Data&Society

Meeting Minutes: August 31, 2015

Attendees

Council Members:

Mark Andrejevic, Pomona College
Solon Barocas, Princeton University
Geoffrey C. Bowker, University of California, Irvine (co-PI)
danah boyd, Data & Society / Microsoft Research (co-PI)
Kate Crawford, Microsoft Research / New York University (co-PI)
Alyssa Goodman, Harvard University
Rachelle Hollander, National Academy of Engineering
Barbara Koenig, University of California, San Francisco
Arvind Narayanan, Princeton University
Helen Nissenbaum, New York University (co-PI)
Paul Ohm, University of Colorado Law School
Frank Pasquale, University of Maryland
Seeta Peña Gangadharan, New America / Data & Society
Sharon Traweek, University of California, Los Angeles
Matt Zook, University of Kentucky

Staff:

Emily F. Keller, Project Coordinator
Jacob Metcalf, Postdoc

Council members unable to attend: Alessandro Acquisti, Edward Felten, Eric Meslin, Alondra Nelson, Latanya Sweeney

Agenda

1. National Science Foundation (NSF) Big Data principal investigators (PI) meeting (logistics and substance) – led by Geof Bowker
    • Design forms of engagement and figure out boundaries of what we can contribute. Brainstorm about types of interventions.
2. Substantive conversation around case study – led by Kate Crawford
    • Based on sample case study by Arvind Narayanan and Bendert Zevenbergen, "No Encore for Encore? Ethical questions for web-based censorship measurement."
    • What is the best format/structure for a data ethics case study?
    • What type of review process is needed for quality control?
    • Does publication indicate review and approval by the Council? Who would like to join a review committee?
3. Name and structure for expanding the network – led by danah boyd
    • What is working or not working about this proposal?
    • Who are the most interesting people to include? Help us identify your best students and junior scholars for upcoming paid opportunities.
    • Name the network to represent the Council and also stand on its own; membership title should signal participation for participants' CVs.
    • Should network members have their names and photos listed on a page of the BDES website, similar to the Council?
    • How should we manage boundaries (loose versus structured)? What commitments should be required? How should nominations be reviewed?

1. NSF Big Data PIs meeting – led by Geof Bowker

RESEARCH ✛ INSTITUTE

Geof Bowker led a discussion about an NSF Big Data PI meeting scheduled for February (originally November). Jacob Metcalf and Geof plan to conduct interviews with attendees to help prepare for the discussion. Projects funded by the NSF's Directorate for Computer and Information Science and Engineering (CISE) have a focus on algorithms.

danah boyd mentioned the Values in Design Council at New York University, asking members to reflect on what worked well in this context. Members advised against "parachuting" into different fields and declaring which issues are important. They emphasized the need to frame issues in ways that make sense to computer scientists, and discussed having the PIs develop scenarios in collaboration with researchers by asking PIs to design scenarios that have worked well for them.

Jacob has done an initial analysis of 72 funded Big Data projects through April/ May 2014 and found many opportunities for introducing ethical discussions. The projects focus largely on security, public planning, and behavior and social media. Jacob will update his analysis before the meeting. Since not all PIs will attend, as many have finished their grants, knowing the attendees in advance is key for planning. The Council is most concerned with CISE programs.

The PIs can be divided into small groups to discuss puzzles and problems, do a deep dive on issues like security, and do mapping work. Provoking a conversation on a topic that the Council can help with is likely to increase buy-in and commitment, rather than presenting external knowledge. Tech folks on the Council should think about how best to engage with tech folks at the meeting. Geof offered to circulate a document on what issues are arising prior to the meeting, to garner feedback from the Council on how best to present and discuss pertinent issues.

2. Case Studies – led by Kate Crawford

Arvind Narayanan described the intended audience for the Encore case study as people with a basic understanding of technology, who may be doing research or data science commercially, and for whom the case study is directly relevant to questions they encounter in their technical work. He said the case study should be readable to someone who is not a technologist, though it will be most useful to someone who is. Council members said the framing of the piece should make it clear if the case study is primarily intended for educational use, and show the appeal to political science researchers early on. They suggested adding a glossary, or definitions for things like institutional review boards (IRBs), to make the case study more usable for people from other fields like law. Audiences in law, computer science, and social science could be asked to develop discussion questions that are framed for different courses, drawing on the people who will use the case studies to help build site modules, and then asking people to share the ways they used the case study in the classroom or lab. Different templates could be created for different audiences. Additional details could be added for non-technical readers, such as highlighting the sites that the script was trying to reach.

To go beyond the technical discussion, social scientists could offer insight on context as well as a qualitative assessment about the meaning of the type of censorship described in the Encore case study. Associated modules and domain-specific questions could be added. Authors could invite specific people from different fields to provide their opinions and generate dialog, such as through commentaries on the discussion questions. They can use interactive formats and non-technical versions to help people explore the issues. Greater context about censorship measurements could be beneficial, such as types of funding and actors. Adding commentary and ideas from the review board could also add perspectives from several disciplines. While the Encore case study is real, the consequences are hypothetical. Historical case studies can show consequences as they play out, and keeping an eye on which potential harms from the case study actually materialize could provide ongoing relevance in the future.

In weighing the tradeoffs of hypothetical versus historical case studies, the Council could draw on the NSF Big Data PI meeting to generate topics, though the PIs might not be able to discuss specific details. Hypothetical case studies can be written to provide a sense of the big picture of an issue, though they might be disconnected from practice, so the Council should only accept those that feel real or that are based on situations that the author has witnessed. Cases should be chosen to address issues that are timely but have a long shelf life. Case studies can have lasting value even if protocols

and approaches change after publication. Historical cases can be turned into hypothetical ones. However this must be done well; for example, Gilbane Gold, a hypothetical case about a whistle-blowing engineer, received criticism that a better understanding of engineering would have solved the problem presented. The facts must be engaging for readers with knowledge of the field. Cases should be robust and contain multiple issues, rather than being confined to a single issue. It is important to specify the issues addressed by each case so that people who are setting up courses like Responsible Conduct of Research (RCR) can use them to help students meet funder requirements.

The Council will be the strongest force for propagating the call for case studies and should view itself as the front line for getting writers. The open call can also surface up people who are unknown to the Council, including those not working under the label "ethics" who are thinking about these issues. Diverse outreach from Council members will be important.

For follow-up, Mark Andrejevic offered to try out the Council's case studies in the classroom. Solon Barocas and Matt Zook offered to join a review committee. Solon also offered to suggest students with outstanding papers as potential case study authors.

3. Network Expansion – led by danah boyd

As part of its commitment to build a network beyond its own membership, the Council is working to develop a large collection of researchers to help create the next generation of people working in these areas, to extend this work into various communities, and to create new opportunities such as grants. This network expansion will provide a feeder to the Research Coordination Network (RCN) led by Geof Bowker, and will live on beyond the Council. RCN participants will be among those invited to join this digital forum. Communication channels such as Slack or webinars may be used for engagement and seeking feedback on Council activities, in addition to email discussions.

In discussing prior experiences building networks, members provided two examples: a board that sought out junior researchers to provide insight on practice and go beyond policy, but didn't allow them to have a formal say or take their advice seriously enough; and a summer institute on economic geography that takes place every 3-4 years that has similarities to the RCN.

danah asked how the Council should set and maintain boundaries and what groups of topics should or shouldn't be included. Rachelle Hollander suggested that the network could have a topic on the development of big data and inequities for marginalized populations, such as Census type data that might reinforce inequities. Members said the network should be open to people outside academia but should focus on people with a connection to the subject, rather than the general public. The network should be kept as a research community and conversation but include industry people without watering down the focus. Practitioners who are dealing with daily tradeoffs related to ethics should be included, as bridge building between the sectors is core to the project. The commercial sector lacks the constraints of academia and advertisers are not subject to the same oversight structures, providing the potential for a rich discussion. There is also an overlap between industry and academia, including PhDs who go into industry. Network building is a good opportunity to reach out to people and have an infrastructure for diverse conversations.

The Council selected "BDES Network Member" for a membership title, and a length of one year, with potential for renewal.