Legal and ethical issues of combining survey data with new forms of data

Andrew Charlesworth

Centre for IT & Law

University of Bristol Law School

What are new forms of data

- Old forms of data digitized
- Data 'born digital'
- Government transactions
- Government and other registration records
- Commercial transactions
- Internet usage
- Tracking data
- Satellite and aerial imagery

Legal issues

- Data protection, privacy, confidentiality
  - E.g. GDPR/DPA 2018, protection of private information

- Records-specific rules
  - E.g. Re-use of Public Sector Information Regulations 2015, Access to Health Records Act 1990

- Intellectual property
  - E.g. copyright

- Other ownership/contractual claims
  - E.g. Twitter Terms of Use

Principle	Automated analysis of datasets
Collection Limitation	The larger the data collection, the better the potential for identifying interesting correlations.
Data Quality	'Messy data' is fine, it's not clear what is relevant until its analysed, and even inaccurate or incomplete data can be useful.
Purpose Specification	Data may have been collected for a particular purpose, but analysis may indicate further unrelated and previously unknown, but valuable, purposes. Data as collected may not be obviously PII, but analysis of it may identify individuals.
Use Limitation	There may be value in sharing and aggregating data that may not be apparent at the time of collection
Security Safeguards	It may be unclear what security issues if any arise from a particular collection of data, or its analysis.
Openness	Where data is collected and analysed, it may not be obvious that it is PII, and even in circumstances where it is, the researcher may have no way of informing the data subject of its use.
Individual Participation	Data which is anonymous may still be utilised in ways that can cause risk/harm to an individual.
Accountability	How and when might a researcher to be held accountable and for what?

Big vs. Data Privacy

Ethics

- Ethical standards in research seek to do three things.
  - to protect the rights and legitimate interests of research participants (e.g. by normally requiring informed consent).
  - to create trust in both researchers and research outcomes (e.g. by endorsing or rejecting practices affecting the actual or perceived reliability or trustworthiness of research data).
  - to preserve the research environment (e.g. by seeking to prevent practices which might restrict future access to research participants or reduce the effectiveness of future research).

Real Life

- A postgraduate researcher presenting a paper at a conference describes collecting and mining the conversations on Islamic chatrooms seeking words that might indicate extremist conversations, without either informing or gaining the consent of the participants, on that grounds that the conversations are 'in the public domain'.

- A professor in Computer Science, while discussing new data mining techniques within vast datasets of interaction data, e.g. internet search terms, pronounces himself amazed at the leeway that the law permits him when analysing data relating to individuals, or their activities, in ways that he himself would consider to be 'intrusive'.

Case Study I

- Population Health
  - Study of antibiotic use in livestock – implications for human health
- Study uses traditional survey methods
  - Questionnaires + interviews/focus groups
- Study also uses various datasets from veterinary practices, agricultural suppliers, government datasets and retailers
  - Data protection issues
  - Confidentiality issues
  - Commercial confidentiality

Case Study II

- Potential risks to study participants (farmers)
  - Breach of personal data
  - Disclosure of commercial information
  - Inconsistency between reporting outcomes and apparent use of antibiotics
- Major concerns on the part of information suppliers about GDPR
  - General lack of knowledge of the law
  - Misapprehensions about the need for consent/re-consent for GDPR and ethics purposes
- Potentially addressed via use of Privacy Preserving Record Linkage:
  - Vatsalan et al (2013) A taxonomy of privacy-preserving record linkage techniques *Information Systems* 38(6): 946-969.

Pragmatics

- The EU is keenly aware of the potential impact of the GDPR on research, both public & private.
  - Existing framework for UK research compliance remains broadly the same.
  - BUT there is scope for divergence in exemptions and derogations between EU Member States, with implications for cross-border research collaborations.
- Certain areas of research are likely to see greater scrutiny than others - big data – Facebook/Cambridge Analytics

The Nightmare Scenario

- "Inquiry into sex abuse files blunder"
  - http://news.bbc.co.uk/1/hi/uk/1519889.stm
- Fortunately, for the University and researchers concerned, this was 18 years ago.
- Could it happen today?
  - Research ethics review process
  - Research methods and ethics training
  - Data management plans
  - Institutional awareness of 'human research'
  - Centralised data storage and archiving
  - Secure data transfer / authorised 3rd parties
  - Disposal of equipment policy

Pragmatics II

- As a social sciences researcher and REC member, my key issues are:
  - Ensuring training and education of researchers;
  - Building DP (and FOI) into data management plans;
  - Awareness of research under my remit involving personal data;
  - Input into, and awareness of, institutional policy;
  - Accountability and research governance;
    - Documentation, audit trail, responsibilities, outcomes.
  - Understanding the sometimes fuzzy line between 'ethical research' and 'lawful research'
    - Consent as both a legal and ethical requirement, and the alternatives to consent.

Difficulties

- Administration vs Academics – two-way communication/co-operation
- Management perceptions of risk and value - REC seniority/workload
- 'Cut and paste' ethics applications
- Differences in discipline cultures
- Differences in scale
- External projects
- Expertise gap in RECs

Pragmatics III

- Different academic disciplines face varying challenges to existing practices - avoid 'one-size-fits-all' solutions and 'quick fixes'.
  - RECs already address DP (to varying degrees) – they can identify discipline-specific risks, good practice and problems with suggested 'solutions'
  - As RECs have varying expertise in DP, it is important that they are not simply used as 'gatekeepers', but rather trained/developed as 'facilitators'.
- Research training, research data management plans, workflow
  - 'privacy by design' and 'privacy impact assessment' elements
- Accountability
  - We do forms and training, but do we really do accountability?