



Considering fairness in a broader sense

Human-machine interactions based on big data and AI are changing administration, supervision and enforcement

Useful developments, but using 'unfamiliar' technologies in government may impact:

- citizens
- companies (both as data source and as co-developer)
- civil society (unions/ngo's)
- governments themselves

What do 'lay' citizens consider as fair and/or trustworthy?

Can we convince citizens on the project level that our AI project is fair and trustworthy?

Designing trustworthy and fair big data and AI-based supervision

Different approaches, but EU HLEG on AI provides starting point

Trustworthy AI should be:

(1) lawful - respecting all applicable laws and regulations

(2) ethical - respecting ethical principles and values (includes fairness)

(3) robust - both from a technical perspective while taking into account its social environment



Designing trustworthy and fair big data and AI-based supervision

Ethics and robustness are subdivided in 7 principles

1. Human agency and oversight
- 2. Technical Robustness and safety (bias reduction)**
3. Privacy and data governance
4. Transparency
- 5. Diversity, non-discrimination and fairness (preventing bias)**
- 6. Societal and environmental well-being (preventing harm)**
7. Accountability



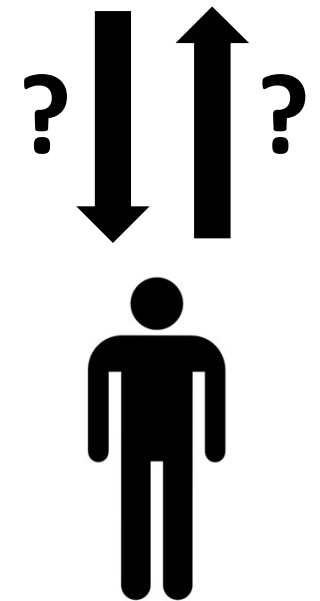
What we noticed in interviews

Potential disconnects:

Governmental activities on project-level

NGO scrutiny either on macro-level (tendency to group governments) or on salient examples

Information asymmetries



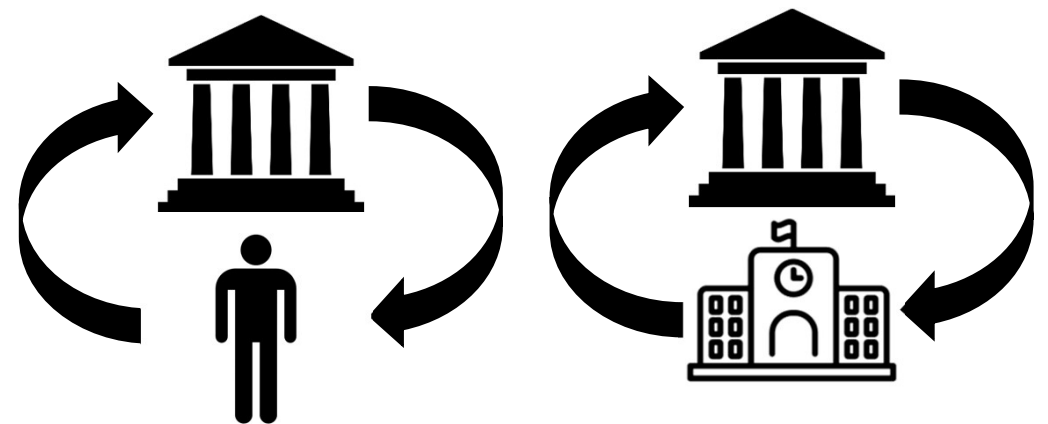
A disconnect between governmental action and citizen perceptions?

Reputations and audiences (companies, 'lay citizens', vulnerable groups)

IT literacy (knowledge about data, AI and its applications and limits)

Pre-existing levels of trust

Privacy concern



How to test this?

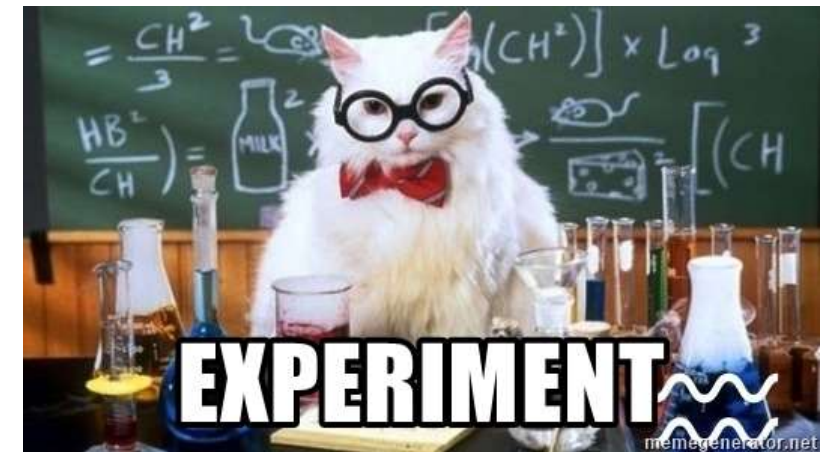
Experiment!

Experiment 1	Experiment 2
Information on legal compliance	Information on human-in-the-loop
Information on no-harm and explainability	Information on fairness & non-discrimination
Information on data-gathering	Information on technical robustness

Experiment 1: 12 groups

Experiment 2: 8 groups

Data coming in, so these are preliminary results



Two groups from experiment 1 as examples

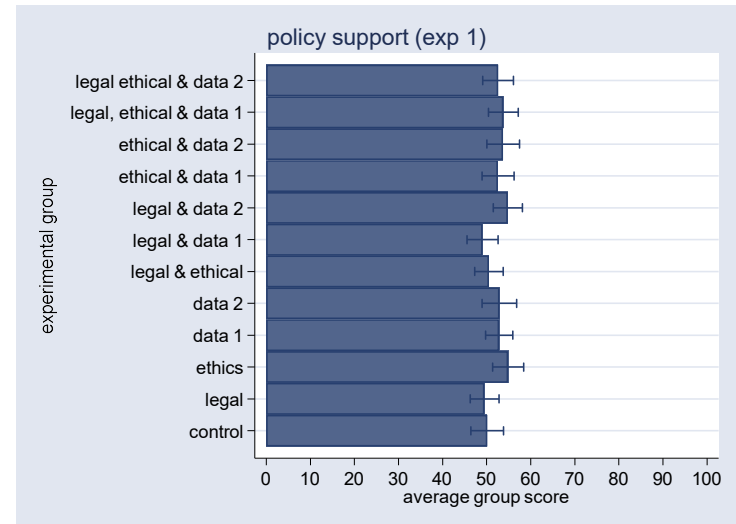
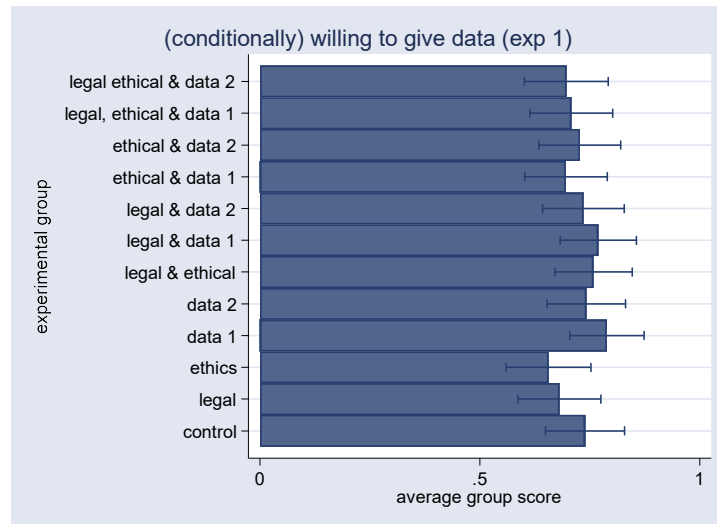
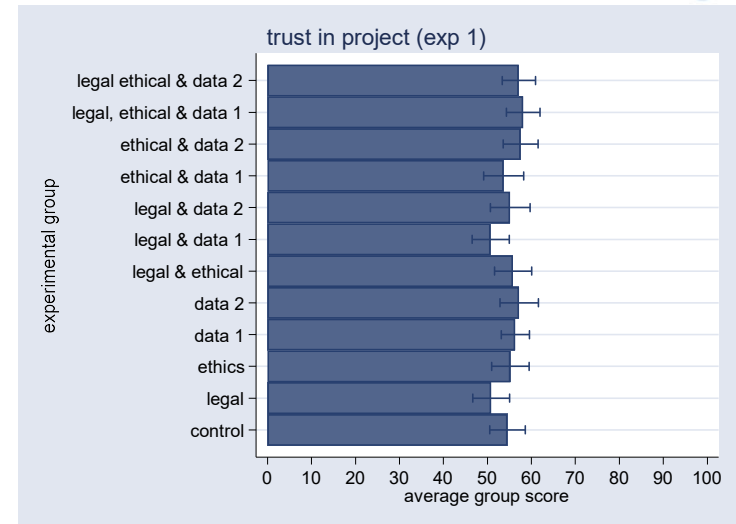
Control group:

Experimental group (legal & ethical)

<p>Artificiële intelligentie (AI) in federale overheidsprojecten: Overheidsorganisaties werken steeds meer digitaal. Daarom zet de federale overheid tegenwoordig in op artificiële intelligentie (AI). Het gaat om meerdere projecten, waaronder:</p> <ul style="list-style-type: none"> • Een gezamenlijk project met Waalse en Vlaamse overheden om met artificiële intelligentie schade aan wegen te herkennen. Doordat de computer op foto's putten in wegen herkent kan onderhoud efficiënter georganiseerd worden. • Het controleren van belastingaangiftes. Met diverse gegevens wordt de kans voorspeld dat iemand fraude heeft gepleegd in zijn of haar aangifte. Door te focussen op aangiftes met een hoge kans op fraude kunnen inspecteurs sneller misstanden opsporen. • Het volgen van stromen van mensen via hun GSM om tijdens evenementen te voorspellen waar nooddiensten (zoals ambulances) nodig zijn. Hierdoor kunnen nooddiensten beter inspelen op snel veranderende situaties. <p>Onafhankelijke experts stellen echter vragen bij de privacy en databeveiliging. Ook is artificiële intelligentie dermate complex dat het niet altijd duidelijk is hoe een computerprogramma tot een bepaalde conclusie komt.</p>	<h3>Baseline</h3>	<p>Artificiële intelligentie (AI) in federale overheidsprojecten: Overheidsorganisaties werken steeds meer digitaal. Daarom zet de federale overheid tegenwoordig in op artificiële intelligentie (AI). Het gaat om meerdere projecten, waaronder:</p> <ul style="list-style-type: none"> • Een gezamenlijk project met Waalse en Vlaamse overheden om met artificiële intelligentie schade aan wegen te herkennen. Doordat de computer op foto's putten in wegen herkent kan onderhoud efficiënter georganiseerd worden. • Het controleren van belastingaangiftes. Met diverse gegevens wordt de kans voorspeld dat iemand fraude heeft gepleegd in zijn of haar aangifte. Door te focussen op aangiftes met een hoge kans op fraude kunnen inspecteurs sneller misstanden opsporen. • Het volgen van stromen van mensen via hun GSM om tijdens evenementen te voorspellen waar nooddiensten (zoals ambulances) nodig zijn. Hierdoor kunnen nooddiensten beter inspelen op snel veranderende situaties. <p>Onafhankelijke experts stellen echter vragen bij de privacy en databeveiliging. Ook is artificiële intelligentie dermate complex dat het niet altijd duidelijk is hoe een computerprogramma tot een bepaalde conclusie komt.</p>
<h3>Legal</h3>	<p>De federale overheid erkent dat er juridische kanttekeningen zijn. Daartoe heeft de federale overheid verschillende onafhankelijke datajuristen in dienst genomen die de projecten zullen controleren. Ook wordt gezorgd voor een wettelijke basis die bepaalt waarvoor overheidsorganisaties artificiële intelligentie wel of niet mogen gebruiken.</p>	
<h3>Ethical</h3>	<p>Om te zorgen dat artificiële intelligentie burgers ten goede komt, heeft de federale overheid intern afspraken gemaakt. In deze afspraken staat dat:</p> <ul style="list-style-type: none"> • Overheidsorganisaties altijd moeten zorgen dat de beslissingen van hun artificiële intelligentie compleet uitlegbaar zijn; • Artificiële intelligentie altijd in het belang van burgers ingezet moet worden. Zo mag de federale overheid artificiële intelligentie niet inzetten voor het opsporen van kleine fouten van burgers, maar wel voor ernstige fraudezaken. 	

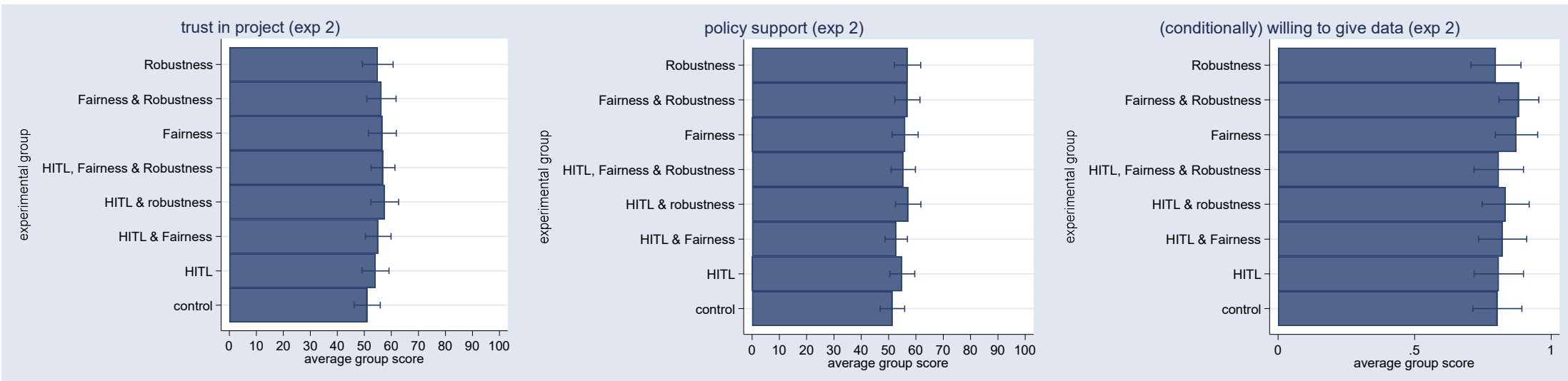
Results experiment 1

- Little differences between experimental groups
- No significant differences from control group
- Some (limited and inconsistent) significant differences between intervention groups
- Information on fairness, legal compliance and data-gathering has little influence



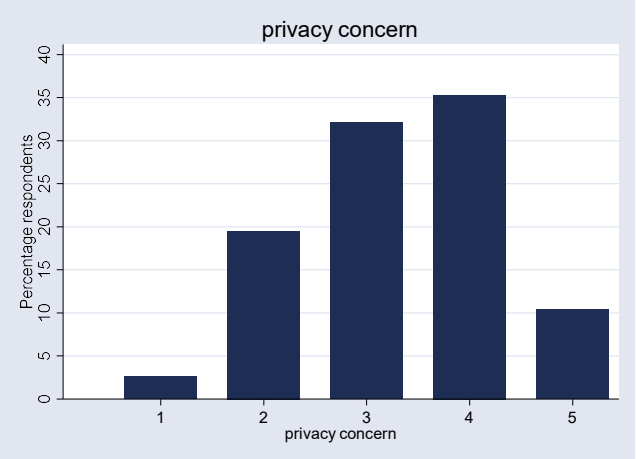
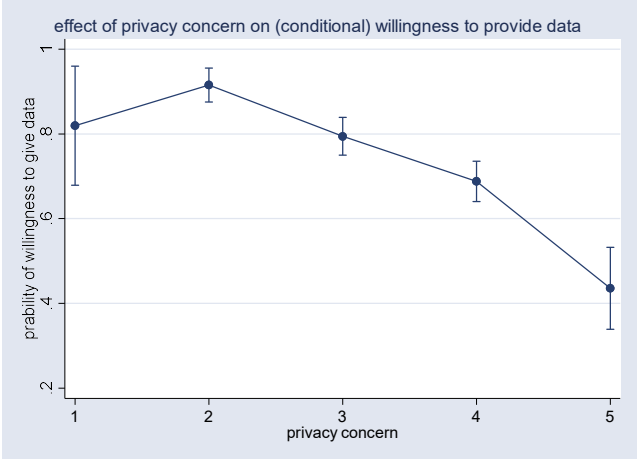
Results experiment 2

- Very similar, also with regard to fairness & non-discrimination as an intervention



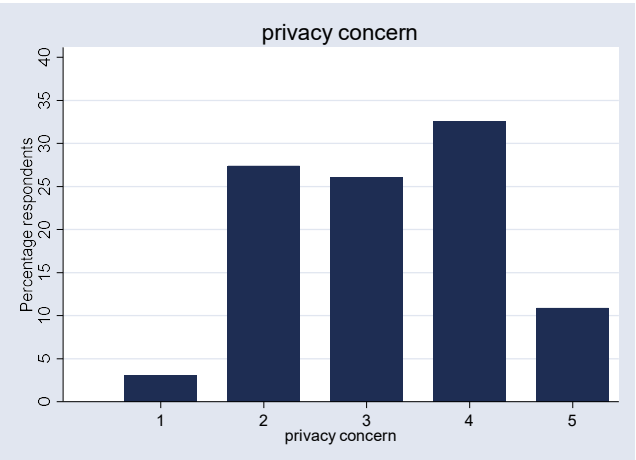
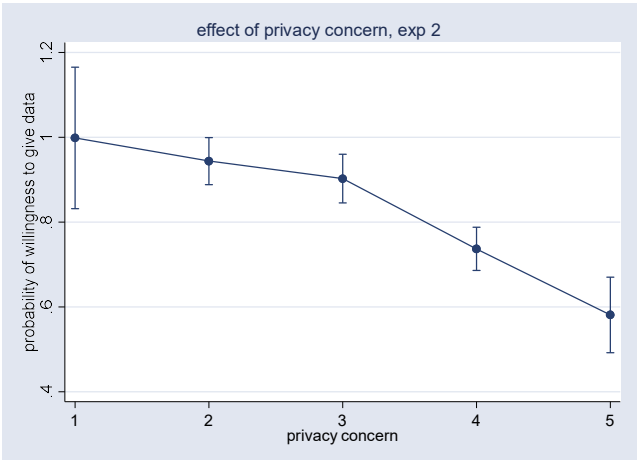
Survey variable results

Exp. 1



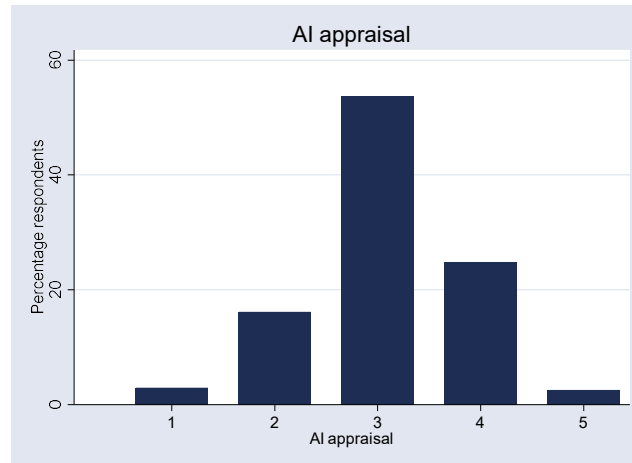
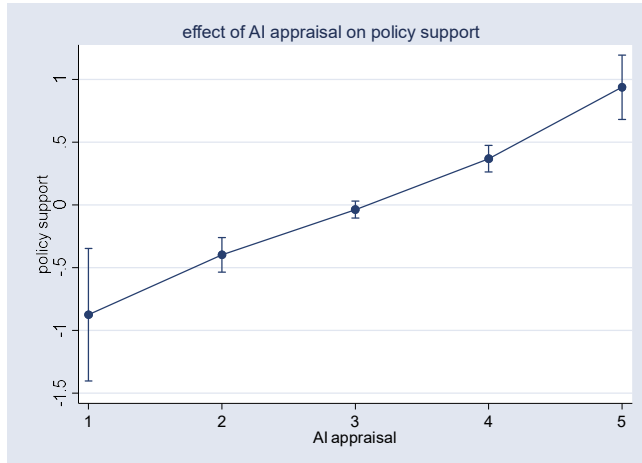
Privacy concern, experiments 1 and 2

Exp. 2



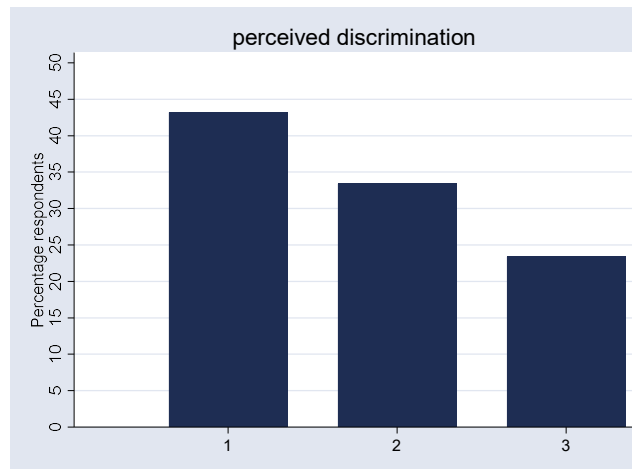
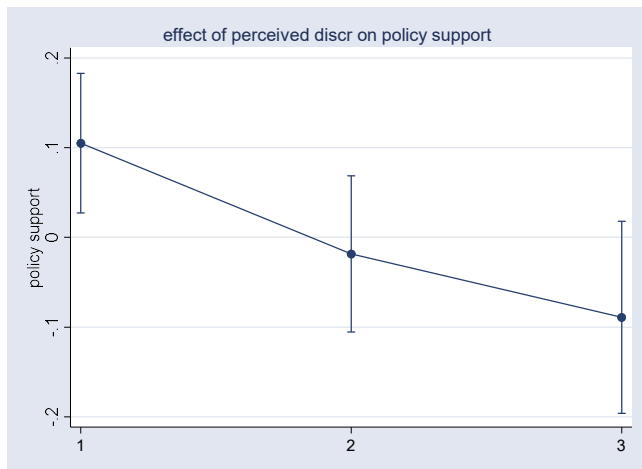
Preliminary conclusion: prior attitudes seem to explain more than experimental interventions

Survey variable results



Some less certain results (we will need to verify these)

Experiment 1. AI appraisal



Experiment 2. Membership of group that is discriminated against (self-reported)



In conclusion

Fairness and trustworthiness are good aims from a good governance viewpoint

However, on the project- and organizational level our research has the following implications:

- do not expect wonders of AI/project design in terms of citizen trust/support
- prior attitudes seem more important than information on measures
- fishbowl transparency on fairness has its limits due to bounded rationality of citizens
- many citizens are wary of how, where, when and for what purpose their data is used
- these attitudes have real implications for governance