

Academic
Data Science
Alliance

Academic Data Science Alliance Data Science Ethos Lifecycle

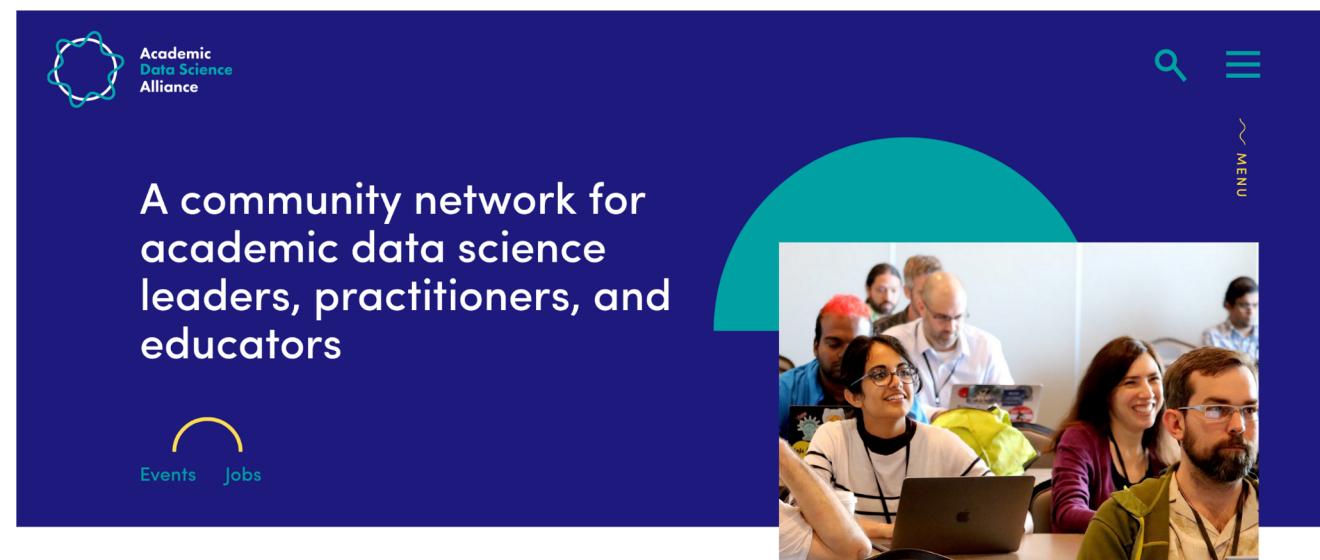
Micaela Parker - Executive Director
academicdatascience.org

July 2022 ~~~ AI/ML Ethics Workshop

About ADSA

a **community-building organization** that supports university researchers and educators to **learn, use, and teach** data-intensive methodologies and responsible applications

An emerging professional society for academic data science



The screenshot shows the ADSA website homepage. At the top left is the ADSA logo and the text "Academic Data Science Alliance". On the right are search and menu icons. A large teal circular graphic overlaps the main content area. The central text reads "A community network for academic data science leaders, practitioners, and educators". Below this are "Events" and "Jobs" buttons. To the right is a photograph of several people at a conference table, and the number "2" is in the bottom right corner.



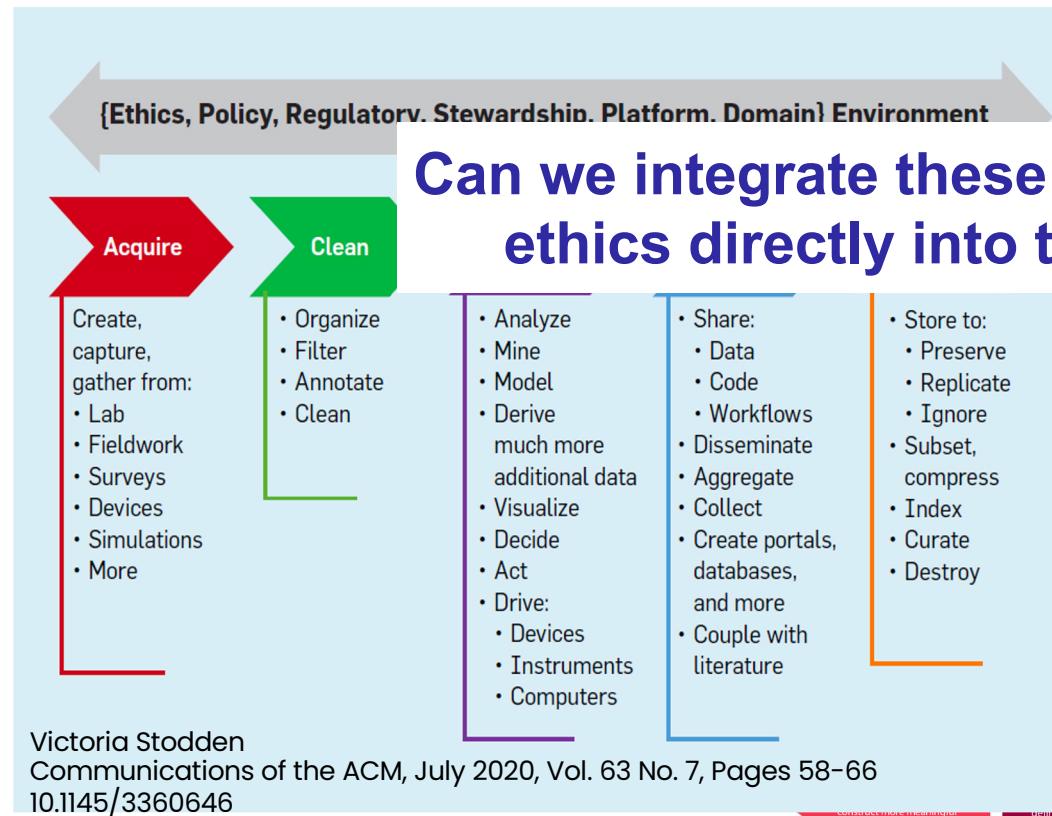
Vision

A community of researchers and educators who take responsibility for a just, equitable future where data science approaches are thoughtfully applied in all domains for the benefit of all.

Mission

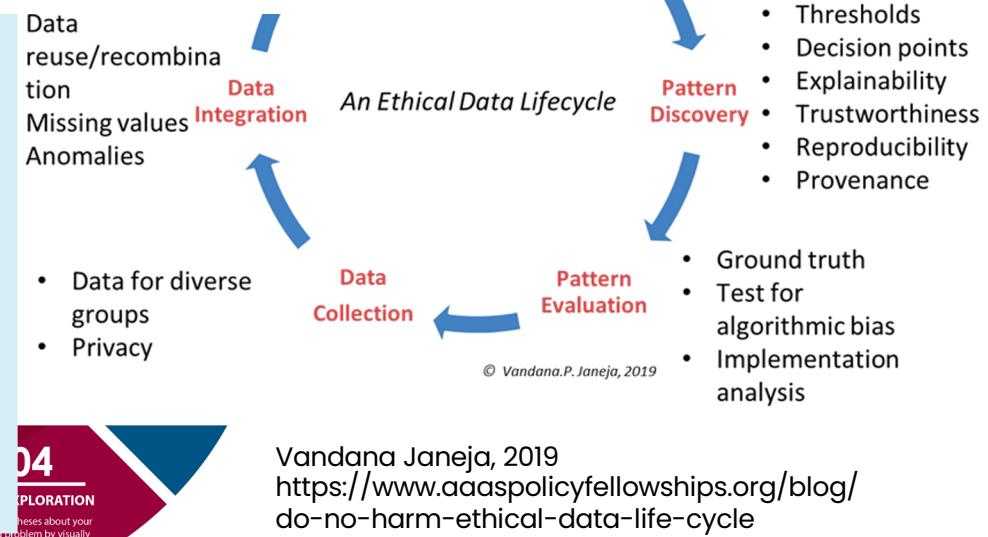
To connect, engage, and empower data science researchers and educators to discover, build, and share resources and opportunities.

Data Science Lifecycle + Ethics



Can we integrate these human contexts and ethics directly into the data lifecycle?

- Selection by demographic groups
- Bias in data selection
- Sampling strategy
- Selection bias
- Geo and time context
- Grouping



Victoria Stodden
Communications of the ACM, July 2020, Vol. 63 No. 7, Pages 58-66
10.1145/3360646

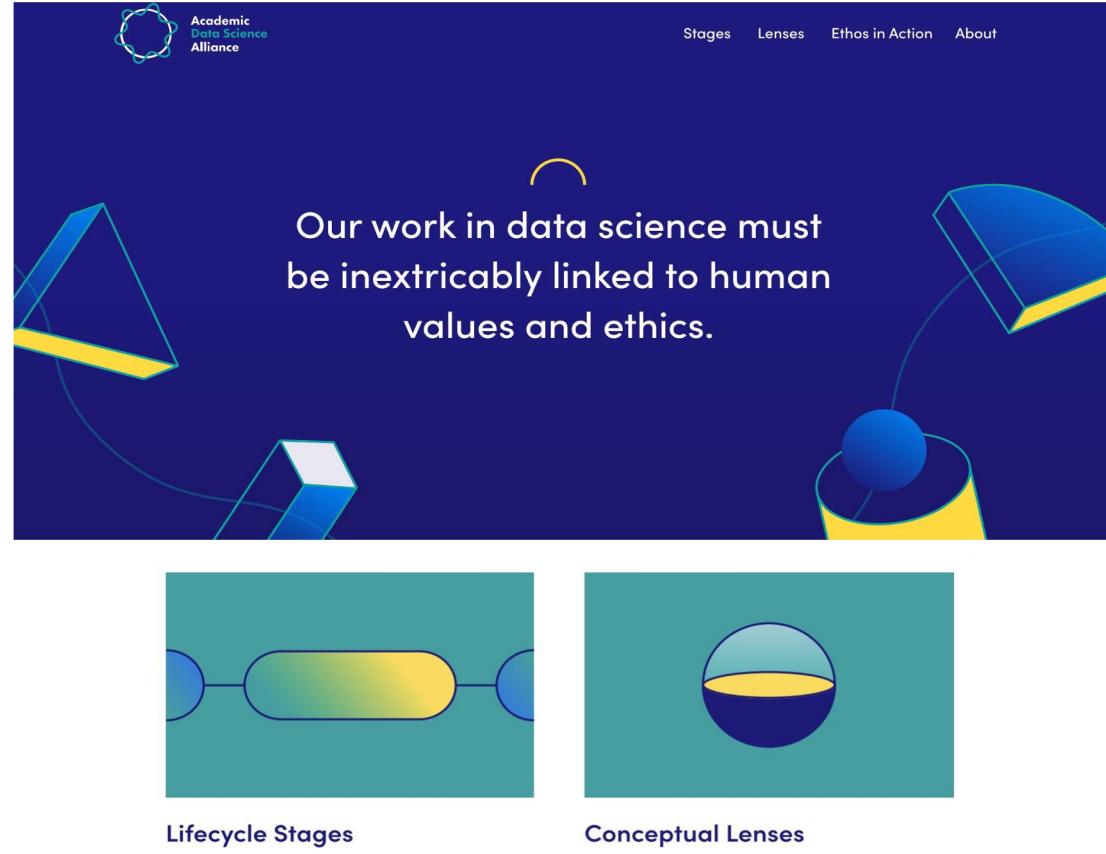


<https://medium.com/co-learning-lounge>

Data Science Ethos Lifecycle

(ethos = the spirit or character of a people)

a more **true-to-life model** of the data science lifecycle that shows how **societal questions** are a constitutive part of the day-to-day work of a data scientist.



Data Science Ethos Lifecycle - Why?

“Teaching responsible workflow in data science, (...) is a topic that invites challenging questions about the intersection of ethics (responsibility) and practice or process (workflow), which is captured in the term ethos.

- What does it mean for a data science workflow to be responsible?
- To whom or to what does a data scientist owe responsibility?
- How does the data scientist's responsibility relate to their workflow?
- And how can attention to the data science workflow support the teaching and learning of responsible data science?”

in Boening-Liptsin, Tanweer and Edmundson, 2022



Data Science Ethos Lifecycle - Who?

Authors and contributors

- Charlotte Cabasse-Mazel, Chair (dhCenter UNIL-EPFL)
- Margo Boenig-Liptsin (IEA de Paris)
- Cathryn Carson (UC Berkeley)
- Caren Cooper (NC State University)
- Alycia Crall (The Carpentries)
- Tiana Curry (ADSA)
- Ari Edmundson (UC Berkeley)
- Anna-Maria Gueorguieva (UC Berkeley)
- Eva Newsom (UC Berkeley)
- Carlos Ortiz (UC Berkeley)
- Stephanie Shipp (University of Virginia)
- Anissa Tanweer (University of Washington)
- You?

JOURNAL OF STATISTICS AND DATA SCIENCE EDUCATION
2022, VOL. JUST-ACCEPTED, NO. JA, 1-24
<https://doi.org/10.1080/26939169.2022.2089411>



Data Science Ethos Lifecycle: Interplay of ethical thinking and data science practice

Margarita Boenig-Liptsin^a, Anissa Tanweer^b, and Ari Edmundson^c

^a Institut d'Études Avancées de Paris; ^b University of Washington; ^c University of California, Berkeley



Data Science Ethos Lifecycle - What?

Goal: support data scientists to find their way to responsible data science practice

Framework: interactive online tool

Audience: data sciences practitioners and educators

Approach: make explicit considerations of salient human contexts and consequences at every stage of the lifecycle

Modality: cases studies



Stages

Each data science project goes through distinct stages.
Select a stage to explore its ethical dimensions.

Question Discovery Analysis Modeling **Interpreting** Sharing

Looking through conceptual lenses:

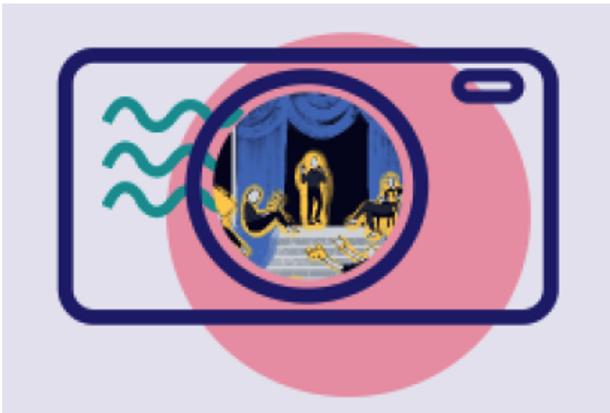
Positionality

Sociotechnical Systems

Power

Narratives

Ethos Lenses



- **Narrative:** prominently circulated or widely held beliefs about why the world is as it is (and what can be done to change it)
- **Sociotechnical systems:** the ways in which people and technologies are irrevocably entangled
- **Positionality:** the ways in which knowledge and actions are inescapably informed by a person's relative social circumstances
- **Power:** the asymmetric capacity to structure or alter the behavior of others



Data Science Ethos Lifecycle - How?

1. Evidence-based research

"Scholars have also focused specifically on **the merits of the lifecycle for addressing and teaching data ethics** (Janeja 2019; Keller et al. 2020). The lifecycle is useful for data scientists to "**develop ethical critical thinking** while analyzing data" (Janeja 2019) and to "implement an ethical review at each stage of research" (Keller et al. 2020). Both Janeja and Keller et al.'s approach shows **how ethical reflection paired with each stage of the lifecycle can help to keep important human questions close to the technical material and design decisions** at hand. "

in Boening-Liptsin, Tanweer and Edmundson, 2022



Data Science Ethos Lifecycle - How?

2. Consider human context at each stage with each lens

Example

Stage 1: Research Question



The **research question specifies the problem to be addressed**. Questions are **typically identified and described by the stakeholder(s)**, which can include a wide range of people from communities, researchers, activists, government officials, and corporate managers. These questions help to establish the boundaries of the data acquisition process that follows.

Questions to ask yourself during this stage:

Positionality Lens

- How does your positionality affect your understanding of the problem?
- What (cultural/gender) assumptions do you bring to the problem?
- How much agency do you have to redefine the problem?
- Do you have the most relevant expertise to address and frame this problem? Who else might have this expertise? Who is considered a relevant expert?
- Does the project plan incorporate regular checks, discussions, and documentation about the ethical dimensions of the project?



Positionality

The capacity of being aware of how the opportunities and limits of one's identity, expertise, and power inform one's perspectives, aspirations, and actions, and are shaped by one's relationships and environment.

Data Science Ethos Lifecycle - How?

3. Case Studies

- **Census Case Study**
 - census collection is very expensive and time-consuming
 - mobile phone records are being explored as a data source that could allow governments to quickly and cheaply estimate population and economic status

⚠ **Sociotechnical Systems:** phones are intertwined with people's personal lives, privacy concerns

⚠ **Narratives:** a utopian narrative about digital traces of human activity (the data are about customers and markets, not citizens)
- **Cook County Property Assessment System**
 - Chicago Tribune uncovered systemic bias and corruption that resulted in a regressive property taxation system perpetuating racist housing practices in Cook County. Property tax assessment is a data-intensive process. Cook County is revising their system. What do they need to consider?

⚠ **Positionality:** biases are propagated when social inequalities are reproduced in the data, or through proxies (attributes that correlate with protected classes, e.g. race, age, disability)
- **Crowd the Tap**
 - An EPA-funded citizen-science project that promotes access to safe drinking water by empowering individuals and groups to investigate the piping infrastructure that deliver drinking water to their own homes

Data Science Ethos Lifecycle - When?

Timeline

- Soft launch for trusted community feedback: late Summer 2022
- Full public availability: end of 2022

Feedback

Please reach out!

Website will have a feedback form

We welcome new Case studies





Thank you!



micaela@academicdatasience.org
steve@academicdatasience.org

Charlotte Mazel-Cabasse
charlotte.cabasse@gmail.com