

Open Data Day Webinar: Future-Proofing Data Ethics

Cathi Greenwood, March 5, 2024



Cathi Greenwood

- WA state's Open Data Program Manager.
- Certified Data Ethics Professional (certification in process).

Topic:
Collecting demographic data to measure equity.

Tool:
The Open Data Institute (ODI)'s Data Ethics Canvas.









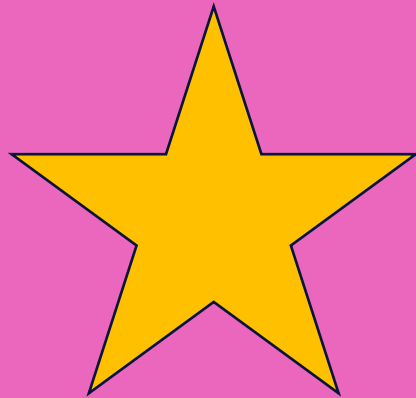
Measure service delivery equity

- Discover discriminatory practices.
- Do people get different service quality for reasons that do not align with the organization's mission or ethics?
- Data is required for measurement.

[Photo by Tim Mossholder on Unsplash](#)



Data Sources



Name/describe your project's key data sources, whether you're collecting data yourself or accessing via third parties.

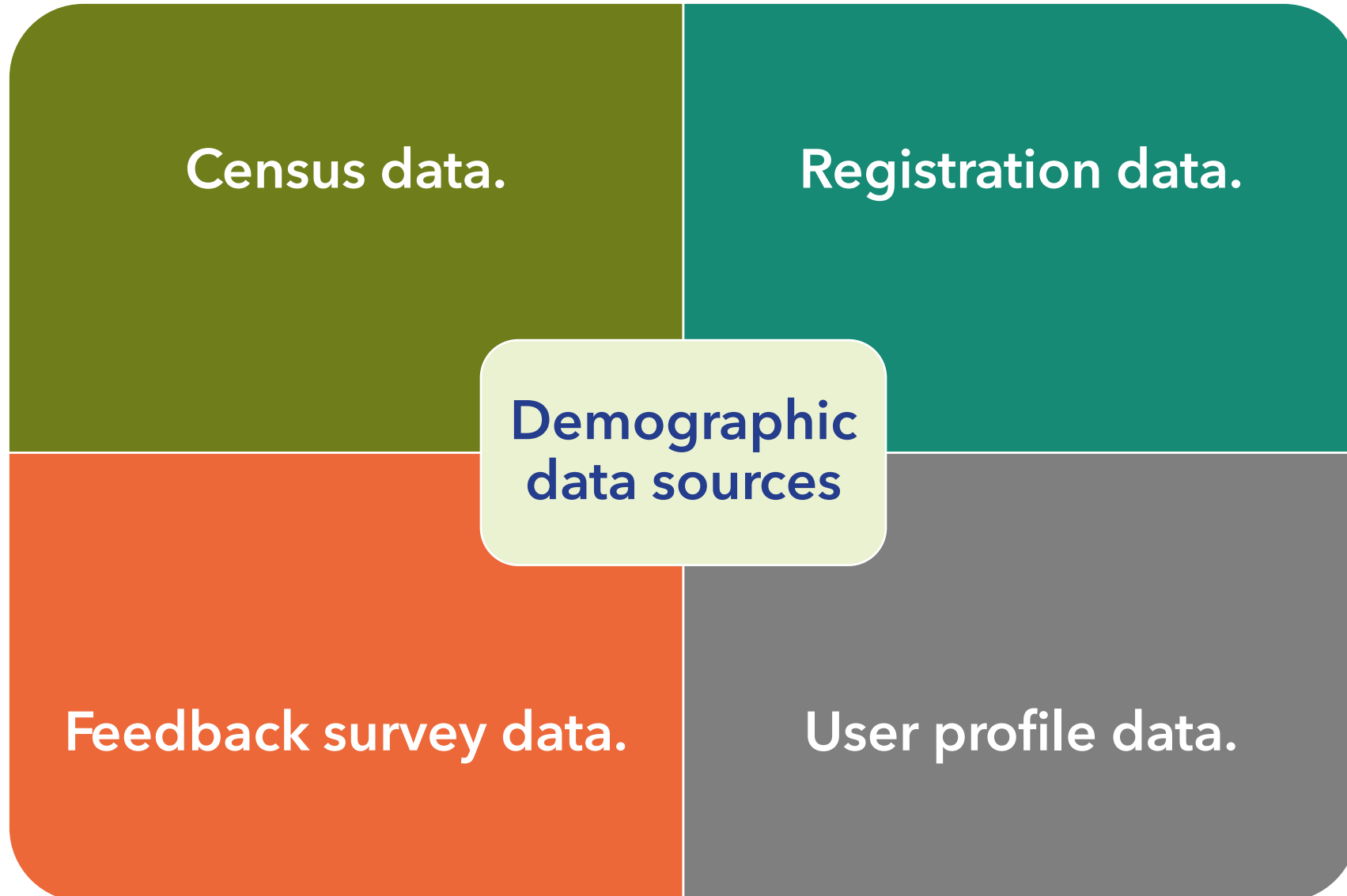


Demographic
data for
measuring
equity



"Word density cloud of sociodemographic characteristics." Image by Jun Li, Melasutra Md. Dali, and Nikmatul Adha Nordin.

Where can we source demographic data?



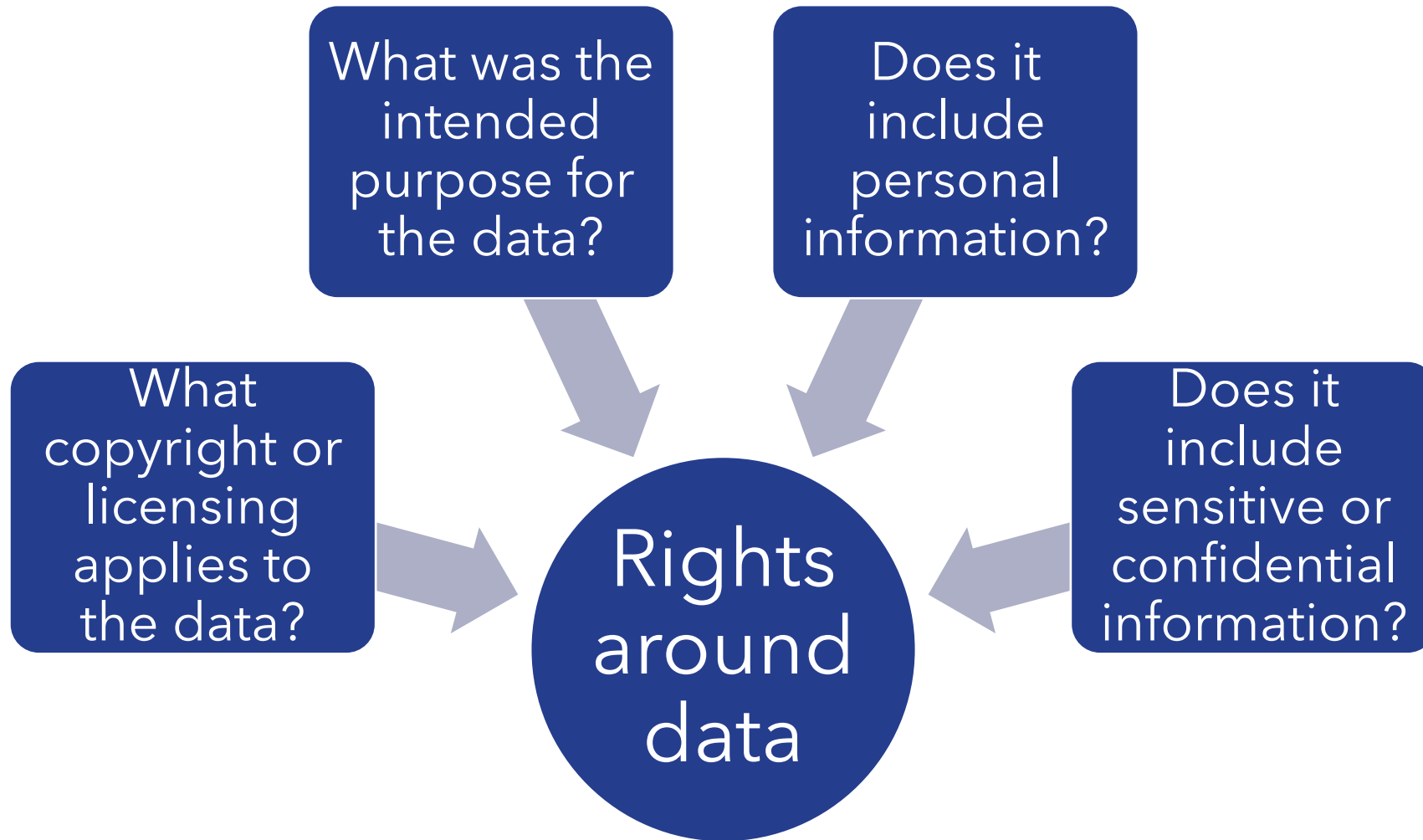
Rights around data sources



Where did you get the data from? Is it produced by an organisation or collected directly from individuals?



Image: <https://theodi.github.io/interactive-data-ethics-canvas/>



Why intended purpose matters

- What data is collected.
 - Gaps in the data lead to wrong conclusions.
- Data subjects' privacy expectations.
 - Who sees the data?
 - Are responses valid for the new purpose?
 - Is data sold without compensating the data subject?
 - Trust.

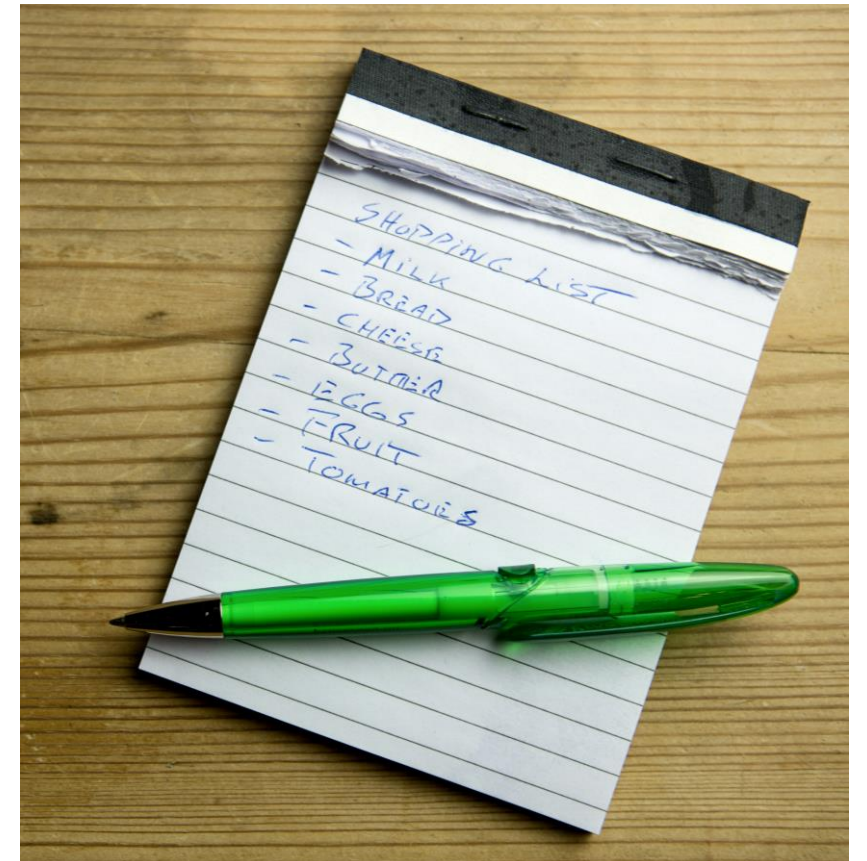


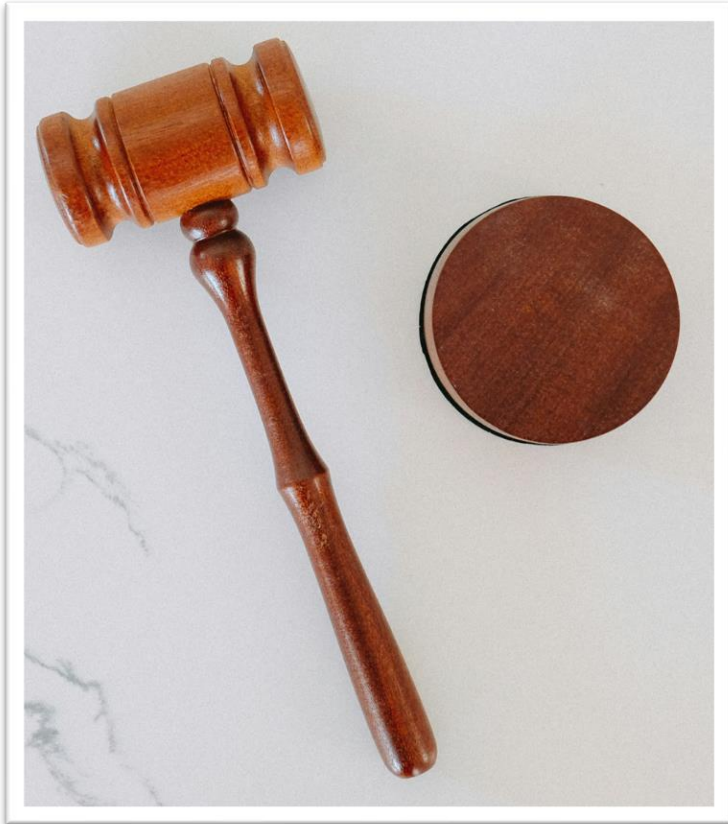
Photo by Torbjørn Helgesen on Unsplash

Use Washington State Privacy Principles

- Demographic data includes personal information.
 - It may also include personally identifiable information (PII).
- Follow privacy principles to protect personal information.



Photo by Hansjörg Keller on Unsplash



Lawful, fair, & responsible use

- Collection, use and disclosure is:
 - Based on legal authority;
 - Not deceptive;
 - Not discriminatory or harmful; and
 - Relevant and reasonably necessary for legitimate purposes.



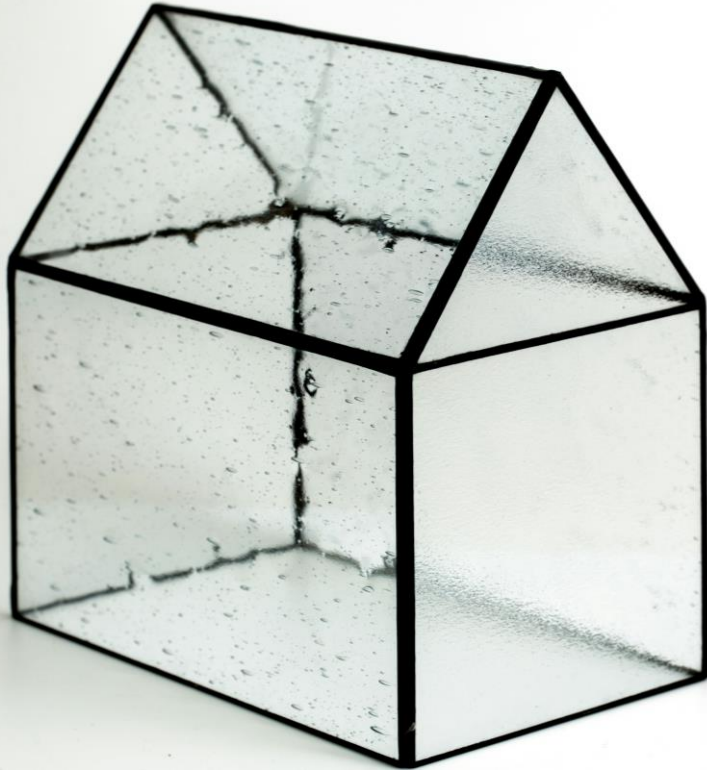
Data minimization

- The minimum amount of information is collected, used, or disclosed to accomplish the stated purpose for collecting the information.



Purpose limitation

- The reasons for gathering information are identified before it is collected.
- Use and disclosure is limited to what is reasonably necessary in relation to the specific reasons the information was collected.



Transparency & accountability

- Transparency means being open and transparent about what personal information is collected, for what purposes, and who it is shared with under what circumstances.
- Accountability means being responsible and answerable for following data privacy laws and principles.



Due diligence

- Taking reasonable steps and exercising care before and after entering into an agreement or arrangement with a third party that includes sharing personal information.



Individual participation

- Give people control of their information when possible.



Security

- Appropriate administrative, technical and physical security practices to protect the confidentiality, integrity, availability and control of personal information.

Limitations in data sources



Are there limitations that could influence your project's outcomes?



Image: <https://theodi.github.io/interactive-data-ethics-canvas/>

Incomplete demographic categories

- Many people do not see themselves represented in demographic questions.

Opt outs

- Many people opt out of responding to demographic questions.

Invisible populations

- People with rare characteristics are excluded from anonymized data.

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?



Image: <https://theodi.github.io/interactive-data-ethics-canvas/>

Executive Order 22-04 Washington State Pro-Equity Anti-Racism (PEAR) Playbook

- Agencies **must gather data** and report to the Office of Equity on their Pro-Equity Anti-Racist (PEAR) progress annually.

RCW 43.06D.900 Findings - Intent [of establishing the Office of Equity]

- "A more inclusive Washington is possible if agencies identify and implement effective strategies to **eliminate systemic inequities.**"

RCW 49.60.400 Discrimination, preferential treatment prohibited.

- "**The state shall not discriminate against**, or grant preferential treatment to, any individual or group on the basis of race, sex, color, ethnicity, or national origin in the operation of public employment, public education, or public contracting."

RCW Chapter 42.56 Public Records Act

- Transparency laws to **provide data to keep the public informed.**

Ongoing Implementation



Are you routinely building in thoughts, ideas and considerations of people affected by your project? How?



Image: <https://theodi.github.io/interactive-data-ethics-canvas/>

Data sources

- Who needs to know about them?
- What do they need to know?

Rights around data

- What training is needed?
- What business processes?

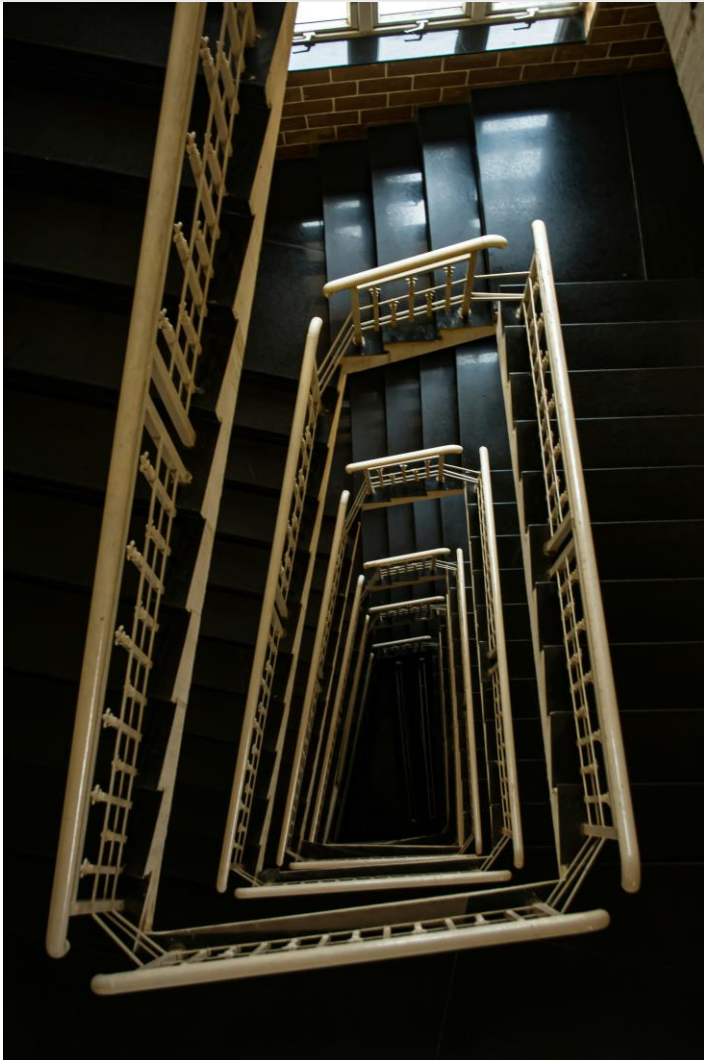
Ethics & Law

- How can we comply with law?
- How can we be ethical with the data?

Ongoing implementation

- How are you routinely building in thoughts, ideas, and considerations of people affected by your project?





Routine review

- Regularly explore alternative methods to measure equity that do not put residents' data at risk.

Privacy principle

- Lawful, fair, & responsible use

Pros

- Ensures that best practices are being followed.
- New opportunities that may be more effective or less risky are considered.

Cons

- Research takes time.

Long-term perspective

- Encourage a long-term perspective of how the data can be used for good and how it can cause harm if misused.

Privacy principle

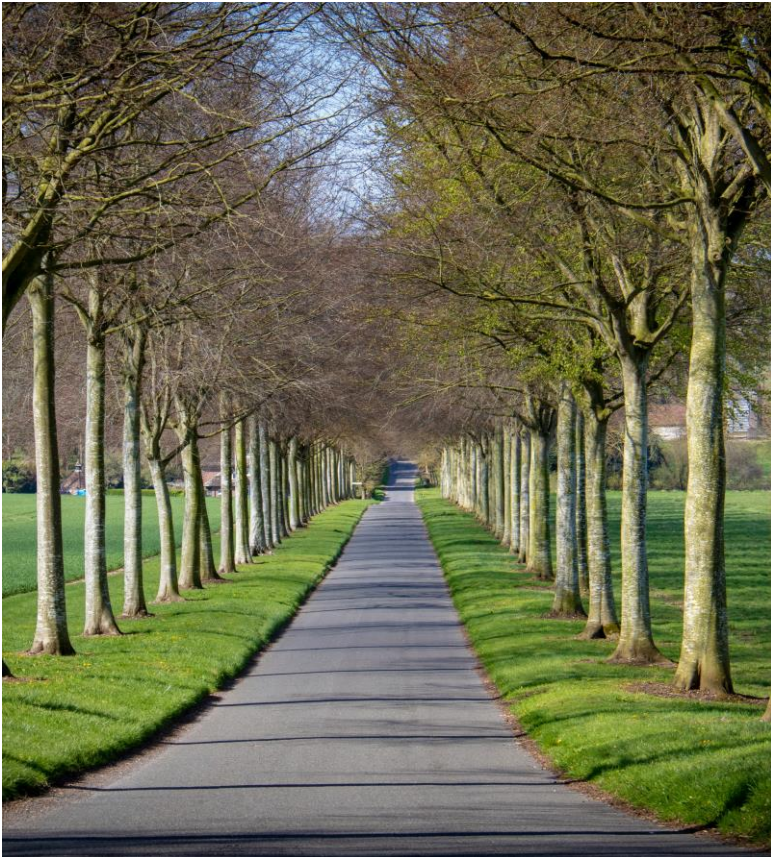
- Transparency & accountability

Pros

- Foresee and prevent potential problems.
- Foresee and maximize potential benefits.

Cons

- Buy-in can be challenging.



Ongoing implementation

- What information or training might be needed to help people understand data issues?





Metadata

- Maintain metadata on all collected data that identifies why it is being collected, under what authority, what its retention is, how it is meant to be used, any limitations in the data, and what the data does and does not mean.
- Regularly review this for whether the data collection should continue and whether it is being used or retained inappropriately.

Privacy principle

- Purpose limitation

Pros

- Single source of answers.
- Everyone knows what data the organization has and how it's used and protected.
- Useful enterprise-wide.

Cons

- Must establish processes to maintain metadata₃₂

Ongoing implementation

- **What systems, processes, and resources are available for responding to data issues that arise in the long-term?**





Data governance board

- Institute a data governance board that regularly reviews reported issues related to the data, and reviews whether the data use is effective.

Privacy principle

- Transparency & accountability

Pros

- Established responsibilities.
- Routine processes.
- Organization leaders have the impacts of this data at the front of their minds.
- Useful enterprise-wide.

Cons

- Time and commitment to establish a board.



Data stewards

- Assign data stewards and empower them to ensure the data is used and protected appropriately.

Privacy principle

- Transparency & accountability

Pros

- Established responsibilities.
- Routine processes.
- One or more staff has authoritative knowledge about the data and has executive-backed responsibility for the data.
- Useful enterprise-wide.

Cons

- Time and commitment to establish data stewards.³⁵

Metrics & Dashboards

- Create metrics and dashboards that demonstrate that the metadata and data lineage are complete and that retention schedules are followed.

Privacy principle

- Transparency & accountability

Pros

- Monitor health and pace of improvements.
- Quickly identify data management process issues to correct them while the problem is small.

Cons

- Time to define and build effective measures.

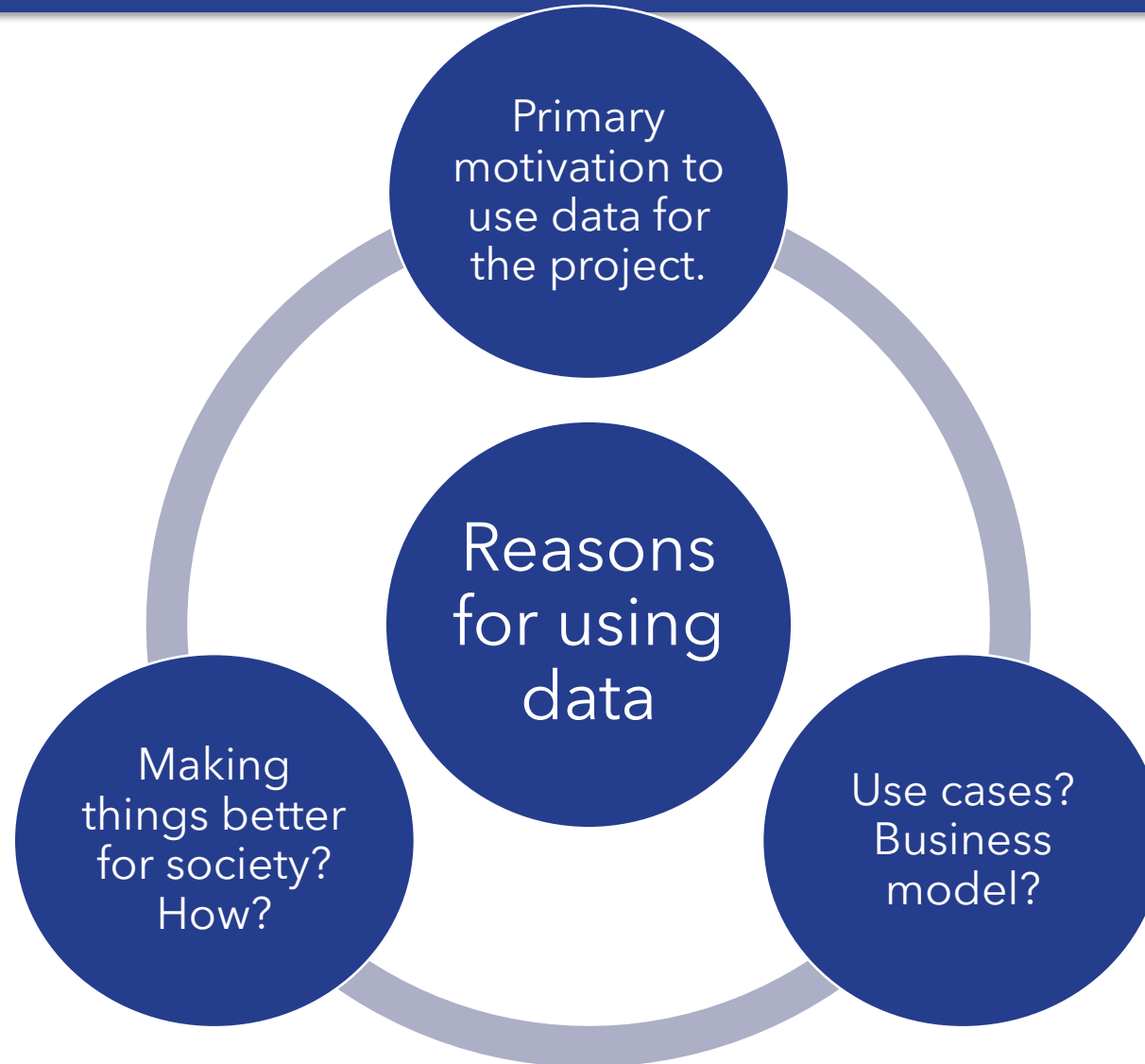
Your reason for using data



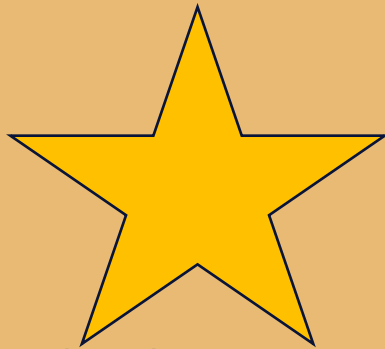
What is your primary purpose for collecting and using data in this project?



Image: <https://theodi.github.io/interactive-data-ethics-canvas/>



Positive effects on people



Which individuals, groups, demographics, or organisations will be positively affected by this project? How?



Who will the project affect positively?

- State residents, vendors, & employees.

How to measure, communicate, and increase positive impact

- Inclusive design.
- Measures reflect residents' definitions of equity.
- Public dashboard.
- Train state employees on measures and intended outcomes.

Positive effects on people:

- What methods will you use to measure, communicate, and increase positive impact?





Use the data

- Regularly review the data for gaps in service equity.
- Congratulate those who found the gap.
- Institute measures to close the gap.
- Monitor progress on closing it.
- Adjust as needed.
- Celebrate when it's closed.

Privacy principle

- Purpose limitation

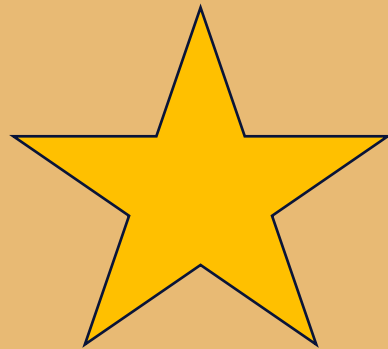
Pros

- Get the benefits that make the risk worth taking.
- Ensure the data is collected for a purpose and used for that purpose.
- Do not put people at risk and then give them no benefit for it.

Cons

- None.

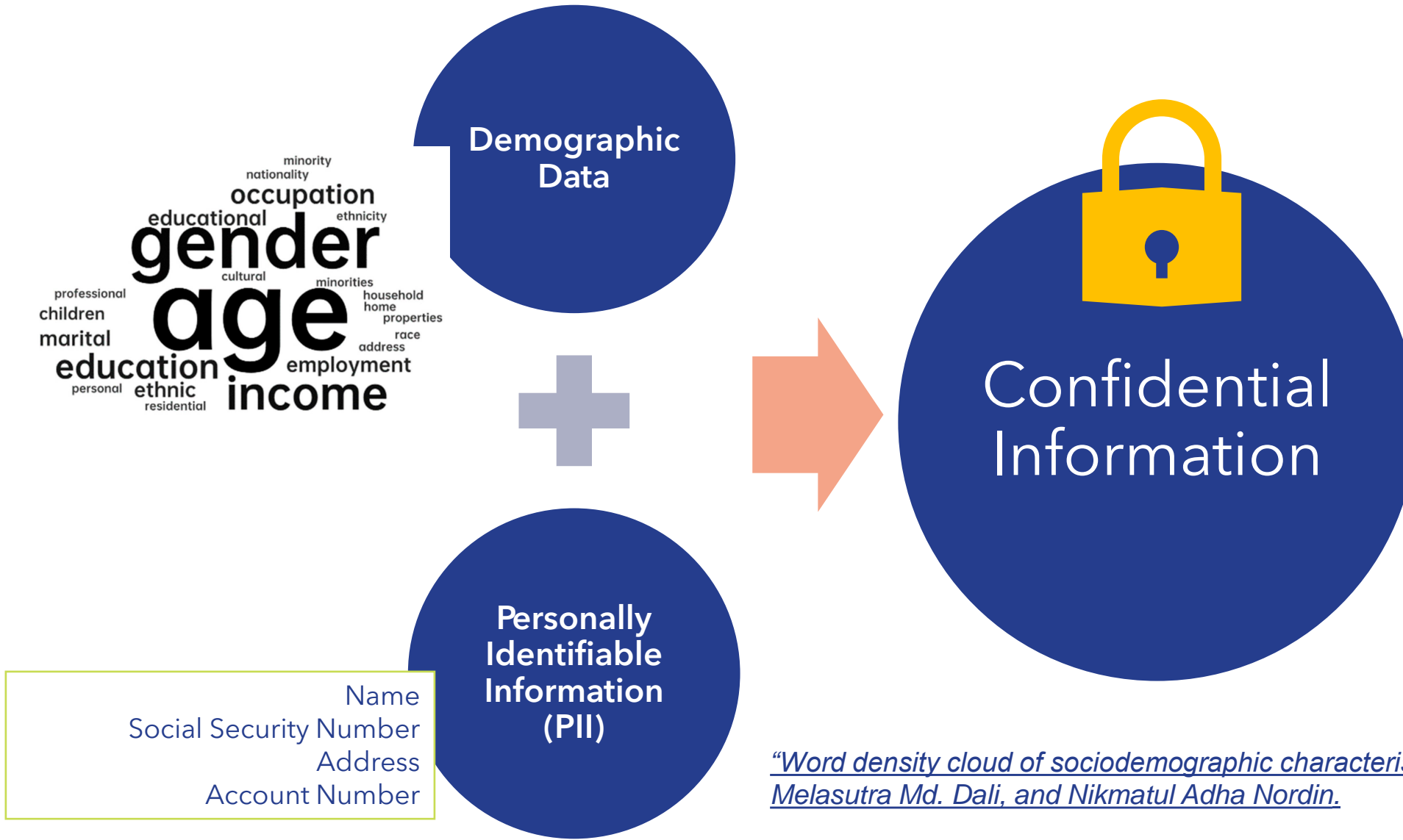
Negative effects on people



Who could be negatively affected by this project?



Image: <https://theodi.github.io/interactive-data-ethics-canvas/>



“Word density cloud of sociodemographic characteristics.” Image by Jun Li, Melasutra Md. Dali, and Nikmatul Adha Nordin.

What is demographic data misuse?

Using data for something other than its intended purpose.

Putting someone at risk by revealing their information.

Conflict with the privacy notice.

Targeting a demographic for a reason that doesn't align with the mission.

Identity theft.

Who can be negatively impacted? How?

- Members of vulnerable demographic groups.
 - Increase vulnerability if identities cannot be kept secure.
- Whistleblowers.
 - May be targeted, criticized, and shut down for bringing attention to risks.

Could the data be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)?

- Yes. In many governments, historically demographic has been used to target populations for deportation, internment, and medical experimentation.

Negative effects on people:

- How are limitations and risks communicated to people?



Meaningful informed consent

- Provide privacy practices and potential risks in plain language in the data subjects' primary language and in formats and forms that all residents can understand.
- Get new consent if risks or uses change.

Privacy principle

- Individual participation.

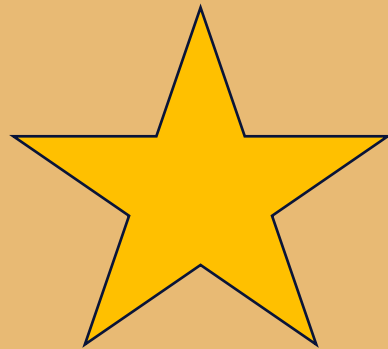
Pros

- Increase data subjects' confidence when providing personal data.
- Minimize reputational risk.
- Increase the effectiveness of the project.

Cons

- Time.

Minimising negative effects



What steps can you take to minimize harm?



Image: <https://theodi.github.io/interactive-data-ethics-canvas/>

Data reuse access methods

- Data breach.
- Public records request.
- Data sharing agreement.
- Open data or 'scraping' web pages.
- Access by researchers and other staff.
- Government mandate.

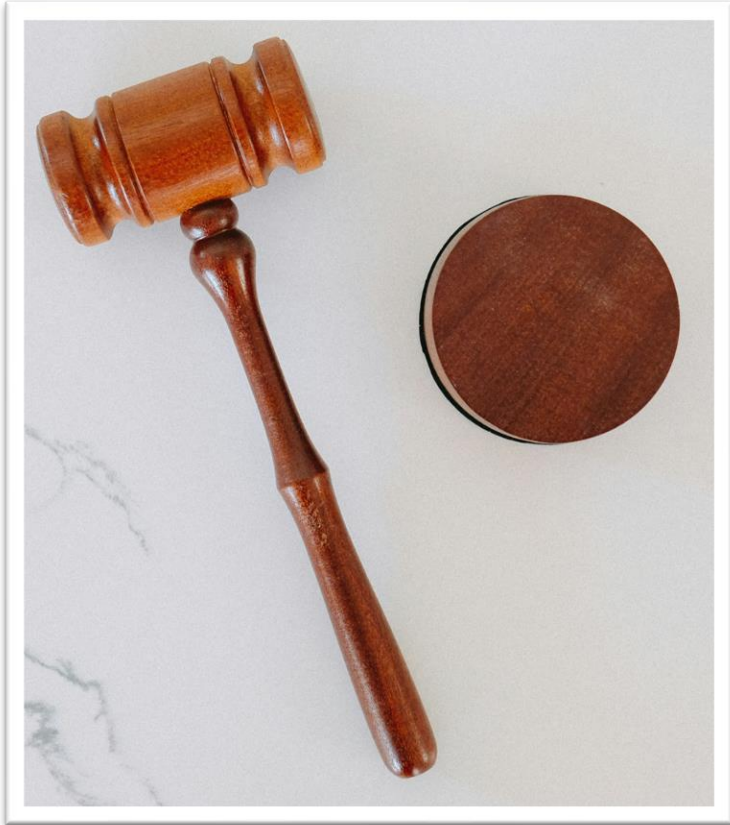
Photo by [Ryoji Iwata](#) on [Unsplash](#)



Minimize negative effects:

- What steps will you take to minimize negative effects?





Use laws

- Explicit legal protection from disclosure, sharing, and mandated sharing of the data that is being collected.

Privacy Principle

- Lawful, fair, & responsible use.

Pros

- Explicit defensible protection.
- Organizations prioritize legal compliance.
- Protect and hold accountable third parties who collect and analyze data on agencies' behalf.
- Increase data subjects' confidence when providing personal data.

Cons

- Takes a long time to establish new laws.
- Hard to change.
 - Lack of flexibility when circumstances change.

Law Recommendations

- Follow existing laws by categorizing collected data, storing the data in systems that are secured to store that data category, and training staff who handle the data on how to protect data of that category.
- Pass a law that details how data collected to measure service delivery equity shall be collected, protected, retained, de-anonymized, and destroyed.
- Pass a law that exempts from public disclosure and prevents reuse of data collected to measure service delivery equity.



Institutional Review Board (IRB) or Ethics Review Board

- Institute an IRB for this research that involves human subjects.
- The IRB will establish and review best practices throughout the data lifecycle.

Privacy Principle

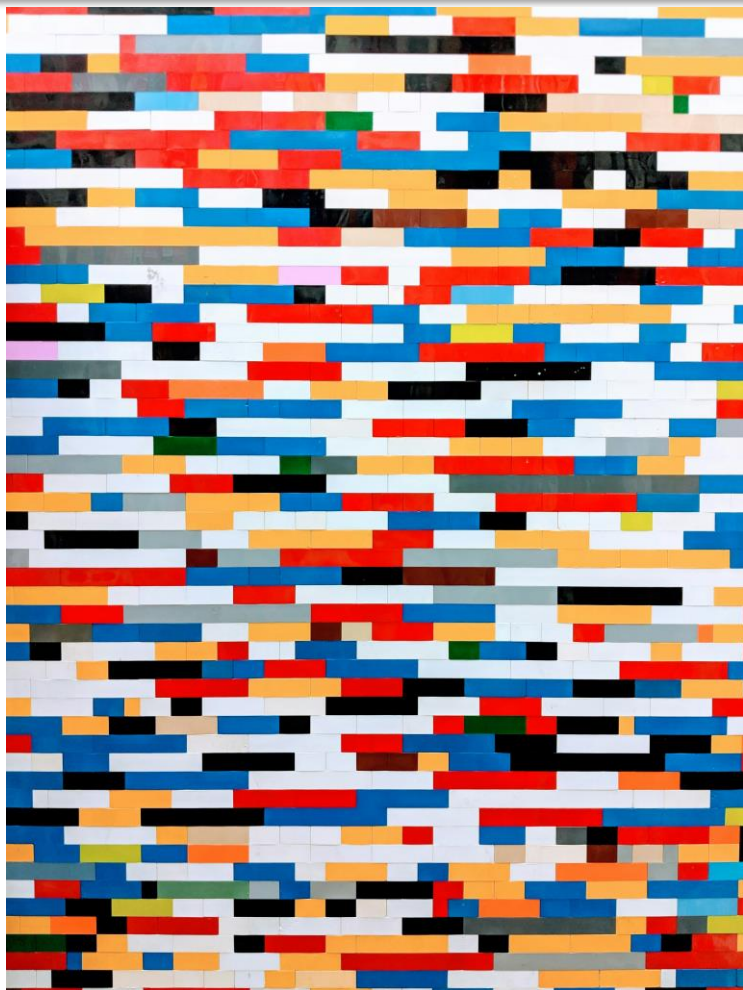
- Lawful, fair, & responsible use

Pros

- Expert direction focused on data subjects' safety.
- Requirements may be easier to prioritize.

Cons

- Time.
- Guidance can be rigid, not adapting to changing data subjects' needs.



Census blocks

- Only collect enough data from data subjects to know which census block they are in.
- Then measure equity using census data.

Privacy Principle

- Data minimization.

Pros

- No demographic data to protect.
- No consent to collect.
- Data source is trusted.

Cons

- May lack important demographic categories.
- Small demographic groups are invisible.
- Some people choose options they don't identify with.
- Some people do not respond.



Retention

- Set retention on the data for as short a time as is needed.
- Ensure the data is destroyed in the system of record and any copies at the end of retention.

Privacy Principle

- Data minimization.

Pros

- Less time when the data is a risk to manage.
- In case of a breach, a public disclosure request, or legal demand for the data, less data is impacted.

Cons

- Takes time to establish and apply detailed plans.
- Requires buy-in from everyone with access to the data.



Statistical Sampling

- Use statistical sampling instead of collecting demographic data from every data subject.

Privacy Principle

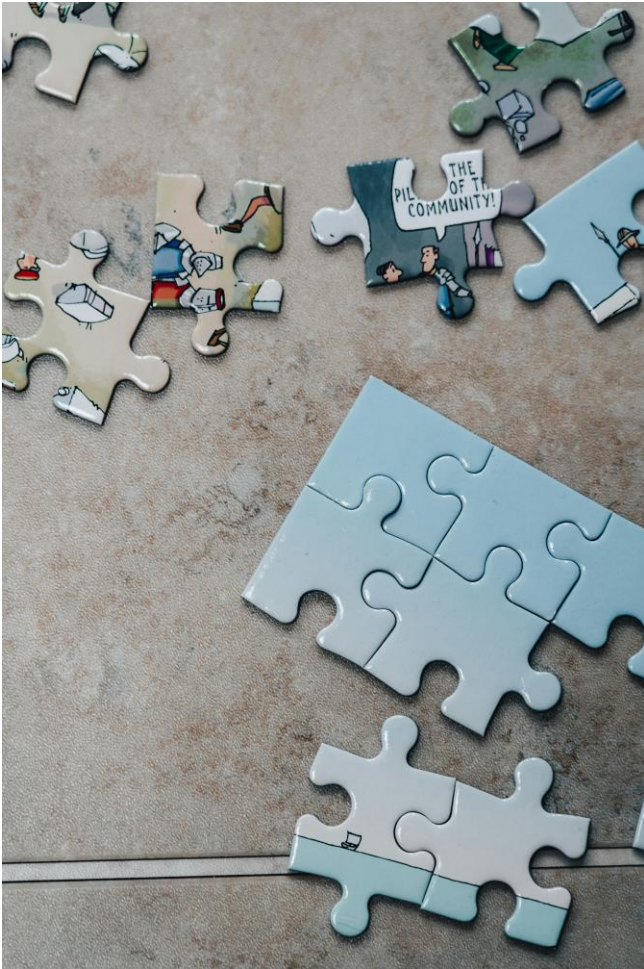
- Data minimization.

Pros

- Less data to protect.
- Efficient data footprint.

Cons

- Takes time and expertise to design effective statistical sampling.



Third-party data collection and analysis

- Ensure the third-party organization is invested in data subjects' positive outcomes and protection from negative outcomes.
- Contractually require the third party to only share anonymized data and to follow retention.
- Ensure the third party can legally follow the contract and not be obligated to share de-anonymized data under public disclosure laws.

Privacy Principle

- Data minimization.

Pros

- Less data to protect.
- The organization cannot disclose data that it does not have.

Cons

- Time to find an organization and draw up an effective contract.
- The organization and the data subjects must trust the third party.
- Requires coordinating with multiple third parties.



Data lineage

- Maintain data lineage on all of the data to ensure everyone can tell how the data is collected, who collects it, what systems it passes through, if and how it is transformed, and where the data goes and when it is destroyed.

Privacy Principle

- Purpose limitation.

Pros

- Ensure sources of data issues can be quickly located and corrected.
- Useful throughout the enterprise.

Cons

- Takes time and resources to design how to store the information.
- Must be maintained.



Access controls

- Institute robust access controls that ensure all access is authorized and tracked.

Privacy Principle

- Security.

Pros

- Data access is limited to those trained to understand the risks and responsibilities.
- If unauthorized data use is detected, the organization can identify who is responsible.

Cons

- Takes time and resources to place effective controls.
- Must be maintained.

Data encryption

- Encrypt data and follow cybersecurity protocols to ensure that any data accessed through a breach is unusable.

Privacy Principle

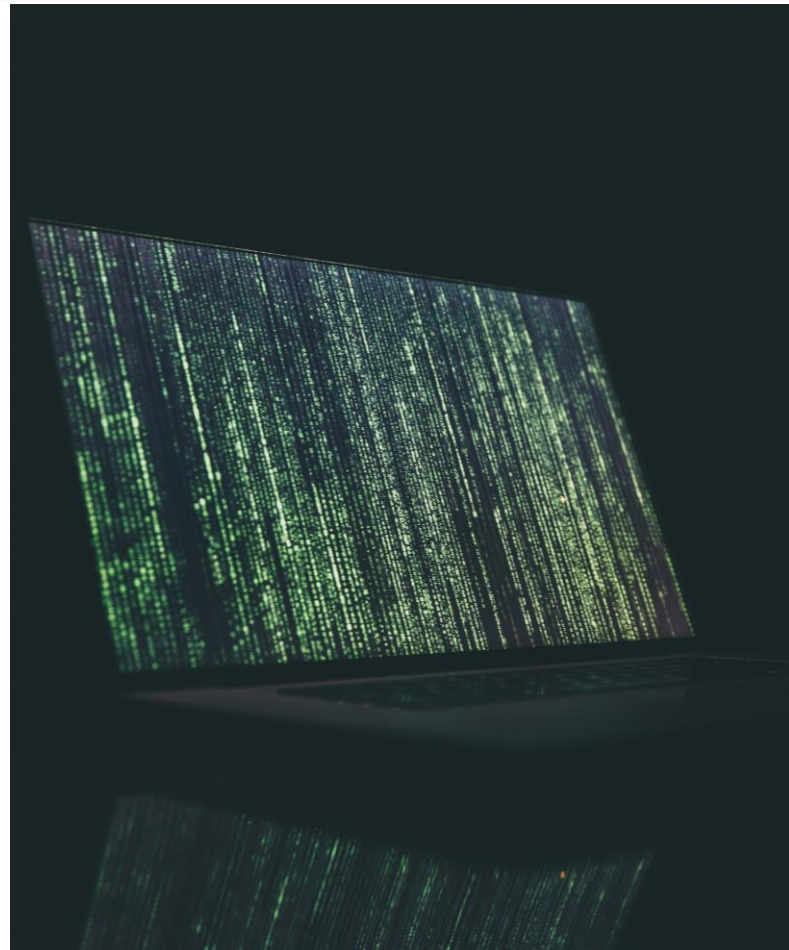
- Security.

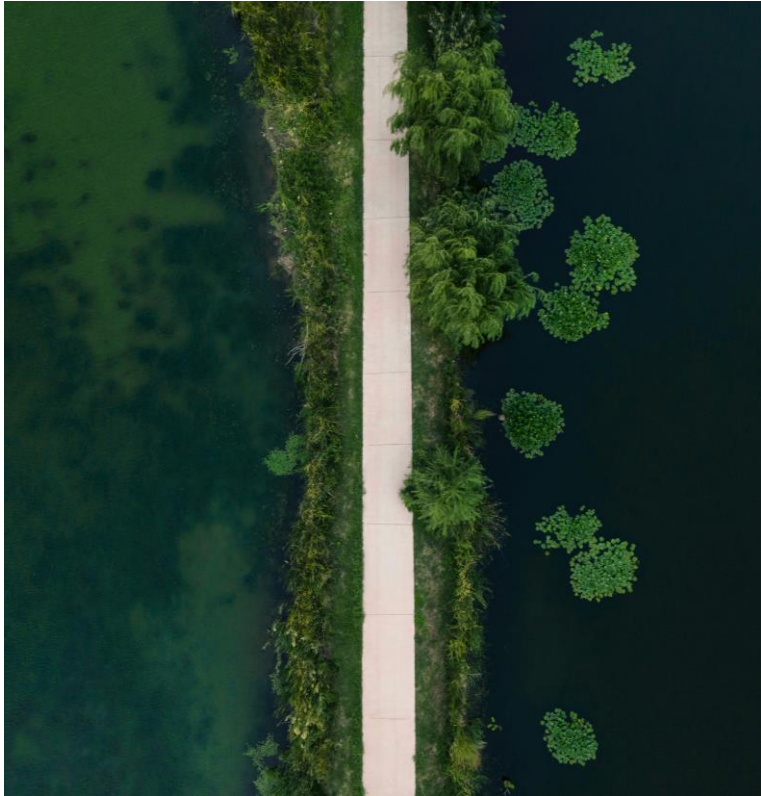
Pros

- Data acquired through unauthorized electronic access is unusable.

Cons

- None.





Collect and store separately

- Collect and store demographic data separate from personally identifiable data.

Privacy Principle

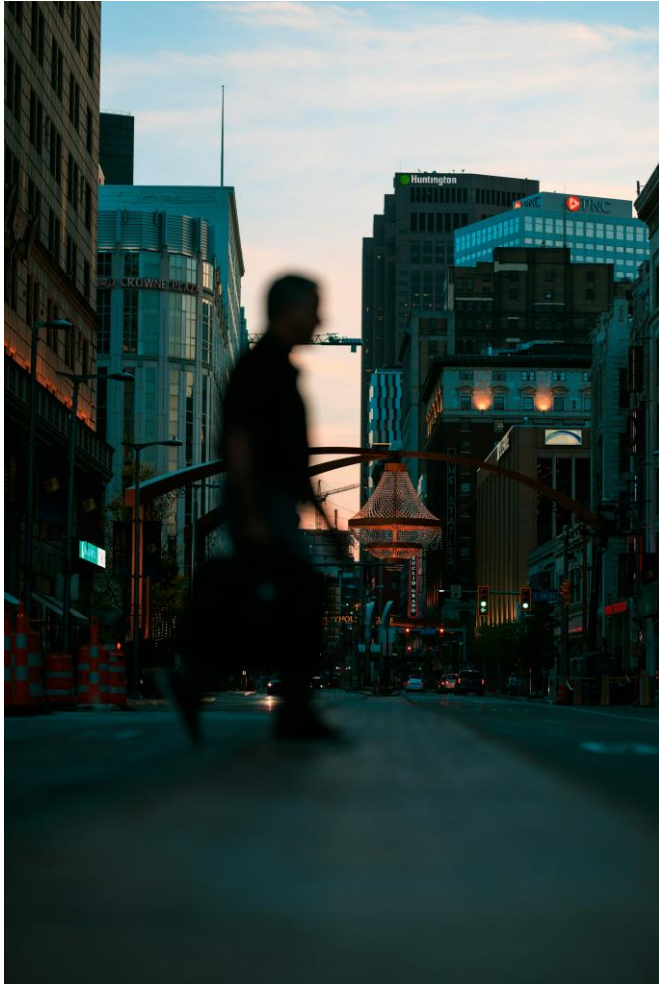
- Security.

Pros

- Personal information about individuals cannot be revealed.

Cons

- May need to collect demographic data multiple times.



Anonymize the data

- Whenever personally identifiable information is not needed, anonymize the data.

Privacy Principle

- Security.

Pros

- Reduced risk.
- Anonymized data is sharable and reusable.

Cons

- May need to collect data again.

Minimize negative effects:

- How will you measure, report, and act on potential negative impacts of the project?



Audits

- Perform regular audits of the data protection methods.

Privacy Principle

- Transparency & accountability

Pros

- Discover any issues while they are small.

Cons

- Time.



Reviews and iterations



How will ongoing data ethics issues be measured, monitored, discussed and actioned?



Reviews & Iterations:

- How will ongoing data ethics issues be measured, monitored, discussed and actioned?



Tools from other sections:

- Institutional Review Board (IRB) or Ethics Review Board
- Data governance board
- Data stewards
- Metrics & Dashboards



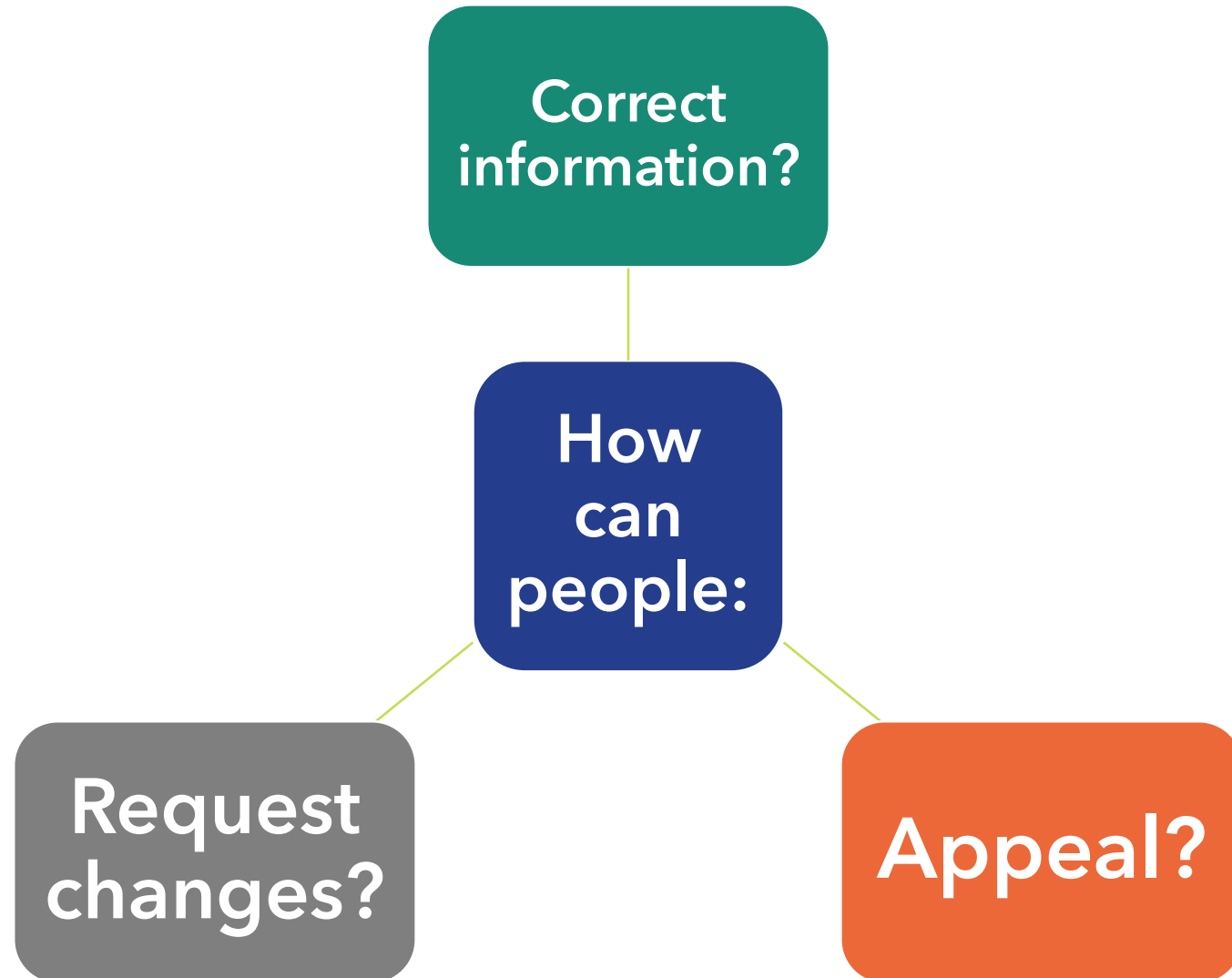
Engaging with people



How can people engage with you about the project?



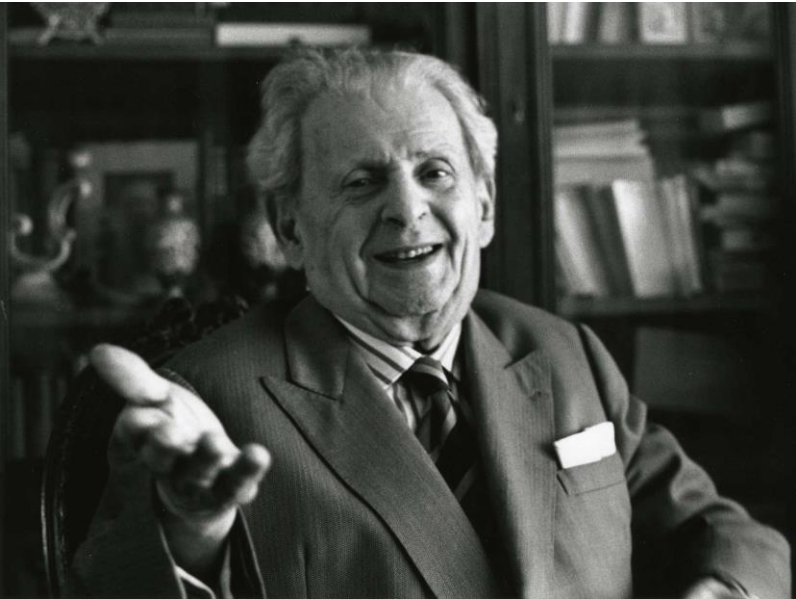
Image: <https://theodi.github.io/interactive-data-ethics-canvas/>



Engaging people:

- How can people engage with you about the project?





Levinasian principles

- See data subjects as Other rather than assimilated into the organization's culture.
- Have an infinite obligation of responsibility to the data subjects.
- Continuously engage with data subjects to understand their perceptions.

Privacy Principle

- Individual participation

Pros

- Increase trust.
- Deliver improved services.

Cons

- Time.

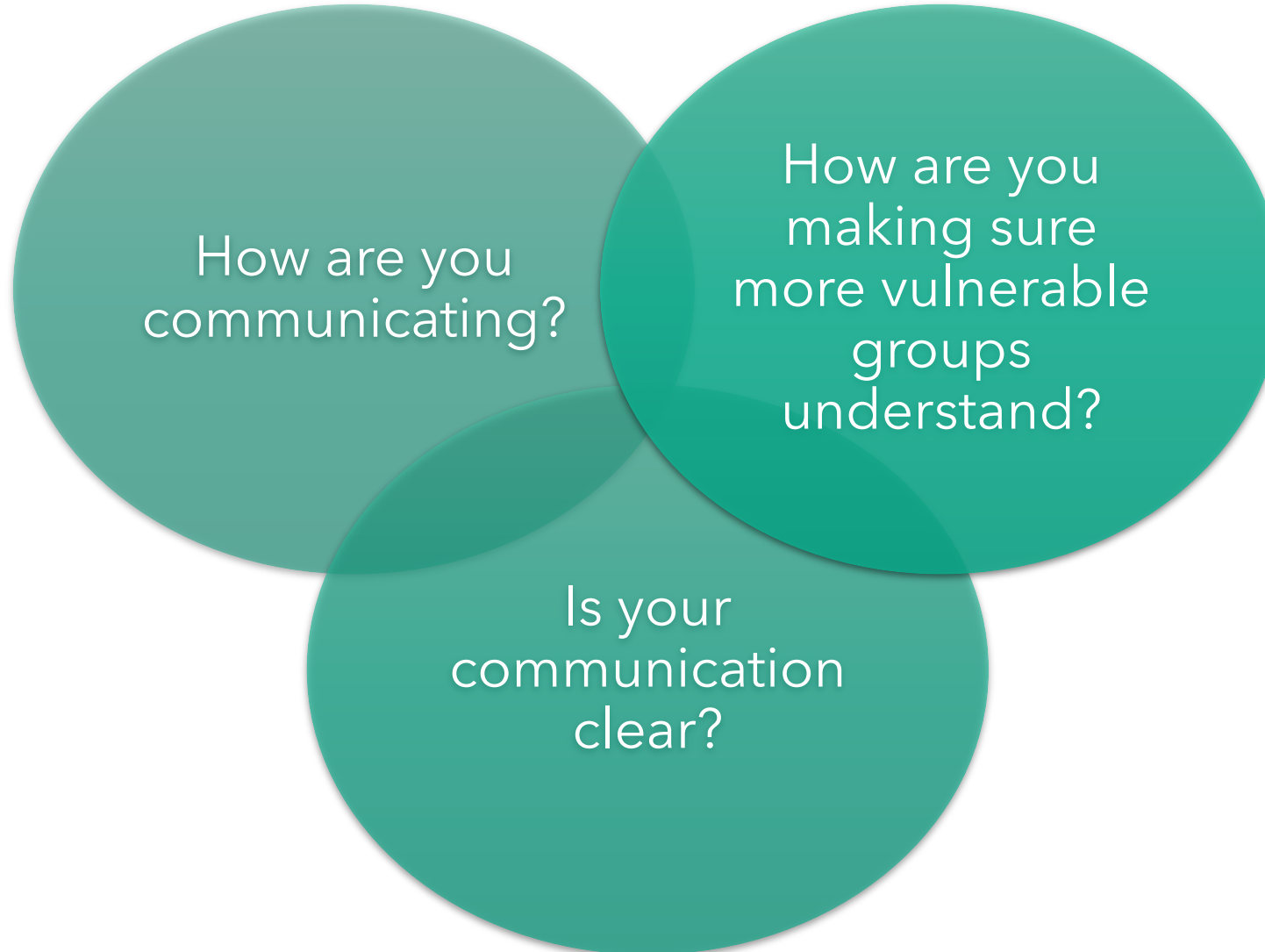
Communicating your purpose



Do people understand your purpose – especially people whom the data is about or who are impacted by its use?



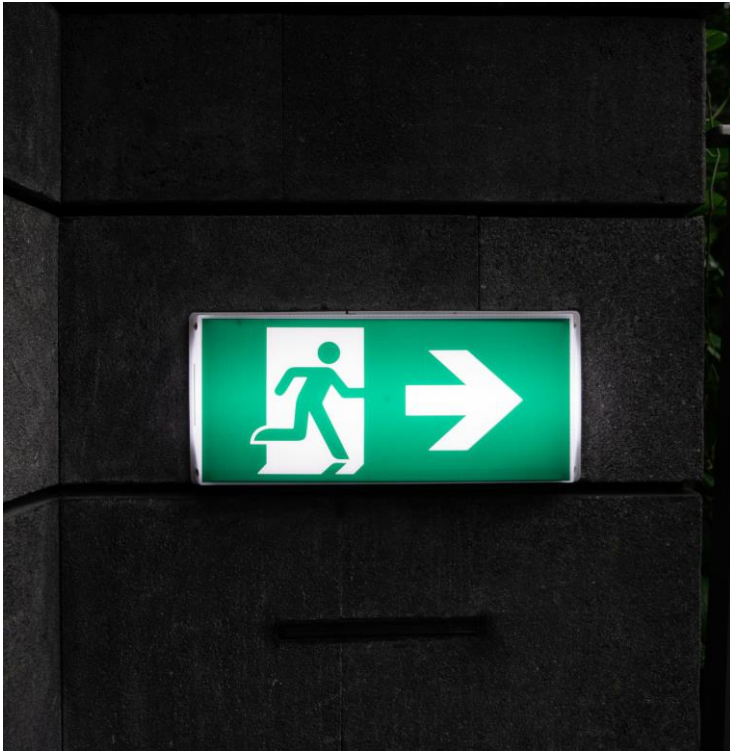
Image: <https://theodi.github.io/interactive-data-ethics-canvas/>



Communicating purpose:

- How will you ensure more vulnerable individuals or groups understand the project and risks?





Allow withdrawal of consent

- Build all systems that collect, process, and use the data with the expectation that data subjects will want to withdraw consent.

Privacy Principle

- Individual participation.

Pros

- Increase trust.
- Carry out withdrawal as a normal and quick business process.

Cons

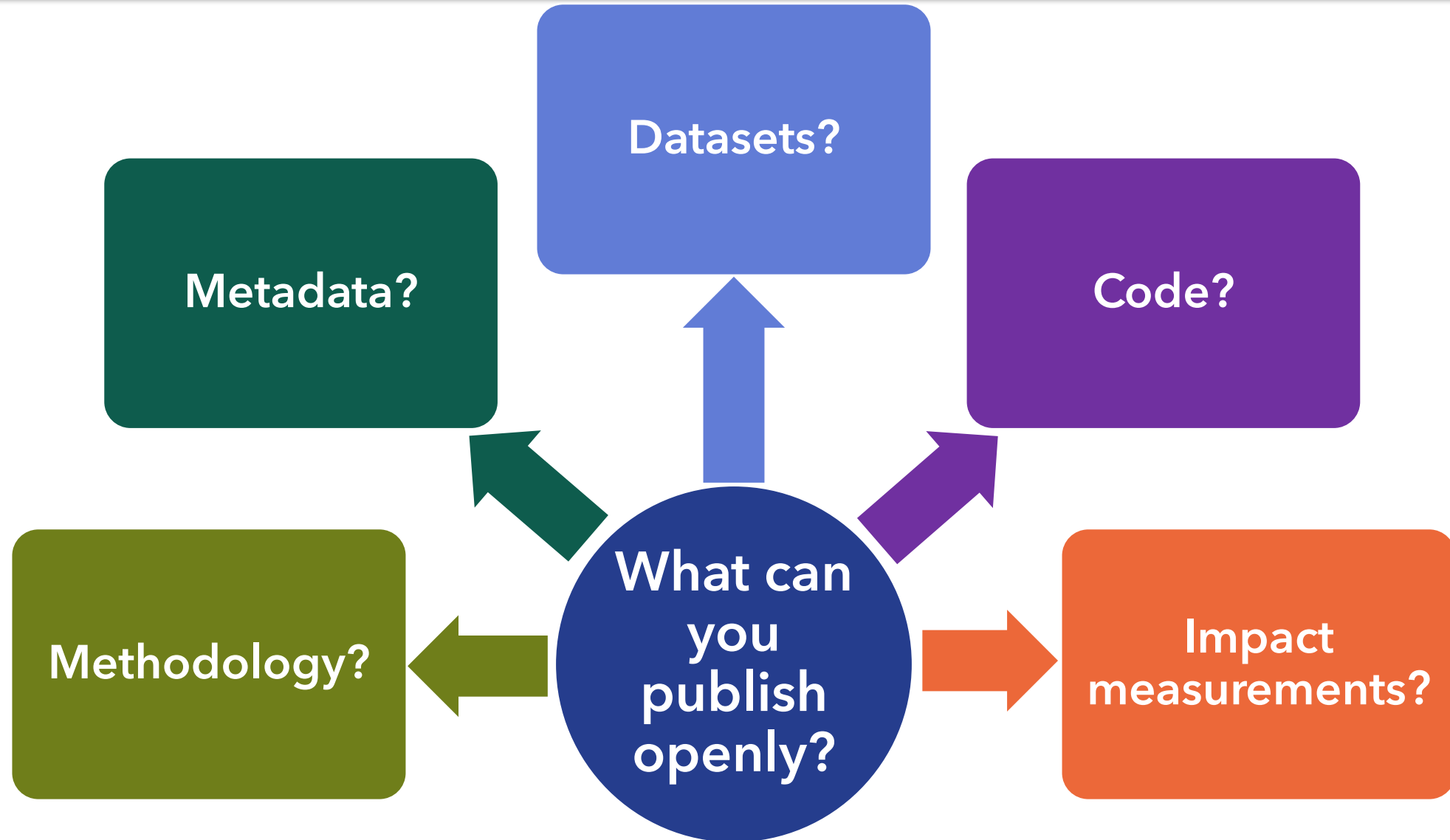
- None.

Openness and transparency



How open can you be about this project? Could you publish your methodology, metadata, datasets, code or impact measurements?





Sharing data with others



Are you going to be sharing data with other organisations?
If so, who?



Image: <https://theodi.github.io/interactive-data-ethics-canvas/>

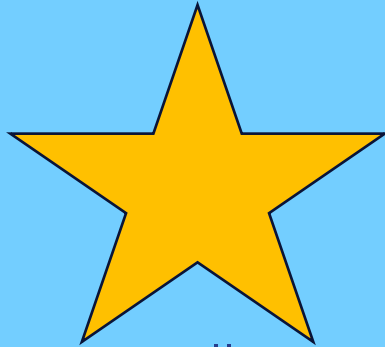
Will you share data with other organizations?

- With whom?

Will you publish any of the data?

- Under what conditions?

Your actions



- What actions will you take before moving forward with this project? Which should take priority?



As you worked through the Data Ethics Canvas:

What actions did you decide to take?

What decisions did you make?

Record them.

Follow through.

Questions? Input? Contact me.

Cathi Greenwood

Washington State Open Data
Program Manager

opendata@wa.gov



**Your state.
Your data.
data.wa.gov**