Protecting confidentiality: the ‘5 Safes’

Five guiding principles:

- **Safe people** - educate researchers to use data safely
- **Safe projects** – consider purposes of all data reuse
- **Safe settings** – locations where data used-SecureLab
- **Safe outputs** - SecureLab projects outputs screened
- **Safe data** - treat the data to protect respondent confidentiality

- Anonymisation addresses **Safe data**
- It is an essential and valuable tool – not the whole story
UK Data Service & anonymisation

• We now do very little anonymisation

• Excellent practices used by key depositors, ONS, NatCen (i.e., in-house micro-data release panels)

• More challenges for single researchers, early career
  • Anonymising after formal end of project, did not budget, research assistants have moved on…
  • Prefer early intervention-Research Data Management Training

We consult with depositors to agree the appropriate degree of anonymisation in conjunction with the level of access control best suited for their data
Access summary – three tiers

**Open**
- available for download/online access under open licence without any registration

**Safeguarded**
- available for download/online access to users who have registered and agreed to an End User or Special Licence

**Controlled**
- available for remote or SecureLab access to authorised authenticated users and outputs checked
Checks for accuracy and completeness

- Typical checks – Safeguarded data
  - Review DIRECT identifiers - names, ID numbers
  - Must also check INDIRECT identifiers
  - Birth date; education; employment; geography (combinations)

- Typical checks – Controlled data
  - Permissible to retain some identifying information (because access very tightly controlled)

- More anonymisation not necessarily better
  - Children in longitudinal studies – month of birth

- Ex. Coding geographic variable by access level:
  - Open data – country
  - Safeguarded – government office region
  - Controlled – post code
Anonymising research data: summary

- **Plan** before collecting data
- **Check** and remove direct identifiers and treat indirect identifiers to reduce disclosure risk
- **Document** your changes

**Balance anonymisation with access control to preserve data usability**
Anonymisation – tools and guides

Existing and emerging tools:

- Statistical disclosure control software e.g., Mu-argus, ARX
- Tools for qualitative data
  - [http://data-archive.ac.uk/curate/standards-tools/tools](http://data-archive.ac.uk/curate/standards-tools/tools)

Published guides:

- ONS Disclosure control guidance for microdata produced from social surveys [http://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/policyforsocialsurveymicrodata](http://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/policyforsocialsurveymicrodata)
UK Data Service guidance on anonymisation

- UK Data Service: [www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation](http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation)

- *Managing and Sharing Research Data* book
Questions

Libby Bishop

ebishop@essex.ac.uk