Meeting #2 Notes

Council For Big Data, Ethics, and Society[1]

April 27, 2015
Data & Society/ 36 West 20th Street, 11th Floor

Schedule:

- 9:00-9:30: Breakfast / Catching Up
- 9:30-10:00: Status Update and Q&A
  - NSF Meeting
  - Website
  - Thoughts from Council
- 10:00-12:00: What's the Role of Ethics in Education? (moderator: danah)
  - 10:00-10:45: Group
  - 10:45-11:30: Breakouts
  - 11:30-12:00: Group
- 12:00-12:30: Lunch
- 12:30-2:30: What is Human Subjects Research in a Big Data Context? (moderator: Kate)
- 2:30-3:00: Break
- 3:00-4:00: Creating a Research Coordination Network & Future Steps (moderator: Geof)

Attendees:

Council Members:

- 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)
- Seeta Peña Gangadharan, New America / Data & Society
- Alyssa Goodman, Harvard University
- Rachelle Hollander, National Academy of Engineering
- Barbara Koenig, University of California, San Francisco
- Arvind Narayanan, Princeton University
- Alondra Nelson, Columbia University (*phone - morning session*)
- Helen Nissenbaum, New York University (co-PI)
- Frank Pasquale, University of Maryland
- Sharon Traweek, University of California, Los Angeles
- Matt Zook, University of Kentucky

Guests:
- Chaitan Baru, National Science Foundation

Staff:
- Jacob Metcalf, postdoc, University of California, Irvine
- Emily Keller, project management, Data & Society

Council members unable to attend: Alessandro Acquisti, Mark Andrejevic, Edward Felten, Eric Meslin, Paul Ohm, Latanya Sweeney

Guiding questions for the day:

Conceptual questions regarding ethics education:
- Data science programs are being started now. What does it mean to integrate ethics into the curriculum from the getgo? What should this look like?
- Existing technical disciplines need to integrate ethics thinking into their education stream. How should this be managed? What should be included?
- Practitioners could benefit from ethics education. How can this be managed in the wild?

Conceptual questions regarding human subjects:
- How do we understand human subjects in big data research? Is our current frame of human subjects too narrowly prescribed? Do we need to start asking different questions?
- Can we imagine a framework for redefining human subjects in light of big data research?

Brass Tacks
- We will be engaging PIs through a big PI meeting. Can we produce something concrete to educate PIs on ethical issues during our portion of the meeting?
- What exactly should go into different kinds of training programs?
- Can we imagine a way of integrating ethics education into datasets themselves rather than treating education as a separable module?

Introductions/ Catch Up:

Helen Nissenbaum initiated the meeting by identifying the need to foster and curate a diversity of engagements with the big data paradigm. She identified the Council as playing an important role in creating institutional foundations and institutional memory that can maintain the continuity of ideas necessary for long-term projects.

Chaitan Baru, Senior Advisor for Data Science in the CISE division at NSF, introduced a new initiative for Regional Big Data Innovation Hubs. These four hubs will be multi-sector, multi-year projects to connect people and will address both general issues (such as ethics) and regionally specific issues (such as agriculture or manufacturing). The hubs will support cyber-infrastructure, community outreach, and

educational programs. Although the hub proposals are still in progress, Columbia University will likely be the Northeast hub and the University of Michigan will be the Midwest hub. Baru stated it is essential to have a "spoke" on ethics. Such spoke projects would be organized under a particular hub (or potentially all hubs), but would directly apply to the NSF for funding. Some members raised questions about how ethics could or should fit in the hub-and-spoke metaphor. For example, if a spoke were dependent on a hub for funding and organization, then landing in the wrong hub would result in ethics projects being stymied—a familiar occurrence in science and engineering ethics projects. The spoke metaphor also suggests ethics and society are secondary, rather than priorities; it was suggested that it may be ultimately more fruitful to work across or between the hubs. Baru noted that all hubs will be focused on education and training, which could be an ideal route for fostering data ethics pedagogy.

Session 1: What's the Role of Ethics in Education? (moderated by danah boyd)

-What are the core challenges in introducing ethics into science and engineering curricula? Into practice?
-Should ethics be integrated into broader courses or be stand-alone?
-How to incorporate (both?) negative and positive ethics frames?

This session was based on the reading: "Pedagogical Approaches to Data Ethics," which examined courses and programs incorporating data ethics in computer science and other fields.

Core challenges to introducing ethics into science and engineering curricula:
1. **Getting students with equal levels of fluency** in both the technical and social sides of the topic.
2. If **research ethics is introduced as strict guidelines** that apply mostly to regulatory mechanisms in university research, people might go to industry to do innovative work by default; or start their own companies as a workaround to the IRB.
3. **Keeping ethics from turning into a compliance framework** and instead make it a necessary part of a successful design process.
4. **Incentivizing and organizing instructors** from multiple campus units. How do you find people on campus with the training to think through and incentivize instructors to teach a class, especially from other departments?
5. **Starting conversations** with people who may be resistant to recognizing the utility of ethics training. One way is to start with worst-case scenarios and then scale back, instead of starting with gray areas.
6. **Finding the best place in a course or curriculum to situate ethics**. Should ethics be at the beginning or end, or presented as a capstone? When is there enough technical expertise to solve ethical problems? Council members shared anecdotal evidence indicating that it is possible to be too early or too late in the curriculum for the lessons to stick. Ira Rubenstein's NYU year-long Privacy Seminar capstone course was very effective and two students' projects will be published.
7. **People have very different perspectives on forms of knowledge production**, which can effectively make ethics seem personal. Members suggested any data ethics curriculum attend to that tension and not just focus on meta or compliance issues, but also examine how epistemology and ethics become personal.

Core challenges to introducing ethics into practice:
1. **Outsourcing approach:** Ethics are sometimes viewed as an afterthought for a lawyer or specialist rather than as part of the technical decision-making process.
2. **Variety of educational backgrounds:** Educating students will not cover all practitioners, as people take different paths into their profession and may have missed essential trainings. Professional education for practitioners is needed. As an analogy, people working in communications don't necessarily study as journalists or take the relavant ethics courses despite journalism having a strong ethical identity.
3. **Convincing tech people that their work has a moral or ethical dimension**: It can be difficult to convince some groups that their work and decisions are ethical in nature. When practitioners use highly technical skills it is harder to see the moral dimension, therefore they may think that ethics come in with what people do with the research later. Furthermore, people with technical expertise may think that ethics are at play in design practices and are another technician's responsibility. Members noted a data/algorithm divide—many people doing pure algorithm work think ethics is elsewhere in the system, not in the purely mathematical algorithm.
4. **Tech corporation's approach to open data:** Corporate values often encourage the pursuit of data as long as they have access to it, without worrying about consequences until later. This indicates a paternalistic attitude in which their technical expertise is seen as granting them carte blanche as to set their own rules as custodians of the data.
5. **Reaching small companies**: Many tech companies have a small staff and they will be harder to reach.
6. **Difference in disciplines:** Disciplines often have very different administrative and pedagogical infrastructures. Also, it is a challenge to understand data from other disciplines, for example making data "public" doesn't mean "sharable" or "usable" because disciplines may have varying attitudes about re-use embargoes. Physics has historically favored keeping data fully public and re-usable, whereas other disciplines with economically valuable data may make data public but put embargoes on re-use.

Best practices for integrating ethics into science and engineering curriculums:
1. **Broad and diverse approach:** Make ethics central in programs, embedded throughout the research process, and present in as many courses as possible. Target new courses as well as incorporating ethics into existing programs.
2. **Interactive discussion:** Emphasize discussion and interaction. Classes should deliberate about the issues and appeal to people who like solving problems, but should not focus on simplistic problem solving. An example exercise would be to assign a hypothetical scenario and have students write a collaborative position paper. Navigating different opinions would help students to see that differing perspectives can also be right.
3. **Avoid oversimplification:** Steer students away from the approach that ethics is a detail to get out of the way. Teach students to consider ethics as part of their work, not something to outsource or have separate experts or a compliance lawyer deal with but to see their work as infused with ethical challenges. Teach substantive thinking.
4. **Move from training to education:** Training and certificate programs suggest right and wrong answers and students may break rules to find the answers. It is necessary to go beyond

compliance and avoidance of a university being sued. Regarding existing compliance requirements, try to make them more substantive.

5. **Multidisciplinary approach:** Wherever possible, students should be asked to work/think with the tools of multiple disciplines.
6. **Tie ethics to quality of work and positive notions:** Focus on fairness, transparency, and accountability rather than negative constraints. Ethics training can help students identify the arbitrariness in what they're normally doing—we can get them to think about the validity of their work through ethics, and help them see how ethics can make good business sense.
7. **Teach case studies and controversies:** Choose data sets that inherently have challenges. This is more effective than abstraction. It also helps them see ethics as more than a problem to be avoided.
8. **Design approach/choices:** Values in design can prevent ethics from turning into a compliance framework. Give advice to deans about how to integrate data ethics in ways attentive to patterns of professional development.
9. **Problem solving/social good approach:** Appeal to a sense of mission and doing good. People might be more receptive to the idea that you can do better for the world versus telling them they are doing something wrong.

Breakouts

Create a plan for what you think should be included and how you think an intervention should be staged to make it possible. Go for both conceptual and brass tacks.
-What advice should we be giving to NSF as they fund in this area?
-What advice should we be giving to Deans as they commit to this area?

-What are the tangible objects that you think can/should be produced to support this? (Curricula, data sets, etc.)
Provide an action plan and discuss.

Ideas:
1. Ethics education activities possibly appealing to companies:
   a. Monthly luncheon discussion. Identify a few companies and help design seed questions.
   b. Shock and awe talks, or just awe talks (the positive side). Examples: Lawrence Lessig, director of the Edmond J. Safra Center for Ethics at Harvard University, and Jonathan Zittrain, co-founder of the Berkman Center for Internet & Society.
   c. CSR activities, including practitioner side, to imagine different possibilities and promote cross-pollination.
2. Discussion forums:
   a. Publishing case studies/articles for discussion (with questions). BDES could put out a call for examples and then write up a few, which we could put on a wiki/ into a manageable list (something people can add to). Alyssa noted the need for a catchy, Twitter-friendly name. Suggested examples: Facebook emotional contagion study, Space shuttle Challenger, students friending teachers on Facebook. Have people debate something outside their purview.

b. Suggested venues for publishing discussion forums included:
    i. Rachelle, who is an editor of *Communications of the ACM*, suggested writing a column describing an exercise and people who are passionate about the topic will respond. This would reach beyond big companies. The ethics column is well read and short. This could include cases of data abuse or access.
    ii. Arvind suggested Princeton Center for Information Technology Policy's [Freedom to Tinker](#) blog, which is widely seen by technologists.
    iii. Trade journals and other newsletters seen by technologists and regulators.
    iv. The D&S website could have a public conversation or mediate an official discussion question of the month via public channels like Twitter. Getting people to comment could be a challenge so particular people could be invited to comment.
        1. Alyssa said this is probably more interesting to people if it's not tailored to them, like Google engineers discussing Turnitin.
        2. Geof said there is a lot of attention to the algorithm side, what about information scientists? Topics could be data classification and values embedded in data, or critiques of metadata.

3. For universities and educational programs:
a. Socratic cafes can create great discussion. Rather than providing a complete picture of an ethics problem, try to inspire conversations with a 2-3 page intro with open questions.
b. A "teach the teachers" approach would involve working with deans, identifying where their faculty can come together to work out ideas about teaching, and providing them with frameworks and mechanisms. Could also host workshops to bring faculty together around data ethics issues.
c. Can try to boost data ethics in general education, which is where many engineers encounter humanistic material. Actively engage in philosophy department and co-construct data ethics within schools.
d. D&S or BDES could aggregate issues such that if you were teaching a course, here are some of the topical areas that you could cover in the syllabus. Classify the cases widely relevant to the field (e.g., open source, open data, using open data responsibly), identify experts in specific areas, and chunk it all together.
e. Finding instructors to teach data ethics may pose a challenge but providing ethics modules that professors can drop into their courses would help spread data ethics into a variety of courses. For example, Shannon Vallor, Associate Professor of Philosophy at Santa Clara University, created a module for technology ethics that anybody teaching data-related topics, primarily technologists, who had not included ethics were able to utilize. The response was very positive and many people added the module to their courses. One class worth of material was the bump that many instructors needed.
f. Case studies by Data & Society should have a template. Come up with a description of best practices for rich case studies and collaboratively produce 5-10 in this model. There are a number of venues where such case studies can be shared, including online databases sponsored by the funding agencies. Rachelle noted that the Markkula Center is revising the Online Ethics Center, which could be a way to collaborate.
g. Suggestions about developing successful case studies for data ethics:

   i. Provide case studies and scenarios (including social justice issues) and/or materials to train teachers. They should also have surrounding material and information so people can make use of it.
   ii. Tailored or basic modules they can drop in.
   iii. Controversial issues to watch for: marginalization and stigmatization, and credit for contributions.
4. Professional societies
  a. Work with professional societies to integrate ethics into annual conferences.
  b. As organizations revise codes of conduct, promote the inclusion of ethics.
5. Ethics education and funding agencies
  a. Look into the history of science and engineering ethics education efforts from the 70's and 80's. There was a lot of effort, interest, and implementation, but many of the initiatives did not stick. It would be helpful for us to know why they fell apart, if they did, before we put a lot of effort into something similar.
  b. Linda Layne, NSF's Program Officer for ethics, noted in a recent article[2] that 90 projects were funded in the EESE program. Contact her for that list before she leaves NSF in August.
  c. CCE-STEM (the successor program of the EESE) could be a source for funding.

What is Human Subjects Research in a Big Data Context? (moderated by Kate Crawford)

-How do we understand human subjects in big data research? Is our current frame of human subjects too narrowly prescribed? Do we need to start asking different questions?
-Can we imagine a framework for redefining human subjects in light of big data research?

This session was based on the reading: "Human-Subjects Protections and Big Data: Open Questions and Changing Landscapes," which drew from the ethical regulations of social science and the humanities to examine human subjects protections in big data research.

What qualifies? Determining whether a data set qualifies as human subjects research is complex, as this may be altered by the added value and knowledge of data aggregation, analytics, and use. Data can become human subjects research beyond its regulation by the IRB. One area that the Council can contribute to is pointing out different ways that this happens. In industry, tracking people's metadata is not considered human subjects research and connecting human subjects to these debates will be controversial. Regulating the ethical use of human subjects data is difficult since technology and other factors are changing so fast.

De-identification and re-identification of data:
Cynthia Dwork has shown the great difficulty of de-identification and Latanya Sweeney has done work on the ease of re-identification. Some data sets may be safely de-identified until they are combined with others that make re-identification possible, much like a chemical reaction in which a final ingredient,

---

[2] Layne, Linda. 2015. "Ethics and the National Science Foundation." *Encyclopedia of Public Administration and Public Policy, Third Edition.* DOI: 10.1081/E-EPAP3-120053328.

added by an unsuspecting researcher, causes an accidental explosion. One example of re-identification is that people who use online dating sites often don't know that other users can paste their photos into google to identify them.

Federal regulatory and access issues:
Pending regulatory changes to the Federal Policy for the Protection of Human Subjects, known as the Common Rule, will impact permissible big data research techniques for projects and institutions that receive federal funding. The changes are currently under intergovernmental review. One of the primary purposes of the revision was to account for human biological materials that could be de-identified to not be human subjects research and how to link this to initial consent, but social and behavioral data rules are affected as well. Council members said there needs to be greater recognition of the value of data analytics in these new regulations. There has been some critique that the changes don't go far enough. National Academies submitted a report on the regulation of social, behavioral, and economic research for input into the changes. The report will be incorporated into the Common Rule but it will take years for this to change and the Council can work to influence it.

Oversight standards can be ambiguous; multiple studies have languished in limbo because no one knows what the oversight should be. Committees that do peer reviews sometimes assist with oversight. There are ambiguities around what types of research are exempt in practice, and researchers would benefit from greater clarity and fairness of definitions. It would be helpful to know what institutions are listed on the Common Rule exemptions list, as well as who runs it and what else is known about it. Consequences for violations are also a bit obscure and one Council member suggested that most researchers that are accused of violating IRBs are later found to be innocent by a Committee on Professional Conduct. There is sometimes a conflict between academic freedom and the IRBs. The risk of over-regulation in academia is that researchers will move to industry, which is limited only by legal issues.

danah boyd described a new advisory board established by the Department of Commerce to address access issues to data from sources such as the Census, the National Oceanic and Atmospheric Administration (NOAA), patent materials, and Gross Domestic Product analyses. The department is looking for ways to responsibly and effectively become the agency of data but the volume of data available creates technical challenges (for example, NOAA produces 2 terabytes of data per day).

Engaging with companies on data ethics:
Companies are significant actors in the area of big data derived from human subjects, who operate outside the purview of IRBs, as their research is neither federally funded nor under the umbrella of academia. Work to influence ethical behavior by companies is dependent on whether they are consumer facing and driven by their reputation or not. Companies with significant assets that are not in the public eye may be restricted by legal issues only, and some of their data activities that have questionable ethics may be fully legal and made possible by available technologies. In the case of public facing companies like Facebook, external pressures increase the chances of engaging in a conversation about ethics. How should the Council approach non-consumer facing companies that feel less external pressure? One Council member noted that the adversarial approach unfortunately works better in his experience, and that non-consumer facing companies also present research challenges in that they are hard to study.

Potential harms and known cases:
What is harm to human subjects in big data research? What harms are the Council trying to mitigate against? Potential harm is easy to prove but actual harm is often debated.

Commonly cited examples of protocol violations include a 2004-2009 study of oxygen levels for extremely premature babies, funded by the National Institutes of Health, which did not properly disclose medical risks to parents, according to the federal Office for Human Research Protections. However, the researchers argued that they were searching for best practices from within the normal but untested range that varied across hospitals, rather than intervening in patient care. Another example is the Lenovo Superfish scandal, a security breach in an adware program that exposed vulnerabilities to customer protections. In this case, researchers made use of flaws in computer systems to generate research, and a consequent study noted the ethical issues involved.

Building data prediction models on sensitive issues can be a danger in itself, and the issues are more complex than just rating research as legitimate or illegitimate.

The complexity of these issues makes it hard to know who should be considered the bad guys. Measuring actual harm requires some sense of who would like to take advantage of the situation by putting data sets together to cause harm. Many important security breaches come from within organizations - would social or technological controls be appropriate for this? Should intent matter when measuring risk and harm?

Crawford mentioned that the Federal Trade Commission started a 10-week Technology and Data Governance Research Fellowship Program to involve students in exploratory projects that help the commission learn about current technological issues relevant to its mission to prevent deceptive and anticompetitive business practices and enhance consumer choice. She said this program will help to examine potential harms and seek to show them empirically.

Information blocking: On the flipside of privacy concerns is the ethical challenge of data restrictions and withholding information that has potential benefits. The Council noted several examples in the medical and healthcare fields. In one case, data about length of patient hospital visits was collected but not released because the hospitals viewed such data as a source of economic advantage for competitors. In another case, pharmaceutical companies were not registering unsuccessful trials despite requirements to do so. These cases include instances when releasing data is legal and HMOs or companies are withholding it while citing patient privacy protections. This information could be used in collaborations between institutions and researchers to benefit patients.

Intent: What ethical principles, such as intent, should determine the data use once it is released? When we label datasets as risky or not, we remove the context of how they came into existence and how they are being used. For example, once passwords and usernames have been released by a hacker, should security researchers be able to touch them, for example, to conduct research to improve password security? Could public shaming or wake-up calls be considered a viable reason for sharing personal data? When security researcher Mark Burnett built and published a database of passwords and usernames from a variety of security breaches over the last ten years, his purpose was research to improve online security, and he claimed the data was already publicly available through search engines anyway. Once a data set

becomes open, even illicitly, what can stop people from using it? Similarly, what are the ethical parameters of using data that came from human suffering, such as medical data on hypothermia treatment derived from Nazi doctor experiments on concentration camp victims?

Data supply chain/ accounting for future uses:
IRBs are designed for advance decision-making, but many of the actions that affect whether data is being handled ethically occur later. How should data uses (sharing, reproducibility, combining with other data sets) be accounted for?

One potential solution is to keep track of the chain of data sharing, including all parties who have accessed the data, to trace it back to the source and report on uses down the line. Anne L. Washington developed this concept of "data supply chains" in collaboration with danah boyd at a January 2014 National Science Foundation workshop. Health data covered under HIPAA that is not de-identified is tracked this way. This model, if extended to other forms of data, could require all entities that acquire data with the intention of generating benefits to society to report back on what they have done and who else they have shared it with. Like the Creative Commons license, uses could be restricted down the line. Tracking where all the datasets are going would reveal a chain of custody to subcontractors and business associates and help enforce goals of participant consent. You can put pressure on the whole supply chain when you understand the different trajectories. How do you trace networks of data?

Collaboration:
One of the ideas being debated in the field is the possibility of universities pooling the proposal writing process, or having a national shared IRB database.

Community oversight and benefit: What role should communities have in research that is derived from and relevant to their lives? What are the mechanisms by which communities can see benefits from the research? In addition to university IRBs, research projects can also undergo review by community IRBs similar to a community board, and Community Benefit Agreements can be written, though doing so is likely to bring a high level of scrutiny. Community governance mechanisms can think through tradeoffs, risks, benefits, and adaptive governance over time in positive way to supplement IRB stasis. This concept can be viewed broadly, from social media communities to international relations.

Community governance is one way to combat concern about a potential "extractive" relationship between data collectors and human subjects that has come up at the World Economic Forum and in the city of Detroit. There is concern about power dynamics between U.S. researchers and human subjects in other countries and regions, such as the global south, in terms of who the beneficiaries of the research are. This issue has been compared to colonialism or natural resource extraction to enrich the U.S. Virtual communities can engage in governance through mechanisms such as Facebook voting, which has historically attracted a very small portion of users.

What should big data benefit to be considered social good? International development? Contributions to social systems? Value for the nation or social welfare? Who is the IRB representing? Are we thinking of the IRB and good research as representing humanity? Is that the wrong conception of what the IRB is

supposed to be doing? IRBs are intended to determine what is "good research," and are not supposed to take social welfare into account.

Individual consent: Should consent go beyond de-identification and anonymity and allow people to abstain completely from being counted as a number in a study that they don't agree with, regardless of whether their identity is disguised? One example is Do Not Track, a program that enables Internet users to opt out of being tracked by websites.

RCN proposal and other next steps (moderated by Geof Bowker)

Geof began with a discussion of the proposal he and others have been developing for a NSF Research Coordination Network (RCN) grant. The purpose of an RCN is, "to foster communication and promote new collaboration among scientists, engineers and educators with diverse expertise and who share a common interest in a new or developing area of science or engineering. By encouraging the formation of new groups and networks, the RCN program will advance fields and create novel directions and opportunities for research and science education." RCN's have an open application cycle, are typically 4-5 years long, and have a budget of $500,000.

The proposal as currently formulated would coordinate a network of data ethics researchers at multiple stages of their careers with the goal of propagating out the research and practices of the Council into a broader community. The proposal is modeled on the Values In Design project run by Geof and Helen, and a successful RCN about digital cultures. The primary activities proposed are:
- Early-career long workshops (annual in summer for 3 or 4 years): Graduate students and postdocs would assemble for a week of mentoring and collaboration, with the goal of producing fundable projects and publications. A select group of participants will be invited to repeat the workshops, in order to build expertise over time.
- Senior researcher agenda-setting meeting (annual for 4 years): Senior researchers will assemble to collaboratively set research agendas. Meeting themes will be coordinated by the host.
- Policy and media workshops at the NSF (twice): Core members of the Council will assemble at the NSF to develop networks with regulatory stakeholders and journalists covering data ethics issues.
- Conference workshops and networking (~20 events): Participants in the early career workshops will commit to coordinating at least 2 workshops or interactive panels at relevant professional conferences.
- International networking: Similar research networks are growing in Europe (located in Ireland) and Australia. This RCN project will facilitate international collaboration with these projects.

The RCN will be housed at the Data & Society Research Institute and supported by two part-time postdoc positions.

After touching on the RCN's details, conversation turned toward broad questions of how to foster research networks in general. Geof indicated that it is important to not just focus on ethics, but to also track the broader social issues of how we research big data. Frank noted that he has been attempting to organize a conference at Yale Law School focused on problem-driven research that is respectful of the

methods of each field. He stated that it would be productive to help justify inter-disciplinary work by building a community of scholars focused on broad problem areas, such as algorithms. Solon noted that he and Arvind have seen that computer scientists have been taking up ethical problems on their own terms and treating them as technical problems to design around (as opposed to mutually constitutive social and technical problems). He suggested such folks would benefit greatly from a sustained and public discussion about research and design practice. Kate emphasized the need to create a safe space in which nascent interdisciplinary projects can develop. The field needs spaces in which we can harness the building of energy and interest around data ethics. Geof discussed the need for an identifiable and familiar label for what we are doing. At the University of California-Irvine and elsewhere, data analytics programs are being built with no social/ethical/cultural wing. An easy to understand label, such as Values in Design, would provide a clear imprint and direction.

The meeting concluded with a next steps to-do list:
1. Rework the human subjects paper (Jacob & Kate)
2. Curated weekly data ethics newsletter/case study (Jacob)
3. Situate the human subjects work within the policy regime, see where D&S could intervene (Jacob)
4. Curation/consolidation point for best papers in X, Y and Z areas (Emily)
5. RCN: (Jacob + Geof)
   a. Collect locations/people for senior meetings and early-career workshops
6. Other kinds of future funded projects?
   a. CCE-STEM
7. Aspen type report for our meeting today? Need to think through curating (Jacob + Emily)
8. Curated topical sources (discrimination, privacy, human subjects, individual vs collective rights, international, data sharing, community level consent/IRB, design practices, data bias/equity, gray big data & what can't be retrieved)
9. 10 Simple Rules? (Matt Zook)
10. How do we find entry points to different audiences? How do we understand their puzzles/challenges?
    a. Invite folks from industry to fall meeting (danah)
    b. Strategy for engaging government agencies (including Commerce, FTC, etc.)
    c. Strategy for engaging with international researchers
    d. Strategy for engaging with community groups (ex: Patients Like Me, mTurk union) (Seeta)
11. Case study collection + art of case studies guide (Solon + Arvind + Jacob)