The Council for Big Data, Ethics, and Society

danah boyd
Data & Society
Research Institute;
Microsoft Research

Geoffrey Bowker
Evoke Laboratory
University of
California-Irvine

Kate Crawford
Microsoft
Research
MIT Center for
Civic Media

Helen
Nissenbaum
New York
University
Information Law
Institute

The Council was started in 2014 to provide critical social and cultural perspectives on big data initiatives. It brings together researchers from diverse disciplines—from anthropology, history, philosophy, physics, economics, computer science, sociology and law—to address issues such as security, privacy, equality, and access in order to help guard against the repetition of known mistakes and inadequate preparation. Through public commentary, events, white papers, and direct engagement with data analytics projects, the Council works to develop frameworks to help researchers, practitioners, and the public understand the social, ethical, legal, and policy issues that underpin the big data phenomenon.

Council reports have examined big data projects supported by the NSF; the history of ethical codes from mid-20th century; debates about personal genomics, the use of student data for personalized learning, and social media algorithms; the history of data management plans as a tool for integrating ethics principles; an examination of data ethics courses and modules in fields such as statistics, business, computer science, and journalism; the application of human subjects protections from the social sciences to big data analytics; and proposed changes to the Common Rule by the U.S. Department of Health and Human Services. Case studies and a white paper are forthcoming.

Policy Recommendations

- **Ensure the Common Rule and other relevant research ethics regulation clearly address regulation of data science**.
  Due to historical quirks in ethics regulation, data science occupies a space between 'research' and 'not research' that leads to substantial confusion and hampers effective and consistent ethics review. Future regulations should directly address this gap.

- **Seek ways to facilitate new approaches to ethics review inside academia and industry**.
  Big data research and industry has the potential to innovate improved models of ethics review, and policy-makers should facilitate this opportunity where possible.

- **Develop mechanisms of ethical assessment tailored to big data research methods and industry practices.**
  There is a notable lack of empirical research measuring potential harms of big data analytics to human subjects.

- **Integrate data ethics concerns into NSF program solicitations and the grant-making process.**
  The NSF grant-making process may be the most effective route to address gaps in ethics research and education.

Pedagogical Interventions

- **Create and distribute high quality data ethics case studies that address difficulties faced by data scientists and practitioners.**
  There is a notable lack of data science case studies available for public use, particularly in the National Online Ethics Center.

- **Develop and support data science curricula with integrative approaches to ethics education.**
  Science and engineering ethics education works best when integrated across a curriculum, rather than as stand-alone units.

- **Train librarians to achieve and promulgate data science and data ethics literacy.**
  University libraries are increasingly the campus hub of data repositories and instructors of data-sharing best practices.

- **Strengthen ethics-oriented activities within professional associations.**
  Computing and data science organizations can play a significant role in setting norms for research and practice. They should update their ethics codes to reflect the specific challenges of ubiquitous data analytics.

Develop cultures of ethics engagement in data science/industry and encourage cross-disciplinary networking

- **Engagement requires hybrid spaces.**
  Advancements in data ethics will require formal and informal spaces where people with wide ranges of expertise can network and collaborate.

- **Build models of internal and external ethics regulation bodies in industry.**
  Industry that utilizes big data analytics faces unprecedented challenges and requires input from both internal and external bodies.

- **Set standards for responsible cross-sector data sharing.**
  Sharing data between industry and academics is a particularly fraught endeavor, but can carry significant upside.

Areas for further research

- Should human data science be regarded as human subjects research?

- What are the quantifiable risks posed by correlative and/or predictive data research?

- Similarly, how should we account for the risk of sharing datasets when we cannot know what auxiliary datasets they will be combined (munged) with in the future? Does the risk differ with public datasets?

- How should data privacy and security scientists approach illicitly gained, publicly-available data?

- What are the options for self-regulation in data science?

- What resources are needed in the university context to encourage engagement with data ethics issues, particularly outside of the IRB?

- How can integrative approaches to data ethics be fostered in classroom environments? What pedagogical resources are needed?

- What are the ecological and environmental impacts of a rise in big data research and industry?

- How can ethical issues be integrated into core technical research?

- What motivates data scientists and their colleagues in industry to pursue ethics processes?

- What is the proper purview of "research ethics" as a topic in the age of big data?