management because BLANK

17. Is there anything else you think we should know about your experience with data management?

18. If you would like to stay up to date on our data work please leave your email address here

Thank you for participating in our survey. Please share the survey link with colleagues! If you have provided your email address we will send you early insights from our consultation and survey process. If you would like more information about the program or about SIMLab please visit www.simlab.org or follow us on Twitter @SIMLab If you have any questions about the survey, please feel free to reach out to kelly@simlab.org

Thank you!