The Ethics of Data in Civil Society
Conference Synthesis
September 2014

Lucy Bernholz and Rob Reich

On September 15 and 16, 2014, one hundred scholars, nonprofit executives, technologists, data scientists, social media and search company managers, policy makers, and philanthropic funders gathered at Stanford University for the Ethics of Data in Civil Society Conference. The conference was hosted by the Stanford Center on Philanthropy and Civil Society's Digital Civil Society Lab, with co-hosts Harvard Humanitarian Initiative and the Brown Institute of the Columbia School of Journalism and Stanford School of Engineering. A list of other conference partners is attached at the end of this document.

This document syntheses the purpose, process and outcomes of the conference. Additional materials from the conference can be found online at stanford.io/ethicsofdata.

Conference Purpose

Civil society is a critical part of democracy. It sits apart from business and government. It is Janus-faced, powered by the private and voluntary action of individuals while producing effects, intentionally or unintentionally, on the public sphere. Civil society organizations mediate between the private sphere of the individual and the public space of the government. It encompasses formal and informal organizations of a million different kinds. Civil society organizations can be enduring or transient, large or small, local, national, or global. What binds them all together is this: Civil Society is the place where we as individual actors (citizens) do things that benefit others. We call this "private action and private resources for public benefit."

A flourishing civil society depends (specifically in the US context but also internationally), on rules, laws, and norms that protect freedom of speech and association – so we can say what we want, affiliate with whom we choose, for activities of common interest, and act without coercion or surveillance.

Most of the rules to-date about digital data and digital infrastructure have been crafted by and for business and governments. But we have different expectations for civil society organizations – although these expectations are not always spelled out. We have, for example, different expectations of how a commercial provider (Comcast or AT & T), a government provider (such as a local utility company) and a nonprofit (such as a library) will handle our data when we contract with them to get online. For the commercial provider to user our usage patterns to offer services seems normal, whereas we probably expect that a public system would focus more on equitable distribution and access then on innovation. We also expect different

things from them with regard to the content we actually make use of once we are online – as the fight over net neutrality or our expectations for censorship-free libraries show.

The questions of data ownership, governance, and consent are particularly interesting to us in the context of our work on digital civil society. We're starting on the road to understanding the roles of digital data in shaping civil society and how we can express and protect those assumptions so that civil society continues to thrive. Doing so now is particularly important – as the technological tools will just keep changing and the rules have not begun to catch up.

Our hypothesis is that it's not enough for civil society actors to participate in debates about digital policies and practices – though that is very important.

The underlying question shaping the conference was whether the ethical questions raised by digital data and infrastructure are common across the many domains in which civil society is active?

The conference structure and substance was designed to explore whether common ethical concerns can be articulated across all of the domains of civil society.

If the answer to the question is yes, we think that civil society will need to actively create three kinds of codes that will protect a right to exist in digital environments and allow individuals and organizations to thrive with digital tools:

1. Software codes - That don't necessarily do what commercial code does
2. Social codes – Organizational practices that manifest integrity and trust
3. Legal codes – Rules that guide nonprofit and philanthropic use of digital data as well as address the issues of data donations

If the answer to the question is no, then the types of codes identified above will still be needed, but will have to be crafted within specific domains while being able to handle digital data that flows across domains and sectors.

The conference participants all brought additional and different questions. Some have experience based on specific types of digital data – from satellite imagery to financial transaction data, medical records to mobile phone tracking.

Some were trying to figure out cross-sector partnerships built around data – sharing the digital data from cities to nonprofits, or protecting data collected for human rights activities from being used in criminal proceedings.

Others were focused on data lifecycle issues – questions about holding on to data in perpetuity or deleting data from the public record.

The conference included a deliberately diverse group, not fully representative of civil society but surely reflecting different missions within health, education, disaster relief, journalism, humanitarian aid, civic technology, and more. Participants represented academia, policy makers, technologists, activists, creators, users, and funders of digital data and social good.

Conference Process

Allen Gunn of Aspiration facilitated the event in an open source conference format. The time together (1.5 days) was focused around workstream sessions on topic pre-identified by the planning committee (Planning Committee and Partners: Appendix A) and on additional topics generated by participants. These culminated in a final feedback session in which all small group work was presented to the conference of the whole and commitments to carry work forward were made. (Agenda: Appendix B) Brief descriptions of the workstreams are provided below.

Digital civil society ecosystem mapping

Two groups worked on questions about governance and accountability for digital data in civil society. Key questions for the groups were: Who governs what and what new forms of accountability are emerging?

In the group "Code Share: Digital Ethics Across Domains" Madeleine Taylor and Anne Whatley facilitated a discussion that produced a matrix for locating digital data use cases across sectors. They visualized the different providers and consumers of digital data in civil society, the relationships between them, and the points of negotiation that present themselves in those exchanges.

Ben Stokes facilitated a group that sought to create a public-facing credibility systems that community organizations would use to learn about, improve, and attest to their digital data practices. Modeled on Creative Commons, online educational badging systems, and verification systems the group produced prototypes of "Ethical Badgers" – icons that would simplify the communication about how digital data are collected, stored, used, shared, and destroyed.

Ethics of technologists, activists, organizations, communities

Two groups worked specifically on an issue that that arise from the working relationships between professionals and volunteers, a situation that distinguishes civil society organizations from business and government. The role of volunteers also came up in other groups and warrants more attention.

 Aman Ahuja led a group that asked specifically "Who is responsible for what and to whom? How do we know? How do they know?" The group discussed how data scientists can communicate the ethical choices embedded in data science to their NGO partners, and how community organization professionals can better express

the ethical challenges inherent in their work. The group crafted both a communications process and documentation protocol built around README files that would allow these conversations, options, and choices to be shared and built upon.

Another group, led by Jake Porway, focused on the ethical choices built into algorithms of all kinds. The first phase of the group's work resulted in articulating the nature of algorithmic functions and the ethical crossroads that are often hidden within the "black box" of software. The second phase of the group's work created a process for making algorithms more visible and transparent to lay users. The process, which they dubbed "AlgoCompass," would apply to algorithms developed within and for civil society work. The group also argued that increasing public literacy about data and algorithms was itself a role for civil society and that the "AlgoCompass" might become something civil society applies to commercial and government programs on behalf of individual and community rights. .

Ethical choices along the data lifecycle

Heather Leson and Patrick Vinck facilitated groups approaching the ethical challenges through the lens of the data lifecycle. Their goal was to identify what types of ethical challenges arise when and how can community groups, especially those that rely on volunteers, address them in time? This group, "Data Cycling: From Choices to Consent," produced a checklist for nonprofits to use in planning and implementing a data-driven process. The checklist is intended to highlight ethical choices, signal upcoming decision points, and allow organizations to prepare for unexpected opportunities and consequences.

Principles and Practices

Building from the pre-conference provocation paper, these groups attempted to frame ethical codes of practice for civil society. They began by wrestling with what such codes of practice might accomplish, for whom, and how? One group took on the code writing challenge directly. Led by Susan Liataud and Andrew Woods, the "Crafting a code of data ethics for civil society" group presented two draft codes of ethics for feedback at the final session of the conference.

Approaching the same idea of a code of conduct from the perspective of avoiding harm, Nathaniel Raymond facilitated a workgroup that built a "responsibility for harm" management checklist. Similar to the data cycling checklist, the model is intended to equip front-line aid providers and their organizations to imagine potential harm, prepare for it, mitigate it, and recover from it (when needed). The example draws specifically from humanitarian aid scenarios, next generation prototypes could be built to focus on other domains.

Researchable futures

Professor Rob Reich, co-leader of the Digital Civil Society Lab, facilitated a discussion looking specifically at the role that scholarship might play regarding digital data and civil society. The group, "Researchable futures," crafted a research agenda for scholars regarding ethics, digital data, and civil society. What questions should be asked? What disciplinary or methodological approaches are important? At the end of the planning the group presented a multi-disciplinary research agenda and a book proposal, as well as several topics for scholar/practitioner/policymaker working groups.

Ethics across the tech stack

Benetech, a nonprofit technology firm serving advocacy and human rights workers, is modeling the creation of purpose-built technology for civil society. Their Martus project involves data collection, storage, use, and destruction capacities designed specifically to meet the needs of human rights advocates. Lessons from their work were extrapolated into other discussions and informed the development of several checklists for responsible data use.

Urgent Issues Forum

In addition to the ways in which civil society organizations need to think about their use of digital data and infrastructure, the conference also address the ways in which these tools, particularly data analytics, predictive data, and remote sensors are being used by policy makers and in fiscal decisions to shape the domains in which civil society works. Generated from the participants, the "Urgent Issues forum," drafted a 24-month agenda for civil society organizations and funders of the greatest digital data threats to freedom of association, expression, and individual privacy.

Discussion with Kate Crawford

Rob Reich and Lucy Bernholz hosted a breakout conversation with Kate Crawford of NYU, MIT and Microsoft Research. Her research on the discriminatory nature and uses of digital data and the ways digital tools are reshaping journalism cuts across many of the workstream topics addressed during the conference. The discussion touched on the catalytic role that Edward Snowden has had on general awareness of government and corporate data practices. She also touched on the need to continue building not only awareness, but a much more nuanced literacy among citizens about "how data work," – what they are, how they are created, how they can exacerbate bias, and how to counter balance their growing role in policymaking. Her expertise on the cultural implications of data technologies allowed insight into the practical value of producing ethical codes of practice (Crawford contributed to the Bellagio Framework on [Big Data, Communities and Ethical Resilience](#)).

In addition to the workstreams noted above, conference participants also shared information about and advanced their own work in a number of areas. These discussions and scholar/practitioner partnerships included:
- Ways to advance data literacy at the community level;
- The implications of digital data on social justice, equality, and marginalized communities;
- Opportunities to engage marginalized communities in shaping practice and policy;
- Practical ways to improve working partnerships between data scientists and community groups, including discussions of ethics, liability and power;
- Demonstrations of functional prototypes for getting informed consent for sensor-based individual health information (e.g. disease research drawing from wearable technologies);
- Practical, economic and legal issues related to data philanthropy;
- Data standards as a governing tactic for nonprofit partnerships; and
- Improving and systematizing data partnerships between industry and academia/researchers.

Conference Outputs

Raw material from the workstreams was presented for feedback to the entire conference at the conclusion of the event. Notes and photographs of the prototypes are still being compiled and will be made available as they are finalized. A working list of outputs includes:

- Class curriculum on trust in conflict zones to be taught at Stanford Fall 2014
- A "responsibility for harm" checklist
- A framing document on different types of consent (carried forward by Responsible Data Forum and others)
- A prototype for analyzing the ethics of algorithms ("AlgoCompass")
- A 24 month "urgent issues" agenda
- A research agenda (Partially going forward by Digital Civil Society Lab)
- At least one book proposal (Possibly going forward by Digital Civil Society Lab)
- Opportunities to craft common practices for data donations between commercial enterprises and research institutes (Possibly going forward by Digital Civil Society Lab)
- Mock-up nonprofit Terms of Service agreements to align with their missions (carried forward by Responsible Data Forum, Digital Civil Society Lab, and others)
- Two draft codes of ethics for data in civil society
- A set of tools for making ethical decisions across the data lifecycle (carried forward by Responsible Data Forum and others)
- A data "badger" for ethical management of data in civil society
- A matrix for locating use cases within and across sectors
- A process for data scientists and nonprofits to articulate and document the ethical choices they made in building apps, making visualizations or analyzing data sets

Blog Posts by Participants

Heather Leson
textontechs.com/2014/09/data-cycling-from-choices-to-consent/
textontechs.com/2014/09/infusing-ethics-into-data-projects/

Christopher Wilson
https://www.theengineroom.org/modeling-consent-policies-for-civil-society-data/
https://www.theengineroom.org/the-big-picture-key-milestones-and-questions-for-ethical-data-projects/

Dave Sessions and VitalWave
http://vitalwave.blogspot.com/2014/10/the-tension-between-those-who-have-data.html

DataPop Alliance produced this summary of the conference. It is available here:
http://www.datapopalliance.org/blog/2014/10/28/ethics-of-data-in-civil-society-conferencehighlights-and-perspectives

Conclusion and Next Steps

Civil society is distinguished from business and government by the use of private resources to public benefit, the role of voluntary labor, the fluidity of participant action, and an unstated, but valued, expectation of trust and integrity. These characteristics were present in each of the distinct working sessions. How these elements shape ethical practices for using digital data may hold the answer to the conference's organizing question – is there a common set of values to be articulated?

From the perspective of civil society – individuals, associations, nonprofits, philanthropic and social investment funders – the conference confirmed the importance of consent, ownership, governance, privacy, freedom of choice and free assembly as focal areas for considering the implications of digital data and infrastructure. These are areas requiring research, practice development, and policy analysis.

All of the materials and ideas produced at the conference were made available to participants to carry forward in their work. As noted on the output section, we are aware of some of these moving ahead by partners and participants. The Digital Civil Society Lab is moving ahead with some of the ideas generated by the conference, as they align with our current capacity. The research agenda and book proposal, in particular, will allow us to expand our network for scholarly research. Other ideas from the conference fit into our goals for applied research and policy development. We will continue to monitor the uptake of conference ideas by partners, participants and the public.

Appendix A: Planning Committee and Partners

Planning Committee
Isaac Baker, Harvard Humanitarian Initiative
Lucy Bernholz, Stanford Digital Civil Society Lab
Rahul Bhargava, MIT Civic Media Lab
Jean-Francoise Blanchette, UCLA
Brittany Card, Harvard Humanitarian Initiative
Alison Cole, Open Society Justice Initiative
Leetha Filderman, PopTech
Allen Gunn, Aspiration Technology
Mark Hansen, Columbia University School of Journalism, Brown Institute
Sandra Soo-Jin Lee, Stanford School of Medicine
Heather Leson, Humanitarian OpenStreetMap
Emmanuel Letouzé, Harvard Humanitarian Initiative
Susan Liataud, Stanford Center on Philanthropy and Civil Society
Annie O'Connor, Vital Wave Consulting
Jake Porway, DataKind
Nathaniel Raymond, Harvard Humanitarian Initiative
Rob Reich, Stanford Center on Philanthropy and Civil Society
Benjamin Stokes, USC
Madeleine Taylor, Network Impact
Patrick Vinck, Harvard School of Public Health
John Wilbanks, Sage BioNetworks
Christopher Wilson, The Engine Room
Andrew Woods, Stanford PACS

Institutional Partners
Harvard Humanitarian Institute
Brown Institute, Columbia University School of Journalism and Stanford School of
Engineering
Aspiration!
DataKind
Pop!Tech

Funders
Fidelity Charitable
Google
Humanity United
Robert Wood Johnson Foundation
Omidyar Network
Open Society Foundations
Vodafone Americas Foundation

Appendix B: Agenda

Stanford PACS Digital Civil Society Lab
Harvard Humanitarian Initiative
and Columbia School of Journalism/Stanford School of Engineering Brown Institute

welcome you to a conference on the

Ethics of Data in Civil Society
SEPTEMBER 15 – 16, 2014
STANFORD UNIVERSITY ALUMNI CENTER
326 Galvez Street
Stanford University

September 15, 2014
11:30 – 12:15 **Arrival, registration, and lunch**

12:30 – 12:45   **Welcome and introductions**
We'll bring the meeting to order with words of welcome and short introductions. Review of agenda and logistics before we move into program discussions.
Lucy Bernholz and Rob Reich, Kim Meredith, Allen Gunn

12:45 – 1:00   **Digital Civil Society** Why ethics, data, and civil society? Why us in this room and why now? (Bernholz and Reich)

1:00 - 2:30   **Interactive Plenary**
Collaborative opportunity for participants to share and compare views on the broad range of issues concerning the ethics of data in civil society. As part of the process, the group will generate a map of open questions, issues and opportunities relevant to the agenda, which will inform subsequent sessions.

2:45 – 4:15   **Workstreams Kickoff**
Participants will opt-in to small working sessions. Each session will address a specific topic associated with the workstreams identified below. Each group will focus on scoping and drafting a concrete deliverable – these will be fodder for future sessions, the science fair on Tuesday, and work beyond the conference. Additional sessions beyond those listed will be proposed during the plenary.

September 15, 2014 (continued)

WORKSTREAM OPTIONS

Additional sessions will be developed during the opening plenary.

- **Digital civil society ecosystem mapping** Who governs what and what new forms of accountability are emerging?
  - "Code Share: Digital Ethics Across Domains" *Madeleine Taylor, Anne Whatley*
  - *Ben Stokes*
- **Ethics of technologists, activists, organizations, communities** Who is responsible for what and to whom? How do we know? How do they know?
  - "The Ethics of Algorithms" *Jake Porway, Aman Ahuja*
- **Ethical choices along the data lifecycle** What challenges arise when and how can we handle them?
  - "Data Cycling: From Choices to Consent" *Heather Leson, Patrick Vinck*
- **Principles and Practice**s: What might ethical codes of practice accomplish, for whom, and how?
  - "Crafting a code of data ethics for civil society" *Susan Liataud & Andrew Woods*
  - "Doing Harm: What it looks like and how it happens" *Nathaniel Raymond*
- **What don't we know?** How might we craft a research agenda for scholars regarding ethics, digital data, and civil society? What questions should be asked? What disciplinary or methodological approaches are important?
  - "Researchable futures" *Rob Reich*

4:15 – 4:30   **Break**

4:30 - 5:00   **Kate Crawford, "What's at stake"** in conversation with Lucy Bernholz

5:00 – 6:15   **Sparktails - Lively conversations with beverages**
Participants will choose from a slate of "spark talk" discussions and get to rotate through the room, joining at least three conversations.

Table   *Digital data and social justice (Seeta Peña Gangadharan)*
Table   *Informed consent (John Wilbanks)*
Table   *Where ethics meet agency (Laura Hudson, Sean McDonald)*
Table   *Responsible data in advocacy (Chris Wilson)*
Table   *Ethical Tech Stack (Jim Fruchterman)*
Table   *Non-Profit Geek Speak (Deb Levine)*
Table   *Data-Pop Alliance: promoting a people-centered Big Data revolution (Emmanuel Letouzé)*
Table   *"Eh – What's the Worst that Could Happen?" (Jason Payne)*
Table   *Standards as governance…or not (Jack Madans)*
Others to be announced

September 16, 2014

8:00 – 9:00   **Arrive, breakfast and coffee**

9:00 - 9:15   **Opening plenary**
                   Review progress and outcomes from Day 1 before charting the course
                   through the remainder of the agenda.

9:15–10:45   **Workstreams continue**
                   Building on outcomes from Day 1 sessions, groups will work to complete
                   outputs to be shared in the afternoon "Science Fair."

10:45 – 11:00   **Break**

11:00 – 12:30   **Elective working sessions**
                   Sessions in this time slot will complement the core workstreams and
                   allow those present to go further on a range of relevant topics generated
                   during Day 1 discovery as well as work to this point. Participants will be
                   welcome to propose additional session topics.

12:30 – 1:30   **Lunch (Birds of a Feather) – Ford Gardens**
                   An outdoor lunch will be served, with themed tables where participants
                   can opt in to a hosted conversation.

1:30 – 3:00   **Science Fair - outcomes from workstreams and other demos**
                   Session facilitators will be invited to present and share the outputs
                   generated in the various workstream activities. In a format similar to a
                   high school science fair, participants will move from station to station,
                   reviewing work done by each group and contributing ideas, questions,
                   and other insights.

3:15 – 3:45   **Where from here: Next Steps**
                   Participants will brainstorm and inventory action items and opportunities
                   arising from the ideas and assets generated during the two days,
                   mapping out paths and plans for ongoing collaboration and follow-up.

3:45 – 4:15   **Closing plenary**
                   We'll end the day with appreciations and reflections from participants
                   and final words from the organizers before adjourning.

Digital data can be beautiful…and creepy. Check out the art presentations in the lobby.
Heather Dewey-Hagborg: *Stranger Visions* (the art and science of discovered DNA)
Woody Powell and collaborators: "Mapping Evaluation Discourse."