Schedule

Time (Pacific)	Duration	Activity
1:15 - 1:30	15 mins	Framing
1:30 - 1:40	10 mins	Introducing the Working Group products
1:40 - 2:30	50 mins	Breakout feedback sessions
2:30 - 2:45	*15 mins*	*Break*
2:45 - 3:15	30 mins	Resume discussion

Important Things to Keep in Mind

- Data has *value* (monetary, property, cultural, intrinsic and personal)

- Constantly question who "owns" versus has "stewardship" over data

  - Researchers have been sued for falsely claiming ownership over data
  - Universities and companies have been sued by individuals and communities
  - Federal government has contractual obligations governing data

- Not everyone shares the same ethic of "open data" as researchers

- Judge "fairness" and "data equity" empathetically (beyond researchers)

- Data is not impersonal

- Data can be re-identified, especially when harmonized

- Data risk is not always centered on the individual; group harms may exist

A Case Example: Akimel O'odham ("Pima")

- **High prevalence of Type II Diabetes (>94%)**
- Lifestyle was more agricultural
- Traditional diet included corn, beans, squash, lean fish
- **"have probably been the most studied group for the causes and consequences of diabetes"** [Pearson 2015]

Indian diabetes study decried Tribe feared doomed; researchers accused of using, not helping, ill

November 7, 1999
Section: DTW
Page: B-01
The Arizona Republic

A Case Example: Akimel O'odham ("Pima")

Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas.

Galanter JM[1], Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, Via M, Hidalgo-Miranda A, Contreras AV, Figueroa LU, Raska P, Jimenez-Sanchez G, Zolezzi IS, Torres M, Ponte CR, Ruiz Y, Salas A, Nguyen E, Eng C, Borjas L, Zabala W, Barreto G, González FR, Ibarra A, Taboada P, Porras L, Moreno F, Bigham A, Gutierrez G, Brutsaert T, León-Velarde F, Moore LG, Vargas E, Cruz M, Escobedo J, Rodriguez-Santana J, Rodriguez-Cintrón W, Chapela R, Ford JG, Bustamante C, Seminara D, Shriver M, Ziv E, Burchard EG, Haile R, Parra E, Carracedo A; LACE Consortium.

Selecting SNPs informative for African, American Indian and European Ancestry: application to the Family Investigation of Nephropathy and Diabetes (FIND).

Williams RC[1], Elston RC[2], Kumar P[3], Knowler WC[3], Abboud HE[4], Adler S[5], Bowden DW[6], Divers J[6], Freedman BI[6], Igo RP Jr[2], Ipp E[5], Iyengar SK[2], Kimmel PL[7], Klag MJ[8], Kohn O[9], Langefeld CD[6], Leehey DJ[10], Nelson RG[3], Nicholas SB[11], Pahl MV[12], Parekh RS[13], Rotter JI[14], Schelling JR[15], Sedor JR[15], Shah VO[16], Smith MW[17], Taylor KD[14], Thameem F[4,18], Thornley-Brown D[19], Winkler CA[20], Guo X[14], Zager P[16], Hanson RL[3]; FIND Research Group.

Data access and usage mimic the biocolonialism and biopiracy concerns of over a decade ago

Guiding Principles for Data Governance and Stewardship

Wilkinson et al. *Scientific Data* 2016. **The FAIR Guiding Principles for scientific data management and stewardship**

The Global Indigenous Data Alliance. 2019. **CARE Principles for Indigenous Data Governance**

Ethics at the ADSA Leadership Summit

ADSA is the *Academic* Data Science Alliance

- What is **our** work to do?
- Who is **included**?
- How do we **walk the walk**?

From the 2018 Leadership Summit

- Ethics needs to be a component of every data science education program.

- A national community of practice and shared resources for ethics in data science education should be formed.

- Ethics should be considered through every step of the data life cycle, and practices and tools should be developed to support this work.

- Institutional Review Boards (IRBs) and insights about ethics from other disciplines represent an important starting point, but they alone are not equipped to handle new potential harms.

From the 2018 Leadership Summit (continued)

- Ethics for data science needs to be understood as requiring new foundational thinking about technology, society, and human beings, and a broad conception that engages issues of human values, broad participation, and justice.

- Opportunities to reflect on ethics should be provided at different stages of data science practitioners' careers.

- The national data science leadership community must actively recruit experts in ethics to define appropriate strategies.

Report available here: L.C. Erickson et al., Ethics Panel Report, 2018 Data Science Leadership Summit. https://docs.google.com/document/d/1tSkKSvJDrhmKzT6BspNr15-bmKLK2_SbuCUQVsrVLJM/edit
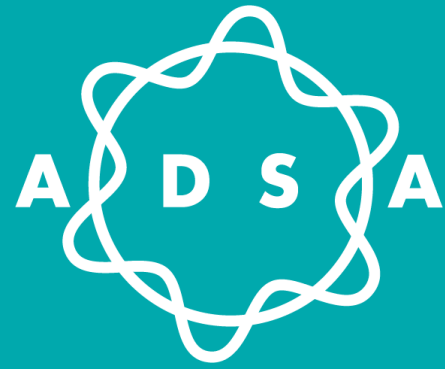
Working Group Perspectives

- Focus on practice

- Ethics is social

- Always contested

Working Group Products

Working Group Products

Draft documents for feedback

- Data Science Ethos Lifecycle

- Data Science Research Administration Document

Presenting the products: Margo Boenig-Liptsin, H. V. Jagadish

Data Science Ethos Lifecycle

A re-envisioned data science lifecycle
that supports practitioners
to consider the human contexts and ethics of data
at each stage of their work.

Data Science Ethos Lifecycle

Reflexivity

Power

Imaginaries

Sociotechnical System

Presented in an online interactive tool with real-world case studies.

Data Science Ethos Lifecycle

A collective work in progress, created by

Margo Boenig-Liptsin, Human Contexts and Ethics Program, Division of Computing, Data Science, and Society, UC Berkeley

Ari Edmundson, Human Contexts and Ethics Program, Division of Computing, Data Science, and Society, UC Berkeley

Anna-Maria Gueorguieva, UC Berkeley undergraduate

Carlos Ortiz, UC Berkeley undergraduate

Micaela Parker, Academic Data Science Alliance

Stephanie Shipp, Biocomplexity Institute and Initiative, University of Virginia

Anissa Tanweer, eScience Institute, University of Washington

Lifecycle mockup by **Emina Musanovic**

Data Science Research Administration

1. Typical procedures for encouraging human subject ethical research practices may be ill-suited to data science research.

2. Data science-driven research creates new types of risk beyond risk to individual human subjects.

3. Research data must be managed to provide adequate governance, sharing, and reproducibility.

4. Currently, administration of data science research is typically not encompassed by a single institutional oversight body.

Data Science Research Administration

Small working group developed a draft document

Phil Bourne, U. Virginia

H. V. Jagadish, U. Michigan
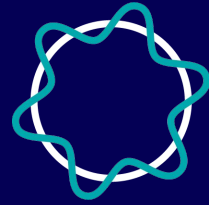
Laura Norén, Obsidian Security

Daniel Susser, Penn State U.

Document is at:

https://docs.google.com/document/d/1-fq-_plSw8JlL99UJJGdCBtr0VXNF_S5pqlzUo5XXPg

We want your input!!