IN5000 – Qualitative research methods

Privacy (law) and research

Gisle Hannemyr, Ifi

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 (both will be replaced by GDPR in May 2018)
  - European Convention on Human Rights
  - UN Declaration of Human Rights
  - The Nuremberg Code
  - The Belmont Report
  - The Declaration of Helsinki

GDPR
(General Data Protection Regulation)

- In may 2018, GDPR (an EU regulation) will replace both the *EU privacy directive* and the *Norwegian personal data act.*
- In broad terms, the GDPR will regulate data protection in ways that are similar to the legislation it replaces.
- However, there is no longer a need to obtain a permit ("konsesjon") to process sensitive personal data.
- However, all restrictions and data security requirements pertaining to processing sensitive personal data remains, and it is your responsibility to observe them.

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/

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" $\Leftrightarrow$ "digital". Analogue recording is not considered "electronic" for legal purposes.

- or -

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

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.

Is just listening processing of personal data?

- No
- That means that transcribing is better than recording if privacy is important.

Avoiding processing personal data

- As informatics students, you are usually not interestred in *personal data,* but some other aspect during data collection (e.g.: UX).
  - Do not *record* personal data (e.g. when taping an interview, do not ask the user about personally identifying information).
  - Be careful about using video or images that reveal identity (e.g. the user's face).

Your thesis will be read

Drama i TV 2

I epilogen til sin semesteroppgave skriver Olsen:

"Det er fredag ettermiddag den 3. februar 2012. Vi er fire journalister som står bøyd over en laptop i 5. etasje i TV 2-bygget på Karl Johan. Laptopen er av sikkerhetsmessige grunner ikke koplet til nettet, noe som skyldes en minnepinne vi er i ferd med å dytte inn i USB-porten. Flere tusen taushetsbelagte politidokumenter og en rekke bilder er ifølge vår kilde lagret på minnepinnen.

> **Vi er fire journalister som står bøyd over en laptop i 5. etasje i TV 2-bygget**

– Faenihelvete! sier vi i kor i det et bilde av Breivik, sittende i en stol på Utøya like etter pågripelsen, dukker opp på skjermen. Vi løper ned en etasje til vaktsjefen.

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:

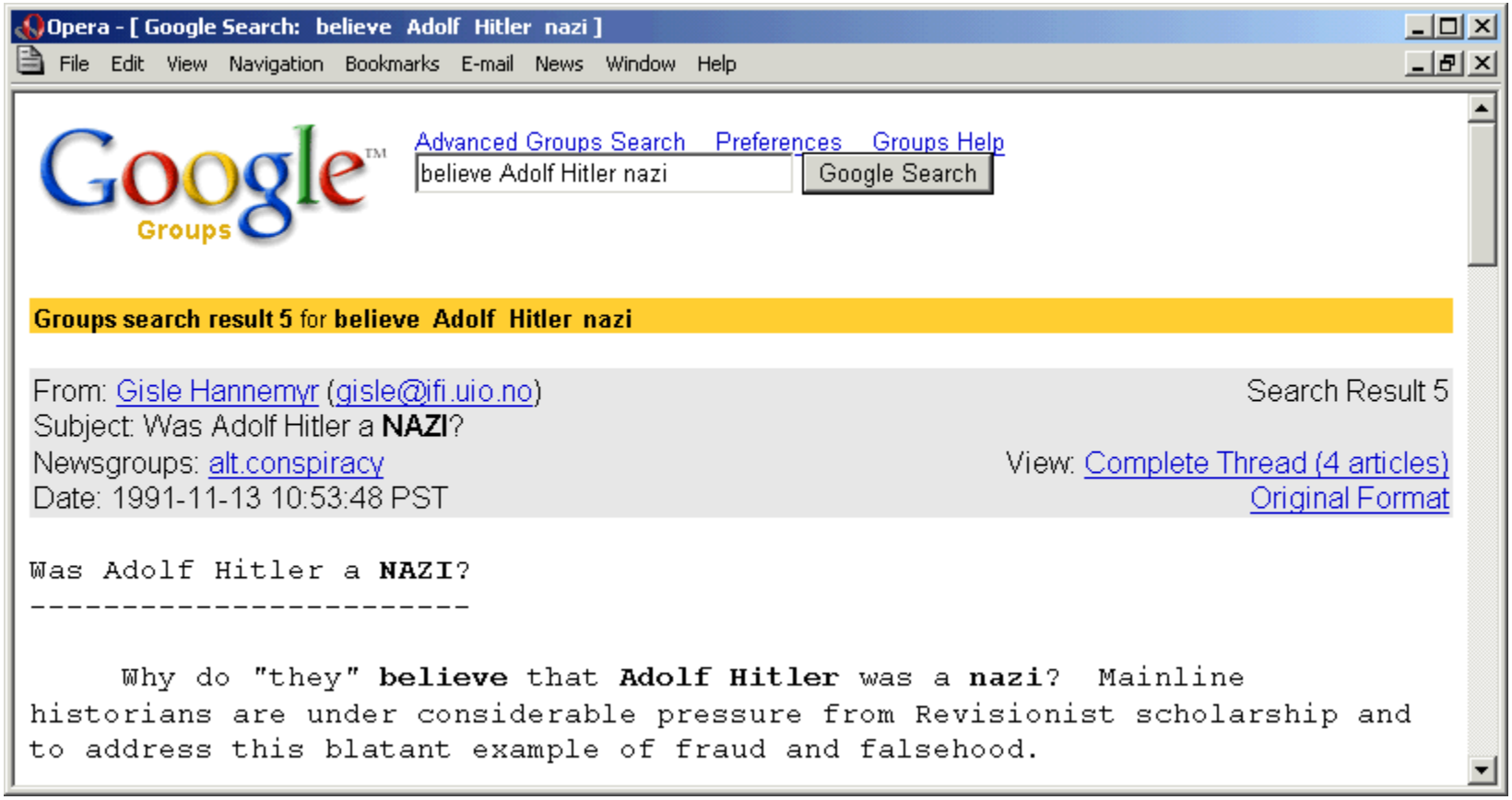
"[Jeg har] valgt å anonymisere både deltakere og grupper i den grad det er mulig i denne oppgaven. Jeg har laget fiktive navn til gruppene, og tatt bort de riktige navnene til opphavsmennene for siterte postinger. Istedenfor ekte aktørnavn har jeg brukt psevdonymer med fiktive fornavn. For at postingene 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.