Ethical frameworks and governance

Kate Powis and Libby Bishop
Research Data Management Team
UK Data Service – University of Essex

Supporting civil society organisations to deliver insights and impact from data: an ESRC and UK Data Service workshop
17 October 2017
Outline

• Introduction
• Questions arising from case studies
• Your story
• Exploring your stories – guided group work
Your ethical issues

1. What is an “ethical issue”?
2. Whose responsibility is ethics?

3. Dilemmas – benefits of data sharing, but real risks (small to severe) of harm to stakeholders or CSO
   • Improve services, meet reporting requirements, but risks: loss of reputation, loss of trust, breach of confidentiality, other harms
   • Tensions: vocational and professional approaches “objective” research/evaluation vs. data for advocacy
   • Uses of representations: can be inaccurate, exploitative
   • Question of consent – how is it understood?
Questions to guide group work

• What are “ethical issues” in data?
• Whose responsibility are they in your organisation?
• Can the documents provided help you think through your approaches to ethical issues?
• What are your takeaway points – what can you do next?
4. A set of morally reasonable expectations about the governance and use of data should be determined in accordance with four principles:

- the principle of respect for persons
- the principle of respect for established human rights
- the principle of participation of those with morally relevant interests
- the principle of accounting for decisions
Guidelines for good practice:
UK Cabinet Office – Data Science Ethical Framework

Six key principles: at a glance view

1. **Start with clear user need and public benefit**
   - Data science offers huge opportunities to create evidence for policymaking, and make quicker and more accurate operational decisions. Being clear about the public benefit will help you justify the sensitivity of the data (principle 2) and the method that you want to use (principle 3).

2. **Use data and tools which have the minimum intrusion necessary**
   - You should always use the minimum data necessary to achieve the public benefit. Sometimes you will need to use sensitive personal data. There are steps that you can take to safeguard people’s privacy e.g. de-identifying or aggregating data to higher levels, querying against datasets or using synthetic data.

3. **Create robust data science models**
   - Good machine learning models can analyse far larger amounts of data far more quickly and accurately than traditional methods. Think through the quality and representativeness of the data, flag if algorithms are using protected characteristics (e.g. ethnicity) to make decisions, and think through unintended consequences. Complex decisions may well need the wider knowledge of policy or operational experts.

4. **Be alert to public perceptions**
   - The Data Protection Act requires you to have an understanding of how people would reasonably expect their personal data to be used. You need to be aware of shifting public perceptions. Social media data, commercial data and data scraped from the web allow us to understand more about the world, but come with different terms and conditions and levels of consent.

5. **Be as open and accountable as possible**
   - Being open allows us to talk about the public benefit of data science. Be as open as you can about the tools, data and algorithms (unless doing so would jeopardise the aim, e.g. fraud), provide explanations in plain English and give people recourse to decisions which they think are incorrectly made. Make sure your project has oversight and accountability built in throughout.

6. **Keep data secure**
   - We know that the public are justifiably concerned about their data being lost or stolen. Government has a statutory duty to protect the public’s data and as such it is vital that appropriate security measures are in place.

More detail in annex below
Where are you? What do you do next?

Quick checklist

1. Start with clear user need and public benefit
   - How does the department and public benefit?
     - High public benefit (to society or to an individual)
     - Medium public benefit (to society or to an individual)
     - Low public benefit (to society or to an individual)

2. Use data and tools which have the minimum intrusion necessary
   - How intrusive and identifiable is the data you are working with?
     - Non-personal and therefore non-identifiable
     - Personal but non-sensitive
     - Personal, sensitive data which could be inferred or directly re-identified
   - If identifying individuals, how widely are you searching personal data?
     - Querying against known individuals
     - Querying against a targeted group
     - Speculatively searching for needle in haystack

3. Create robust data science models
   - What is the quality of the data?
     - Representative and unbiased
     - Historical data which is biased and excludes certain groups
     - Inaccurate or missing data
   - How automated are the decisions?
     - Humans making decisions based on analysis
     - Limited human oversight but regularly checked
     - No human oversight or method of checking
   - What is the risk that someone will suffer a negative unintended consequence as a result of the project?
     - Low
     - Medium
     - High

4. Be alert to public perceptions
   - If personal data for operational purposes, how compatible was it with the reason collected?
   - Do the public agree with what you are doing?
     - Very compatible
     - Less compatible but fair
     - Not compatible

5. Be as open and accountable as possible
   - How open can you be about the project?
     - Very open, and make open the tools and data for re-use
     - Open about project but not about data tools
     - Cannot talk about project aim
   - How much oversight and accountability is there throughout the project?
     - Throughout – including the decision made as a result of insight
     - Only at the beginning
     - None

6. Keep data secure
   - How secure is your data?
     - Very secure, with restricted access to a few named individuals
     - Secure and password protected
     - Openly available within the department

*Not all may apply to your project

Tick where you are on the scale

- - - - -

All fine? Go forward!  Some issues? Think carefully  Tricky issues? Extreme care & oversight

Some departments might find themselves at the left hand side of the scale, and others more on the right (blue), reflecting the nature of their department’s work. This does not mean the project should not go ahead, but think carefully about it, and if possible, bring some elements to the green end of the scale.
Best practice for legal compliance

The public benefit of data access should be greater than the harm or distress caused by disclosure

✓ Investigate early which laws apply to data
✓ Do not collect/store personal or sensitive data if not essential to your research/evaluation project
✓ Seek advice from local research office experts
✓ Plan early in project
✓ If need to deal with personal or sensitive data:
   ✓ inform participants about how their data will be used
✓ Not all research/evaluation data are personal (e.g. if effectively anonymised)
More tools, resources and guidance

- Anonymisation Decision Making Framework
  https://www.youtube.com/watch?v=elleJgLMMoA
- Information Commissioner’s Office
  https://ico.org.uk/
- Markkula Center for Applied Ethics, University of Santa Clara
  https://www.scu.edu/ethics/
- UK Data Service – general guidance on legal and ethical issues, including anonymisation (quantitative and qualitative) and consent
  https://www.ukdataservice.ac.uk/manage-data/legal-ethical
  https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation
  https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing
Guidelines and tools for internet based data collection

• British Sociological Association (BSA) Ethics Guidelines and Collated Resources for Digital Research
  https://www.britsoc.co.uk/media/24309/bsa_statement_of_ethical_practice_annexe.pdf

• Recommendations from the AoIR Ethics Working Committee (Version 2.0) Ethical Decision-Making and Internet Research
  https://aoir.org/reports/ethics.pdf
Contact

Collections Development and Data Publishing team
UK Data Service
University of Essex
ukdataservice.ac.uk/help/get-in-touch.aspx