INF5220 –
Qualitative research methods

Privacy (law) and ethics in research

Gisle Hannemyr, Ifi

Overview of lecture

- Regulatory framework for processing personal data.
- Ethics and processing personal data.

Regulatory framework for collecting personal data for research

- Because research may involve use of personal data, there may be legal and ethical guidelines that regulates IR data processing:
  - EU privacy directive
  - Norwegian personal data act (popplyl.)
  - European Convention on Human Rights
  - UN Declaration of Human Rights
  - The Nuremberg Code
  - The Belmont Report
  - The Declaration of Helsinki

Legal requirements in Norway

- The legal requirements for the controller doing research where *personal data* are collected and processed are specified in *Personopplysningsloven* (popplyl.):
  - Main requirement: *All* such research need to be reported on a special form to *Personvernombudet for forskning* (Privacy ombudsman for research). http://www.nsd.uib.no/personvernombud/
- My guidelines about filing a report (in Norwegian):
  - http://heim.ifi.uio.no/~gisle/ifi/pol.html

Norwegian personal data act §2: Definitions

- **personal data** (*personopplysning*): any information and assessments that may be linked to a natural person, (§ 2.1)
- **sensitive personal data** (sensitive personopplysninger) – personal data related to: (§ 2.8):
  a) racial or ethnic origin, or political opinions, philosophical or religious beliefs,
  b) the fact that a person has been suspected of, charged with, indicted for or convicted of a criminal act,
  c) health,
  d) sex life,
  e) trade-union membership.

«Personopplysning» = Personal data

- popplyl: Data that may *directly or indirectly* connected to a physical persion
  - Name
  - PIN
  - IP-address
  - Patient profile of a rare disease + location (mosaic effect - *bakveisidentifisering*)

Norwegian personal data act §2: Definitions

- **processing of personal data** (*behandling av personopplysninger*): any use of personal data, such as collection, recording, alignment, storage and disclosure or a combination of such uses, (§ 2.2)
- **personal data filing system** (*personregister*): filing systems, records, etc. where personal data is systematically stored so that information concerning a natural person may be retrieved, ($2.3)
- **consent** (*samtykke*): any freely given, specific and informed declaration by the data subject to the effect that he or she agrees to the processing of personal data relating to him or her. (§ 2.7)

Actors in Norwegian legal framework

- **controller** (*behandlingsansvarlig*): the person who determines the purpose of the processing of personal data and which means are to be used (popplyl. §2.4).
- **processor** (*databehandler*): the person who processes personal data on behalf of the controller (popplyl. § 2.5).
- **data subject** (*registrerte*): the person to whom personal data may be linked (popplyl. §2.6).

When does the law apply?

§ 11: **Basic requirements for the processing of personal data**
The *controller* shall ensure that personal data which are processed

a) are processed only when this is authorized pursuant to §§ 8 and 9,
b) are *used only for explicitly stated purposes* that are objectively justified by the activities of the controller,
c) are not *used subsequently* for purposes that are incompatible with the original purpose of the collection, without the consent of the data subject,
d) are *adequate, relevant and not excessive* in relation to the purpose of the processing, and
e) are accurate and up-to-date, and are *not stored longer than is necessary* for the purpose of the processing

Conditions for the processing of personal data

- § 8: Personal data (cf. § 2, no. 1) may only be processed if the data subject has consented thereto, or …
- § 9: Sensitive personal data (cf. § 2, no.8) may only be processed if the processing satisfies one of the conditions set out in § 8 and
  a) the data subject consents to the processing,
  …

Consent must be:

- Freely given:
  - No pressure or coercion or linking to favours
- Specific:
  - Usually by signature
- Informed:
  - Purpose of reserach
  - How personal data will be used
  - When personal data will be destroyed or anonymized.

Alternative conditions to consent or statutory authority (a-f):

popplyl. § 8: Personal data (cf. section 2, no. 1) may only be processed if the data subject has consented thereto, or there is statutory authority for such processing, or the processing is necessary in order:

a) to fulfil a contract to which the data subject is party, or to take steps at the request of the data subject prior to entering into such a contract,
b) to enable the controller to fulfil a legal obligation,
c) to protect the vital interests of the data subject,
d) to perform a task in the public interest,
e) to exercise official authority, or
f) to enable the controller or third parties to whom the data are disclosed to protect a legitimate interest, except where such interest is overridden by the interests of the data subject.

The Mosaic Effect

"The Mosaic Effect occurs when the information in an individual dataset, in isolation, may not pose a risk of identifying an individual (or threatening some other important interest such as security), but when combined with other available information, could pose such risk. Before disclosing potential PII [personally identifiable information] or other potentially sensitive information, agencies must consider other publicly available data – in any medium and from any source – to determine whether some combination of existing data and the data intended to be publicly released could allow for the identification of an individual or pose another security concern."
Source: http://project-open-data.github.io/policy-memo/

popplyl: Report form compulsory if:

- Recording or processing of information about individuals by *electronic* means.
  - NB: "electronic" ⇔ "digital".
    Analogue recording is not considered "electronic" for legal purposes.

- or -

- A manual register containing *sensitive personal data* will be created.

popplyl: Permit compulsory if:

- Sensitive personal data is recorded (popplyl. § 33).
- Sensitive personal data (popplyl. §2.8) are data that reveals information relating to:
  - racial or ethnic origin, or political opinions, philosophical or religious beliefs;
  - the fact that a person has been suspected of, charged with, indicted for, or convicted of, a criminal act;
  - health;
  - sex life;
  - trade-union membership.

popplyl.: But permit not compulsory for research if the *privacy ombudsman for research* approves:

1. First time contact to selection of respondents is based upon, either:
   - publicly available data (i.e. data that exists in the public sphere);
   - a responsible person at the insitution where the respondent is registered;
   - initiative from the respondent.
2. The responent has given *valid consent* to all parts of the research.
3. The project is terminated at the time agreed upon.
4. All material collected is destroyed or anonymized when the project is terminated.
5. The project is not joining data from external registers or data bases.

popplyl.: Reuse of data is *not* permitted without new consent

- **popplyl. § 11c**. The controller shall ensure that personal data which are processed … are not used subsequently for purposes that are incompatible with the original purpose of the collection, without the consent of the data subject.

IR ethics, sources:

- Cheltenham and Gloucester College of Higher Education: *Research Ethics: A Handbook of Principles and Procedures.*
- Association of Internet Researchers (AoIR), reports on *Ethical and Legal Aspects of Research on the Internet*
  http://aoir.org/reports/ethics.pdf (2002)
  http://aoir.org/reports/ethics2.pdf (2012)
  - Adapted from biomedical research.

Four major problems

- Is online interpersonal media (social media) in the Public or Private Sphere?
- Covert research/Informed consent
- Protecting anonymity
- Raw data

Covert research methods

- Online research poses in general a risk to individual privacy and confidentiality because of greater accessibility of information about individuals, groups, and their communications – in ways that would prevent subjects from knowing that their behaviours and communications are being observed and recorded (e.g.: a large-scale analysis of postings and exchanges in a USENET newsgroup archive, in a chat room, etc.).

Valid consent is a required for sharing personal data

[P]rivacy is considered widely as a crucial norm in ethical research […] Data arising from research should ordinarily be considered confidential and may not be shared with others without the consent of the researched.

– *Research Ethics Handbook*

Protecting anonymity

[R]esearchers must take care where the alteration of contexts may reveal the identity of data sets hitherto protected. Particular care should be taken with data that arises from covert […] research methods […].

– *Research Ethics Handbook*

Protecting raw data

- Good research practice means that the raw data (for aggregated, pesudonymized or anonymized data that is published) must be available for scrutiny and peer review upon request.
- Solution: Retain the raw data, but pseudonymize records by using numbers instead of real IDs. Make access to RAW data very restricted (encrypted and in custody of a trusted third party, analogous to safekeeping sensitive data accumulated in biomedical research).

Institutional setting

- In biomedical resarch, the institutional setting (i.e. the research clinic) usually has well developed procedures and mechanisms for handling, anonymizing and protecting personal data originating from research.
  - This is taken as given both by the researchers and also by the data subjects (i.e. the patients).
- In Internet research, no similar setting usually exists and has to be constructed by the data controller as part of his/her research framework for each project.

AOIR suggestion:

- Researchers need not obtain informed consent, etc., from subjects if:
  - The data is collected from the public sphere with *no intervention* with the persons whose activities are observed and recorded.
  - The *collection of data* does not include personal identifiers which, if released, could result in reputational or financial harm to the person whose activities are observed.

Handling ethics: MIT "Gaydar" project

"Our analysis demonstrates a method of classifying sexual orientation of individuals on Facebook, regardless of whether they chose to disclose that information. Facebook users who did not disclose their sexual orientation in their profiles would presumably consider the present research an invasion of privacy. Yet this research uses nothing more than information already publicly provided on Facebook; no interaction with subjects was required. Although we based our research solely on public information, only a limited subset of our results, which contain no personally identifiable information, is presented in this paper to maintain subject confidentiality."

Source: Carer Jernigan and Behram F.T. Mistree: *Gaydar: Facebook Friendships Expose Sexual Orientation;* First Monday 14:10; 2009.

Data collected only from the public sphere, but disclosure of personal identifiers could lead to harm for data subjects. The researchers treated their data anonymously, never using real names except to validate their predictions during data analysis. The only copy of the raw data was on an encrypted DVD that was held by their advisor. The project was reviewed ethical review board at MIT and approved.

Why is Internet research so special?

Example: Handling ethics

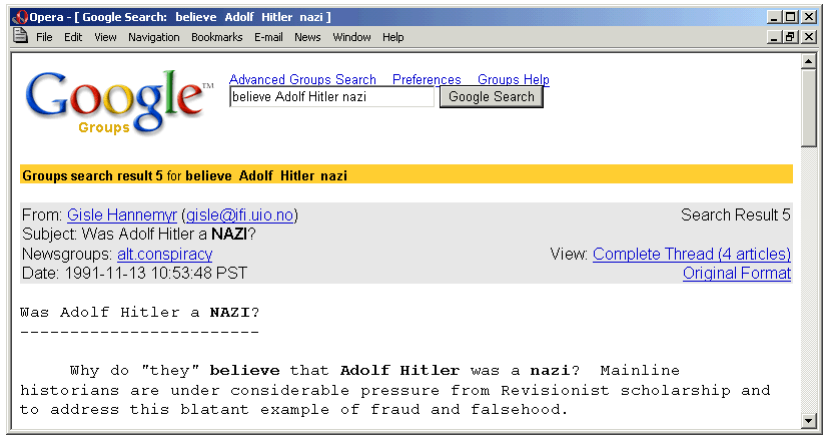
Espen Munch: *En antropologisk analyse av elektronisk nettkommunikasjon,* master thesis in social anthropology at UiO, 1997:

"[Jeg har] valgt å anonymisere både deltakere og grupper i den grad det er mulig i denne oppgaven. Jeg har laget fiktive navn til gruppene, og tatt bort de riktige navnene til opphavsmennene for siterte postinger. Istedenfor ekte aktørnavn har jeg brukt psevdonymer med fiktive fornavn. For at postingene ikke skal bli for lette å spore i News-arkiver, har jeg også fjernet de nøyaktige postingstidspunktene, alt som har med avsenderens epostadresse å gjøre, og eventuelle artikkelnummer."

Pseudonymizing a direct quote

```
From: [John Doe]
Subject: Was Adolf Hitler a NAZI
Newsgroups: [some.newsgroup]
Date: [withheld]
Was Adolf Hitler a NAZI
----------------------
Why do 'they' believe that Adolf Hitler was a
nazi?  Mainline historians are under considerable
pressure from Revisionist scholarship and to
address this blatant example of fraud and
falsehood.
```

… but not very succesfully



Note: Google Groups no longer reveals email addess.

Final words

- The greater the vulnerability of the data subject, the greater the moral obligation of the researcher to protect the data subject from harm.
- Because "harm" is defined contextually, ethical principles are more likely to be understood inductively. That is, rather than universal predicates, doing ethical Internet research requires practical judgment paying attention to context (what in Aristotelian ethics is identified as φρόνησις – *phronēsis* - or practical wisdom).
- When making ethical decisions, researchers must balance the privacy rights of the data subjects with the social benefits of the research and researchers' rights to conduct research. In different contexts the privacy rights of subjects may outweigh the benefits of research.