

Data Basics: Ethically Making & Sharing Data

Cynthia Heider, Public Digital Scholarship Librarian
heider@upenn.edu

Find these slides & other materials at
<https://bit.ly/rdds-data-ethics>!

Themes



Data
Integrity

vs

Bias &
Potential
for Misuse

Security,
Privacy &
Surveillance

vs

Openness

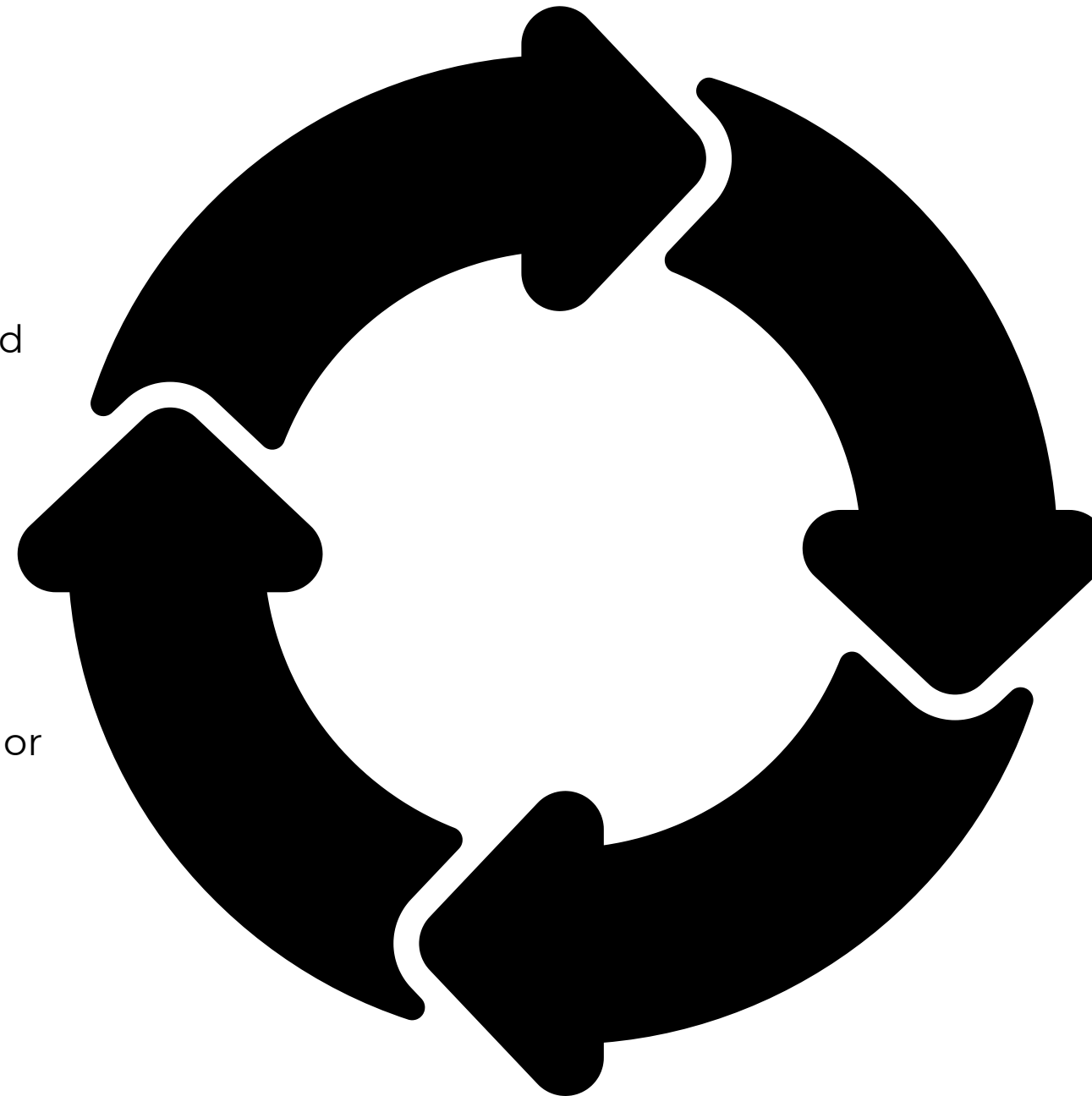
The Data Work Life Cycle

1. Acquisition

The stage at which people decide which data to collect and why (**conception**), determine how to collect them (**instrumentation**), and take action to obtain them (**collection**).

4. Disposition

The stage at which people **destroy** or **archive** data, either completely or partially.



2. Processing & Analysis

The stage at which people decide how to **tabulate** and **interpret** data. This includes determining which data to include or exclude in analyses, how to process the data to create new variables or measures like indexes, and which people will be involved in the process to decide what the data mean.

3. Dissemination

The stage at which people **publish** and **share** data and/or **report findings** from their analysis.

Principles & Precedents

Centering Ethics in Data Work

An Ethical Approach to Research:
“The Belmont Report” on Ethical Principles and Guidelines for the Protection of Human Subjects of Research (1974)



Beneficence

The commitment to maximize benefits and avoid causing harm to the extent possible, even if it is not a formal or legal requirement



Centering Ethics in Data Work

An Ethical Approach to Research:
“The Belmont Report” on Ethical Principles and Guidelines for the Protection of Human Subjects of Research (1974)



Respect for Persons

The responsibility to uphold people's power to make decisions that are in their best interest and to protect people who do not have that power



Centering Ethics in Data Work

An Ethical Approach to Research:
“The Belmont Report” on Ethical Principles and Guidelines for the Protection of Human Subjects of Research (1974)



Justice

The commitment to the fair distribution of burdens and benefits among people



Centering Ethics in Data Work

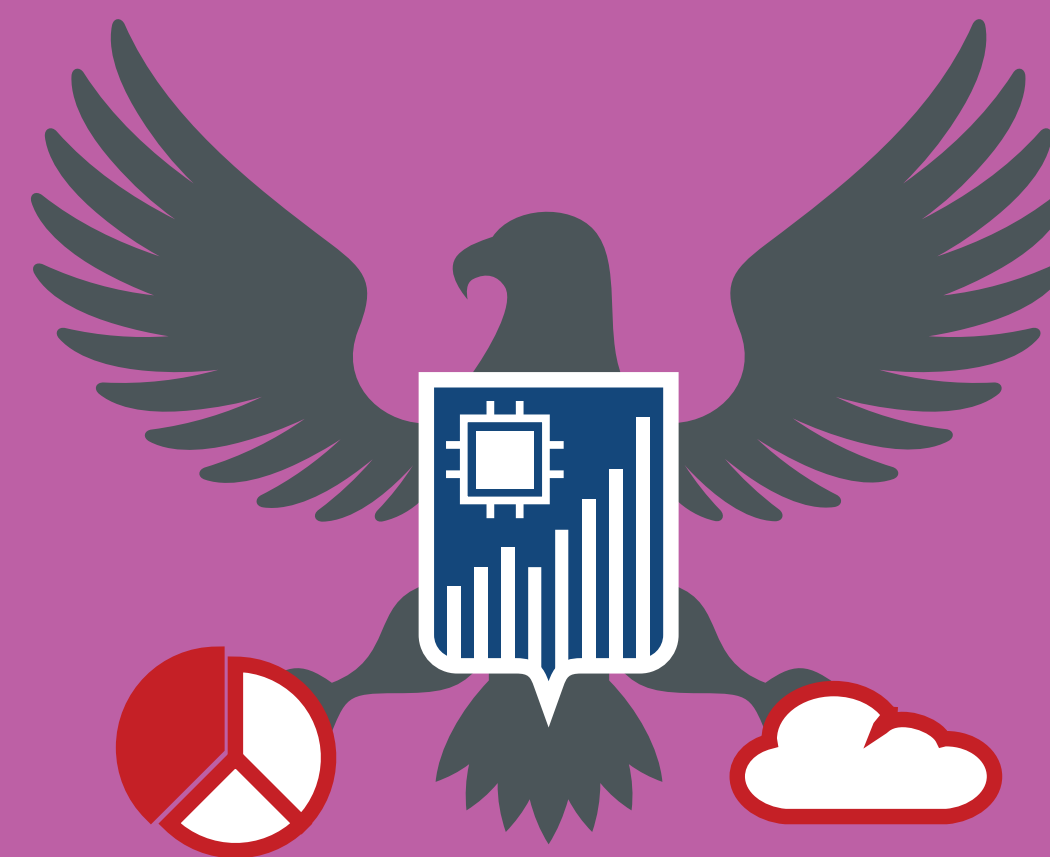
An Ethical Approach to Data:
Federal Data Strategy “Data Ethics Framework”
(2020).

“Data Ethics are the norms of behavior that promote appropriate judgments and accountability when acquiring, managing, or using data, with the goals of protecting civil liberties, minimizing risks to individuals and society, and maximizing the public good.” [Ref]

Uphold applicable statutes, regulations, professional practices, and ethical standards.

Respect the public, individuals, and communities.

Respect privacy and confidentiality.



Act with honesty, integrity, and humility.

Hold oneself and others accountable.

Promote transparency.

Stay informed of developments in the fields of data management and data science.

Centering Ethics in Data Work

A Feminist Approach to Data:
Data Feminism by Catherine D'Ignazio and
Lauren Klein (2020)

“We derived these principles from the major ideas that have emerged in the past several decades of intersectional feminist activism and critical thought. At the same time, we welcome the notion that there are many other possible starting points that share the end goal of using data (or refusing data) in order to end oppression.” [Ref]

Examine Power

Challenge Power

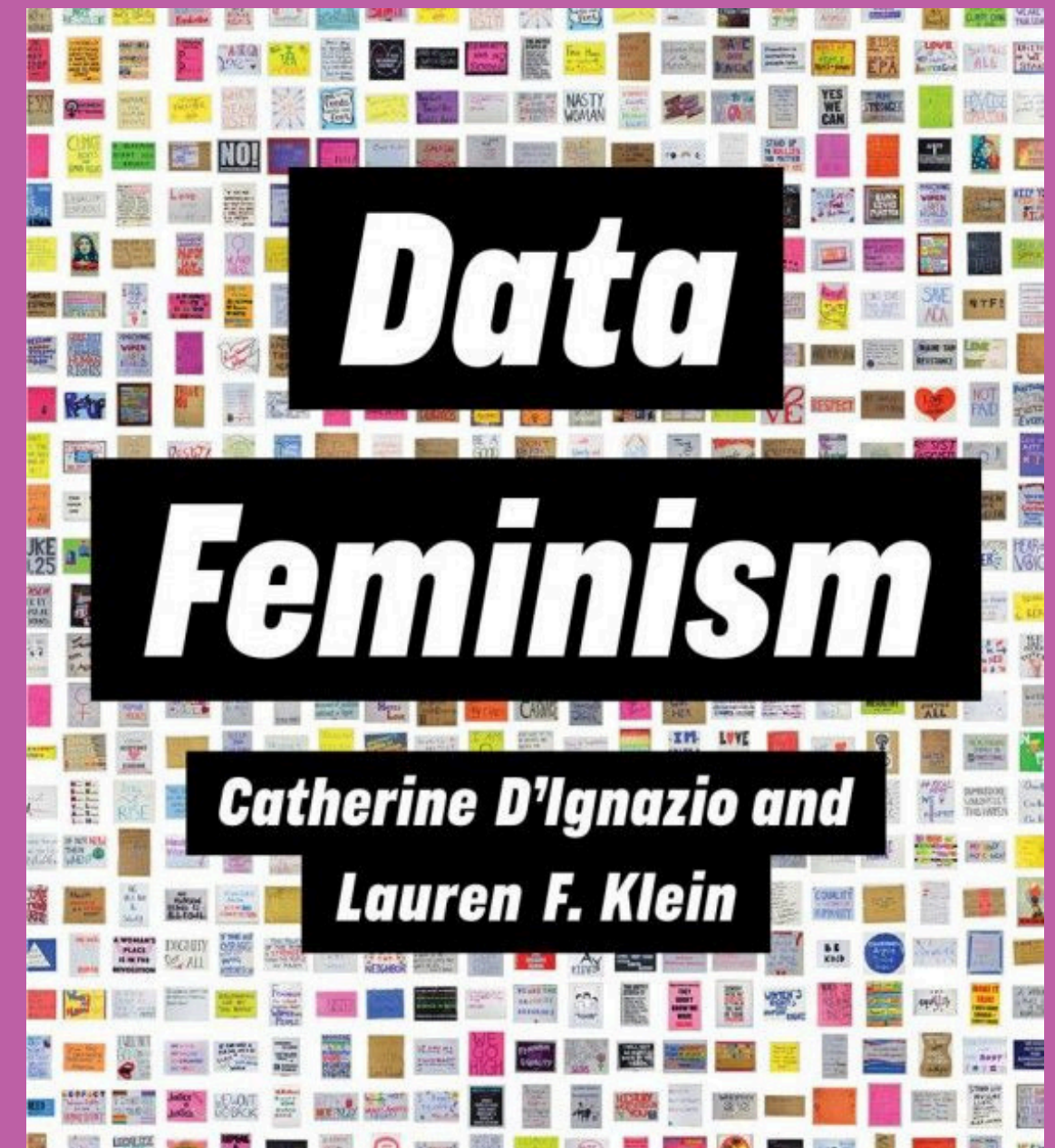
**Elevate Emotion &
Embodiment**

**Rethink Binaries &
Hierarchies**

Embrace Pluralism

Consider Context

Make Labor Visible



Putting It Into Practice

Centering Ethics in Data Collection Work

Implementing principles in the Acquisition phase of the data life cycle



Beneficence



Respect for Persons



Justice

- Involve stakeholders from the start
- Collect only the minimum viable data
- Require informed consent and “opt-in”
- Provide reciprocal value in exchange for data
- Anticipate potential security risks, such as re-identification

Example scenario

A team designing an in-home security system was charged with making it easier to use, while also offering a greater sense of protection. Users wanted the system to monitor activity in and around the home but also preserve privacy.

To understand the line between protection and privacy, the team mapped out all human interactions that occur daily at home. This allowed them to pinpoint times when visual recording was critical for protection and when it wasn't. Inspired by analog camera shutters, the team designed a “privacy shutter” to open and close automatically at specific points.

Centering Ethics in Data Analysis Work

Implementing principles in the Processing & Analysis phase of the data life cycle



Beneficence



Respect for Persons



Justice

- Forefront the human element inherent in data, rather than approach it as an abstraction
- Be transparent about the data's context and limitations
- Seek out and incorporate stakeholders' interpretation of the data
- Identify encoded bias and consider mitigation actions

Example scenario

A large equipment company wanted to explore the tradeoffs between usage-based repair and time-based.

Rather than present graphs and charts to show the difference between the two approaches, the team created two contrasting stories demonstrating what happened in each case. One told of an operator who overspent on parts that were replaced before new ones were needed. In the other scenario, the operator had to send his entire team home due to a part failure. Adding these simple, relatable stories helped everyone in the room understand the real-world implications of a data-driven solution.

Centering Ethics in Data Sharing

Implementing principles in the Dissemination phase of the data life cycle



Beneficence



Respect for Persons



Justice

- Incorporate quality control frameworks such as FAIR, CARE, and domain-specific specifications
- Account for how publication may reinforce inequities or close disparities
- Share to reduce the burden of duplicate data collection
- Return data and research results to stakeholders and contributors in a form they can use

Example scenario

In 2018, the City of Chicago launched CityKey, a single card that could serve as an ID, library card, and public transit farecard. The card was specifically designed for residents of Chicago who might have a hard time acquiring a driver's license or state ID—including undocumented immigrants.

In designing the CityKey, the City Clerk's office wanted to avoid replicating the experience New York City had faced with its own IDNYC card. New York City had to go to federal court to protect the personal information contained in its applicant database, which the Trump administration sought to use for immigration enforcement. Unlike New York's program, Chicago's CityKey database does not retain images of any documents, nor any personal information about CityKey holders.

Centering Ethics in Data Afterlives

Implementing principles in the Disposition phase of the data life cycle



Beneficence



Respect for Persons



Justice

- Empower individuals to order the destruction of their data
- Be transparent about the plans for the data after project conclusion
- Put in place accountability mechanisms to provide redress for harms that may arise from data misuse

Example scenario

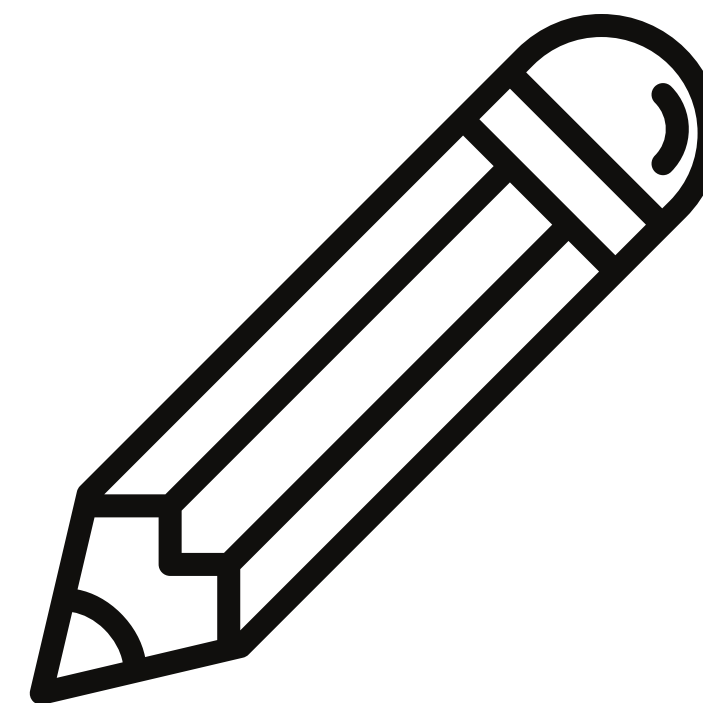
In 2010 Mario Costeja González filed a complaint with the Spanish Data Protection Agency (AEDP) against a local newspaper and Google Spain for claims relating to auction notices mentioning González published in 1998. The notices concerned real estate auctions held to secure repayment of González's social security debts. González contended that these pages were no longer necessary because "the attachment proceedings concerning him had been fully resolved for a number of years and that reference to them was now entirely irrelevant." He sought to have the local newspaper remove the pages or alter them so his personal information was no longer displayed. He also sought for Google Inc. to remove the links to the articles in question so that the information no longer appeared in Google Search results.

The AEDP dismissed the plaintiff's claims against the newspaper, but allowed those against Google.

Tools & Resources

The ODI Data Ethics Canvas

Graphic:
bit.ly/dataethicscanvas
Google Doc:
bit.ly/ethicscanvasdoc



Improve project planning

- Consider potential risks and impacts from the project start.
- Debate about project objectives, intention and impact.
- Consider wider ethical issues, relevant laws and regulation.

Grow impact and trust

- Design better products and services while addressing bias.
- Examine the impact that your data activities can have.
- Mitigate harmful impacts for people and communities.

Complement other ethics guidelines

- Tease out risks around a project, without predisposing an outcome.
- Illustrates how one type of data activity can have many outcomes and consequences, depending on context, purpose and organizations involved.

Manage data ethics in the long-term

- Manage and discuss data ethics.
- Develop ethical frameworks and guidance around your project.
- Raise awareness of data issues in your organisation.

Data Sources

- Name/describe your project's key data sources, whether you're collecting data yourself or accessing via third parties.
- Is any personal data involved, or data that is otherwise sensitive?

Rights around data sources

- Where did you get the data from? Is it produced by an organisation or collected directly from individuals?
- Was the data collected for this project or for another purpose? Do you have permission to use this data, or another basis on which you're allowed to use it? What ongoing rights will the data source have?

Limitations in data sources

- Are there limitations that could influence your project's outcomes?
- Consider:
 - > bias in data collection, inclusion/exclusion, analysis, algorithms
 - > gaps or omissions in data
 - > provenance and data quality
 - > other issues affecting decisions, such as team composition

Ethical and legislative context

- What existing ethical codes apply to your sector or project? What legislation, policies, or other regulation shape how you use data? What requirements do they introduce?
- Consider: the rule of law; human rights; data protection; IP and database rights; anti-discrimination laws; and data sharing, policies, regulation and ethics codes/frameworks specific sectors (eg health, employment, taxation).

Ongoing implementation

- Are you routinely building in thoughts, ideas and considerations of people affected by your project? How?
- What information or training might be needed to help people understand data issues?
- Are systems, processes and resources available for responding to data issues that arise in the long-term?

Your reasons for using data

- What is your primary purpose for collecting and using data in this project?
- What are your main use cases? What is your business model?
- Are you making things better for society? How and for whom?
- Are you replacing another product or service as a result of this project?

Positive effects on people

- Which individuals, groups, demographics or organisations will be positively affected by this project? How?
- How are you measuring and communicating positive impact? How could you increase it?

Negative effects on people

- Who could be negatively affected by this project?
- Could the way that data is collected, used or shared cause harm or expose individuals to risk of being re-identified?
- Could it be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)?
- How are limitations and risks communicated to people?

Minimising negative impact

- What steps can you take to minimise harm?
- How could you reduce any limitations in your data sources? How are you keeping personal and other sensitive information secure?
- How are you measuring, reporting and acting on potential negative impacts of your project?
- What benefits will these actions bring to your project?

Reviews and iterations

- How will ongoing data ethics issues be measured, monitored, discussed and actioned?
- How often will your responses to this canvas be reviewed or updated? When?

Engaging with people

- How can people engage with you about the project? How can people correct information, appeal or request changes to the product/service? To what extent?
- Are appeal mechanisms reasonable and well understood?

Communicating your purpose

- Do people understand your purpose – especially people whom the data is about or who are impacted by its use?
- How have you been communicating your purpose? Has this communication been clear?
- How are you ensuring more vulnerable individuals or groups understand?

Openness and transparency

- How open can you be about this project? Could you publish your methodology, metadata, datasets, code or impact measurements?
- Can you ask peers for feedback on the project?
- How will you communicate it internally?
- Will you publish your actions and answers to this canvas openly?

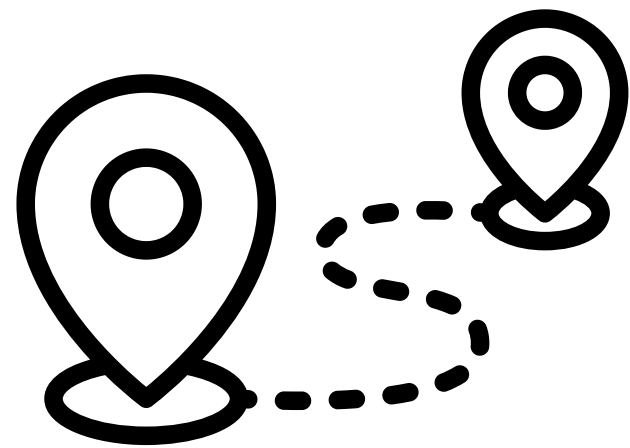
Sharing data with others

- Are you going to be sharing data with other organisations? If so, who?
- Are you planning to publish any of the data? Under what conditions?

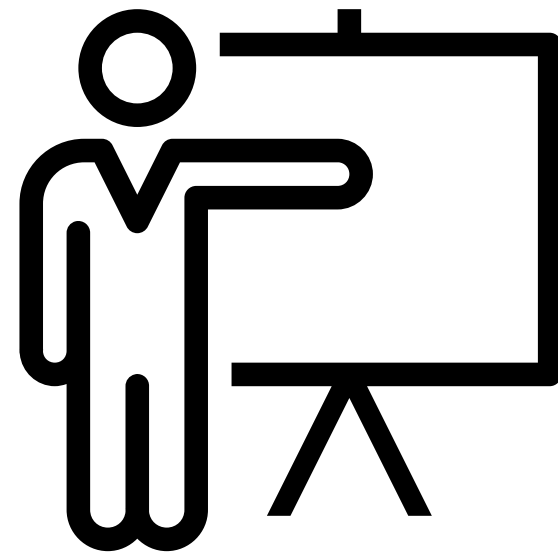
Your actions

- What action will you take before moving forward with this project?
- Will you openly publish your actions and answers to this canvas?

Penn Resources



[Data Management
Resources Libguide](#)



[RDDS Data
Workshops and
Consultations](#)



[Office of Research
Services](#)