DIY Research Data Management Training Kit for Librarians

Ethics and copyright

THE UNIVERSITY of EDINBURGH

Robin Rice, Data Librarian
EDINA and Data Library
Information Services
Learning objectives

• Be aware of ethical requirements that apply to the collection and management of data involving human subjects.
• Understand the difference between privacy and confidentiality and how they apply to the management of research data.
• Know what IPR is and how it applies to your [users’] research data.
Ethics of collecting data on human subjects

• The purpose and nature of the research itself
• The nature of consent obtained (e.g. opt-in versus opt-out participation)
• What data needs to be safeguarded during analysis and destroyed after its use
Which research is subject to ethical review?

- Using human participants or live animals
- Referencing individual subjects (people)
- Keeping identifiers for individuals
EXEMPT from ethics review

• freely available in the public domain, or
• obtained through pure observational studies of public behaviour:
  – are of human action that occurs in a forum open to the general public
  – are non-invasive
  – require no interaction with participants
  – do not identify participants
Some scholarly societies with ethics guidelines

- Association of Social Anthropologists of the UK and the Commonwealth
- British Association of Social Workers
- British Psychological Society
- British Educational Research Association
- British Society of Criminology
- British Sociological Association
- Social Research Association
• Individuals are normally considered to have a ‘right to privacy’ - the state of being free from intrusion or disturbance in one's private life or affairs.

• Research proposals need to indicate how data will be collected about human subjects. Ethical reviews will consider how intrusive this is for the subjects.

• The right to privacy implies the need to obtain consent from research subjects.
Confidentiality and research

• Oftentimes the consent form will include a promise to the subject that the researcher will not disclose confidential information.

• Keeping this agreement implies controls on storage, handling, and sharing of personal data.

• Strategies can include
  – If possible, collect the necessary data without using personally identifying information.
  – If personally identifying information is required, de-identify your data upon collection or as soon as possible thereafter.
  – Avoid transmitting unencrypted personal data electronically.
Intellectual Property Rights

- Intellectual Property Rights (IPR) can be defined as *rights acquired over any work created or invented with the intellectual effort of an individual*. These rights are time-limited and do not require explicit assertion.

- Owners of works are granted exclusive rights:
  - to publish the work
  - to license the work’s distribution to others
  - to sue in case of unlawful or deceptive copying

- It may be possible to waive these rights through a public declaration similar to a licence.
When do data enjoy copyright in the UK?

• When the dataset is an expressive work
  – Even if not, the metadata may enjoy copyright

• What if the data are merely a set of facts?
  – A simple list or spreadsheet might not enjoy copyright
  – The EU established a ‘database right’ (‘sui generis’ right) to protect *substantial investment* made by database producers in obtaining, verifying and presenting database contents.
  – Also, the structure of the database, including the *selection and arrangement* of the database’s contents may meet a test of intellectual creativity bestowing copyright.
Better not count on it though…

- What does “in the public domain” mean? Not everyone on the internet understands copyright is implicit.
- “Insubstantial amounts” of databases may be copied, but there is not a lot of case law to determine how much that is: under 50%?
- “Screen scraping” is a common method of obtaining raw data from online sources. This may be lawful, depending on what gets published.
To be continued…

- Data licensing to be covered in next session
- An advantage of data licensing is ability to be explicit about what can and cannot be done with the data obtained, compared with copyright.
- However, if a licence is breached, any third party who obtains the data is not under its obligation!
- A time-limited embargo in a repository may help data creators to gain time to publish results first.
Last word

• Best to create a public use dataset that you are comfortable about getting used, if putting online.
• And don’t forget the Freedom of Information Act and related legislation may compel you to release data, especially if in the public interest!