

Privacy & NSD

IN5000

23 March 2022

Trenton Schulz



Why do people doing research in informatics need to know things about privacy?



THE CORONAVIRUS
PANDEMIC



This Week in Asia / Health & Environment

Across Asia-Pacific and Europe, Covid-19 has thrown up another risk – an addiction to lockdowns

- Experts say the normalisation of extraordinary Covid-19 regulations raises questions about how readily authorities may embrace illiberal responses to future crises
- From Singapore to Australia, governments have also been criticised for using the pandemic to tackle issues not related to public health

Topic | **Coronavirus pandemic**



Source:

South China Morning Post,
3 April 2021

<https://www.scmp.com/week-asia/health-environment/article/3128141/across-asia-pacific-and-europe-covid-19-has-thrown>

Teo, a dean of a Singapore school, discusses some of the issues of collecting digital data

He [Teo Yik Ying] pointed to [South Korea](#)'s implementation of a sophisticated digital contact-tracing system after its 2015 outbreak of Middle East respiratory syndrome, or Mers, which won international plaudits but also raised privacy concerns due to its reliance on personal data such as phone logs and bank records.

Teo, a dean of a school discusses some of the issues of collecting digital data

“In my opinion, that would be a right advancement by a sensible government. But while a sensible government will want to do this for good reasons, equally one can say a renegade government may abuse some of these privileges as well. The checks and balances always need to be there.”

—Teo Yik Ying,
Dean

Saw Swee Hock School of Public Health
National University of Singapore



Photo from Folkehelseinstitute

The first *Smittestopp* app centralized data and had to delete it after request from the Norwegian Data Protection Authority

FHI stoppar all innsamling av data i Smittestopp

FHI sletter alle data frå appen Smittestopp og stoppar midlertidig innsamlinga av data etter varsel om forbod frå Datatilsynet.

Source:

<https://www.nrk.no/norge/fhi-stoppar-all-innsamling-av-data-i-smittestopp-1.15053310>



Hans Ivar Moss Kolseth

Journalist

Vilde Gjerde Lied

Journalist

Mette Kristensen

Journalist

Ugo Fermariello

Journalist

Publisert 15. juni 2020 kl. 06:52

Oppdatert 26. okt. 2020 kl. 19:32

We will examine the rules around collecting data for research

Lov om behandling av personopplysninger (personopplysningsloven)

Dato	LOV-2018-06-15-38
Departement	Justis- og beredskapsdepartementet
Sist endret	LOV-2018-12-20-116
Ikrafttredelse	20.07.2018
Endrer	LOV-2000-04-14-31
Kunngjort	15.06.2018
Rettet	11.02.2019 (PVF art 40)
Korttittel	Personopplysningsloven
EØS/EU/Schengen	EØS-avtalen vedlegg XI nr. 5e (forordning (EU) 2016/679)

Jf. tidligere lov 14. april 2000 nr. 31. Jf. personvernforordningen, også omtalt som GDPR og PVF.

Kapitteloversikt:

Kapittel 1. Personvernforordningen (§1)

Kapittel 2. Lovens saklige og geografiske virkeområde (§§ 2 - 4)

Kapittel 3. Utfyllende regler om behandling av personopplysninger (§§ 5 - 15)

Kapittel 4. Unntak fra den registrertes rettigheter (§§ 16 - 17)

Kapittel 5. Personvernombud (§§ 18 - 19)

Kapittel 6. Tilsyn og klage (§§ 20 - 25)

Kapittel 7. Sanksjoner og tvangsmulkt (§§ 26 - 30)

Kapittel 8. Uekte kameraovervåkingsutstyr mv. (§31)

[Sted] / [dato]

Vil du delta i brukerundersøkelsen «overordnet tittel»?

Jeg er en student i emnet IN1030 – System, krav og konsekvenser ved Institutt for informatikk ved Universitetet i Oslo. Med dette skrevet ønsker jeg å informere hva prosjektet mitt har som formål, spørre deg om du vil delta i prosjektet, samt berette hva deltagelse vil innebære for deg.

Formål

Formålet med mitt prosjekt er å undersøke [overordnet tema og interesseområde for dine obligatoriske oppgaver]. I forbindelse med at jeg konkret ønsker å lære mer om [forskningsspørsmål eller problemstilling], ønsker jeg å [beskrivelse av brukerundersøkelsen din]. Formålet er å forstå dine behov og ditt syn på temaet, slik at jeg kan [mål med brukerundersøkelse].

Deltakelse

Du blir spurt om å delta fordi du faller innenfor min målgruppe, definert som [målgruppe]. Dersom du velger å delta ønsker jeg å [valgt datainnsamlingsmetode] for min datainnsamling. [Brukerundersøkelsen] vil vare i [ca. tid jf. plan], og jeg kommer til å gjøre [valgt datainnsamlingsmetode].

Frivillig deltakelse

Det er frivillig å delta. Du kan når som helst avslutte eller trekke tilbake informasjon som er gitt. Du kan når som helst velge å trekke samtykket uten å måtte oppgi grunn. Dersom samtykket trekkes vil eventuelle personopplysninger som er innsamlet om deg slettes og det vil ikke innebære noen negative konsekvenser for deg at du velger å trekke ditt samtykke.

Personvern: innsamling, oppbevaring, behandling og bruk av dine opplysninger

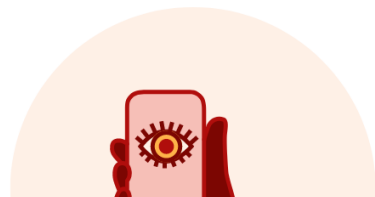
Ingen sensitive personopplysninger (jf. Personvernforordningens artikkel 9 og 10) vil bli innsamlet. Opplysningene vil bli behandlet og lagret på en sikker måte. Opplysningene vil bli anonymisert i transkriberingen og rapporteringen. Opplysningene vil bli behandlet og lagret på en sikker måte. Opplysningene vil bli behandlet og lagret på en sikker måte. Opplysningene vil bli behandlet og lagret på en sikker måte.

We will examine what researchers need to do to collect personal information while following the law

NSD

We ensure that data about people and society can be collected, stored and shared, both safely and legally, today and in the future.

Search for webpages and research data (beta version)



We will examine some ethics around internet research

Photo by [Mati Mango](#) from [Pexels](#)

I am not a lawyer (IANAL), so feel free to examine more



Photo by [Tingey Injury Law Firm](#) on [Unsplash](#)

I won't discuss methods for keeping data secure and private

UNIVERSITY OF OSLO



← Research

[Norwegian version of this page](#)

Services for sensitive data (TSD)

What is TSD?

- provides a platform for researchers at UiO and other public research institutions
- can collect, store and analyze sensitive research data in a secure environment



Research that involves people has several guidelines to regulate processing of data

- Norwegian personal data act (2018)—Includes GDPR
- European Convention on Human Rights
- UN Declaration of Human Rights
- The Belmont Report
- The Declaration of Helsinki
- The Nuremberg Code

Norway implements GDPR through the personal data law *personopplysningsloven* (LOV-2018-06-15-38)

Lov om behandling av personopplysninger (personopplysningsloven)

Dato	LOV-2018-06-15-38
Departement	Justis- og beredskapsdepartementet
Sist endret	LOV-2018-12-20-116
Ikrafttredelse	20.07.2018
Endrer	LOV-2000-04-14-31
Kunngjort	15.06.2018
Rettet	11.02.2019 (PVF art 40)
Korttittel	Personopplysningsloven
EØS/EU/Schengen	EØS-avtalen vedlegg XI nr. 5e (forordning (EU) 2016/679)

Jf. tidligere lov 14. april 2000 nr. 31. Jf. personvernforordningen, også omtalt som GDPR og PVF.

Kapitteloversikt:

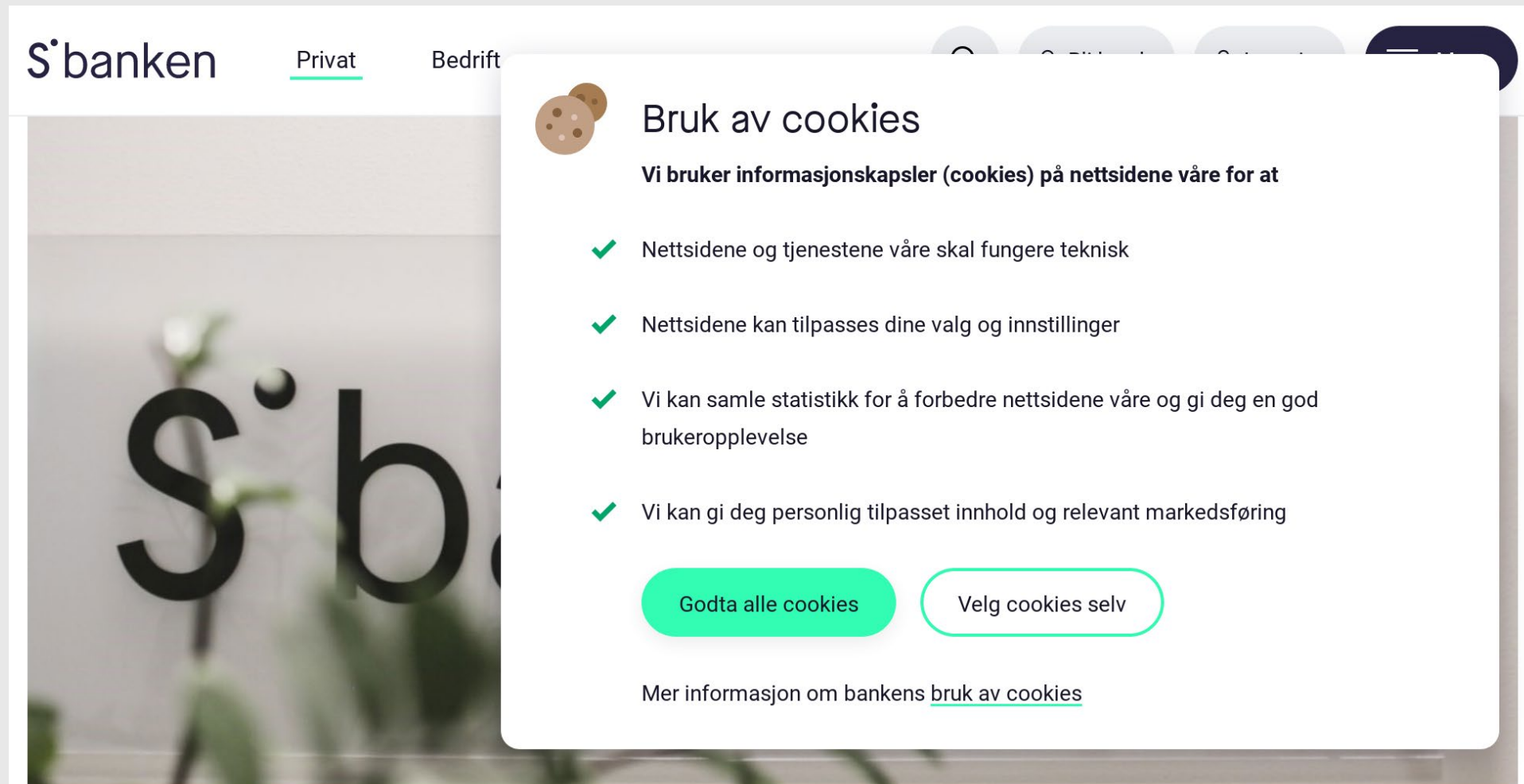
[Kapittel 1. Personvernforordningen \(§1\)](#)

[Kapittel 2. Lovens saklige og geografiske virkeområde \(§§ 2 - 4\)](#)

[Kapittel 3. Utfyllende regler om behandling av personopplysninger \(§§ 5 - 15\)](#)

[Kapittel 4. Unntak fra den registrertes rettigheter \(§§ 16 - 17\)](#)

We experience consequences of the GDPR when we use the internet



The screenshot shows the Sbanken website with a cookie consent dialog box overlaid. The dialog box is titled "Bruk av cookies" and contains the following text and elements:

Bruk av cookies

Vi bruker informasjonskapsler (cookies) på nettsidene våre for at

- ✓ Nettsidene og tjenestene våre skal fungere teknisk
- ✓ Nettsidene kan tilpasses dine valg og innstillinger
- ✓ Vi kan samle statistikk for å forbedre nettsidene våre og gi deg en god brukeropplevelse
- ✓ Vi kan gi deg personlig tilpasset innhold og relevant markedsføring

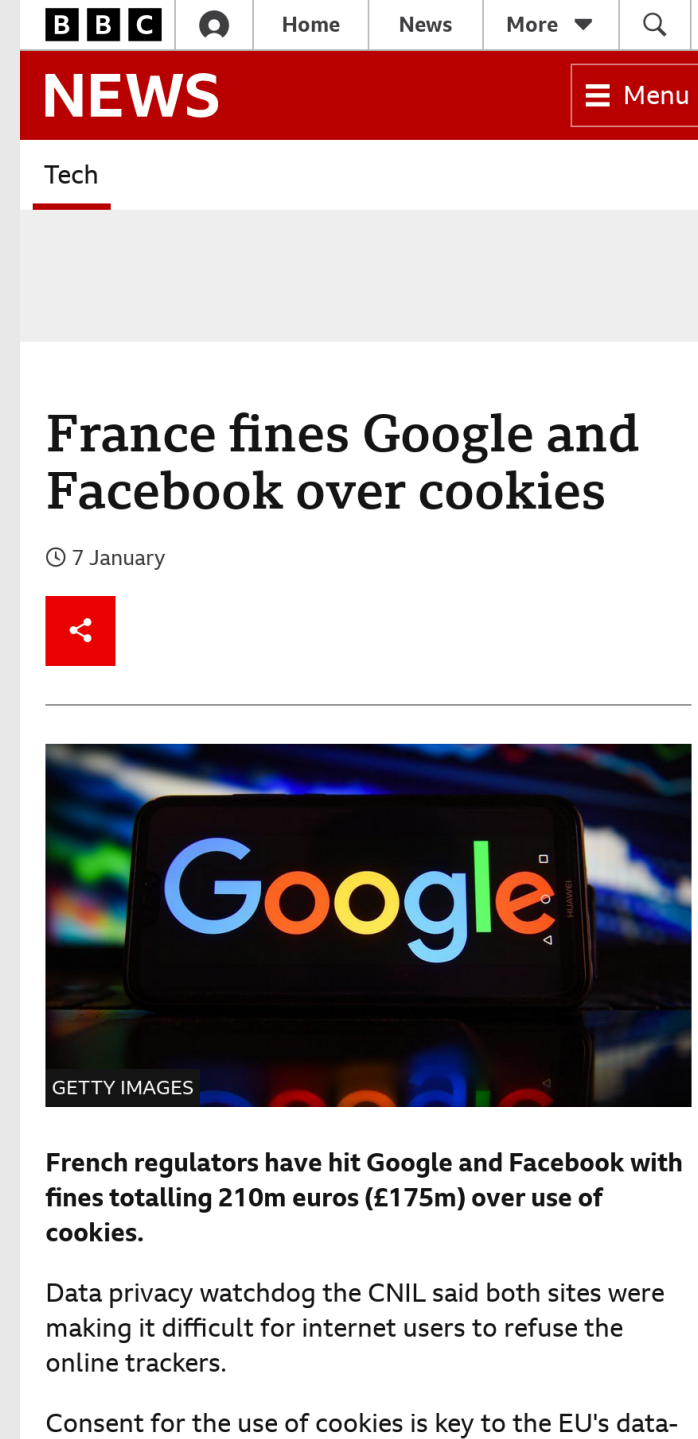
Godta alle cookies Velg cookies selv

Mer informasjon om bankens [bruk av cookies](#)

Companies also get to experience the GDPR

Source:

<https://www.bbc.com/news/technology-59909647>



The image is a screenshot of a mobile news application interface. At the top, there is a navigation bar with the BBC logo, a home icon, and menu items for 'Home', 'News', and 'More'. A search icon is also present. Below this is a red header with the word 'NEWS' in white and a 'Menu' button. The main content area is titled 'Tech' and features a large headline: 'France fines Google and Facebook over cookies'. Below the headline is the date '7 January' and a red share icon. A photograph of a smartphone displaying the Google logo is shown. Below the photo is a sub-headline: 'French regulators have hit Google and Facebook with fines totalling 210m euros (£175m) over use of cookies.' The main body of the article begins with the text: 'Data privacy watchdog the CNIL said both sites were making it difficult for internet users to refuse the online trackers.' The article is attributed to 'GETTY IMAGES'.



NEWS

Menu

Tech

France fines Google and Facebook over cookies

7 January



GETTY IMAGES

French regulators have hit Google and Facebook with fines totalling 210m euros (£175m) over use of cookies.

Data privacy watchdog the CNIL said both sites were making it difficult for internet users to refuse the online trackers.

Consent for the use of cookies is key to the EU's data-

There are several terms one should know when working with data about people

- Data subject
- Personal Data
- Processing
- Controller
- Processor

Data subject (*den registrerte*) is a natural person



Personal data (*personopplysninger*) is any information relating to an identified or identifiable natural person (either directly or indirectly)



Personal data (*personopplysninger*) is any information relating to an identified or identifiable natural person (either directly or indirectly)



Kari Nordmann
Fake street 123
Oslo
Norway



Processing (*Behandling*) is any series of operations done to personal data

Photo by [Christina Morillo](#) from [Pexels](#)

Controller (*behandlingsansvarlig*) is the person who determines the purpose of processing of the personal data and the means



Photo by [Christina Morillo](#) from [Pexels](#)

Processor (*datahandler*) is the person who processes the personal data on behalf of the controller

Photo by [Christina Morillo](#) from [Pexels](#)

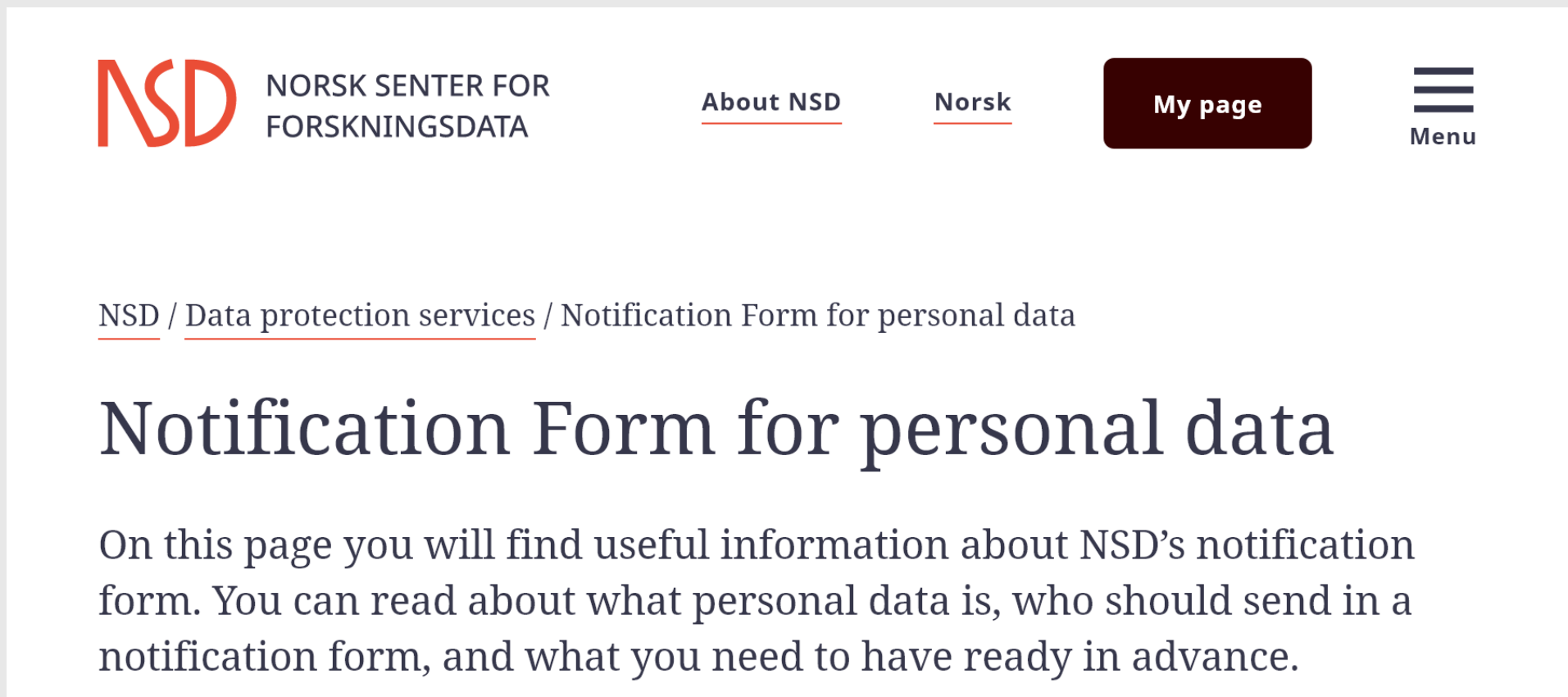
Discussion point

Think about your research projects. Do you think that you will be collecting personal data?
Why or why not?

Are there ways that you think can avoid collecting personal data?

GDPR specifies requirements for the controller doing research that involves collecting or processing personal data

Main requirement for researchers: all research involving personal data must be reported using a special form to the privacy ombudsman for research (*personvernombudet for forskning*) (<https://www.nsd.no>)



The screenshot shows the top navigation bar of the NSD website. On the left is the NSD logo (NORSK SENTER FOR FORSKNINGSDATA). In the center are two underlined menu items: "About NSD" and "Norsk". On the right is a dark red button labeled "My page" and a hamburger menu icon labeled "Menu". Below the navigation bar is a breadcrumb trail: "NSD / Data protection services / Notification Form for personal data". The main heading is "Notification Form for personal data". The introductory text reads: "On this page you will find useful information about NSD's notification form. You can read about what personal data is, who should send in a notification form, and what you need to have ready in advance."

Basic rules for “Should I contact the NSD?”

1. Recording or processing of information about individuals by electronic means.
2. A manual register containing special categories of personal data will be created

Filling out the notification form (*meldeskjema*) is straight forward; just remember...

Plan your study carefully: roundtrips with NSD will take time



Photo by [fauxels](#) from [Pexels](#)

Have your paperwork in order; all documents need to be included

WORKSHOP DESIGNS FOR THE MECS PROJECT

This is a work-in-progress document for workshops at Kampen in the MECS project for autumn 2018. Right now, I can only come up with two, but I'm happy to expand this if possible. *Everything* in this document is tentative and can be changed!

WORKSHOP 1: MATERIALS FOR A ROBOT "SHELL"

{Oppgave (1. utkast):

Hvis du skulle anskaffe en robot til å hjelpe deg i huset, og kunne bestemme hvordan den skulle se ut, hvilke ideer til utforming får du da? }

Oppgave oppdatert, eksempel:

Hvis du skulle få en robot til å hjelpe deg med noen oppgaver i hjemmet, oppgaver som kan gjøre deg mer selvstendig. Hvis du kunne bestemme hvordan den skulle se ut, hvilke ideer til utforming får du da? (Materialer, utforming)

Forestille deg at du skal ha en robot som skal bor hjemme hos deg. Roboten kan styres med lyd eller fjernkontroll eller noen annet, og den kan hente ting for deg (eventuelt finner ting) og bære ting for deg (for eksempel en bok, din telefon, noen andre ting du vil ha med deg). Vi skal kalle det en frakkebord robot. Vi er interessert i hvordan en robot kunne eventuelt ses ut og hva slags materialer man kan bruke for å lage den roboten. Vi skal vise deg noen eksempler som kan brukes for inspirasjon.

Samtykkeerklæring om deltakelse i workshop



Bakgrunn og formål

Vi ønsker å hente idéer på hvordan en robot hjemme skal se ut og hva slags materialer robot bør bestå av. Vi er også interessert i hva slags krav og behov eldre har for å være trygt og selvstendig hjemme. Dette vil vi bruke for å lage prototyper videre.

MECS

Et pågående forskningsprosjekt ved Universitetet i Oslo, Institutt for Informasjonsteknologi (IKT) i sammenheng med det å bo trygt hjemme. I tillegg vil vi på denne måten demonstrere mulighetene for økt sikkerhet og personvern til hjemmet for mennesker som bor hjemme. Prosjektets formål er å undersøke bruk av informasjons- og kommunikasjonsteknologier (IKT) i sammenheng med det å bo trygt hjemme. I tillegg vil vi på denne måten demonstrere mulighetene for økt sikkerhet og personvern til hjemmet for mennesker som bor hjemme. Prosjektet vil på denne måten demonstrere mulighetene for økt sikkerhet og personvern til hjemmet for mennesker som bor hjemme.

Processing the form takes approximately 30 days: **no data collection** until the form is processed



Photo by [Anete Lusina](#) from [Pexels](#)

When in doubt, talk with your advisor or contact the NSD

The NSD is there to help you comply with the law; there are not there to play “gotcha”.

Vike and L'orange Fürst pointed out several issues with the GDPR, NSD, and data collection for anthropologists



If you research changes and you need to collect different data, you need to inform about changes




The screenshot shows the NSD (Norwegian Supervisory Authority for Data Protection) reporting interface. The header is dark blue with the NSD logo and the text 'MELDESKJEMA FOR BEHANDLING AV PERSONOPPLYSNINGER'. Below the header, there are dropdown menus for 'Norsk' and 'Trenton W. Schulz'. The main content area is white and contains the following elements:

- Breadcrumbs: [Meldeskjema](#) / [Close the Gap](#) / Meldinger
- Section Header:

Meldinger
- Referansenummer: 116844
- Status: Vurdert
- Action Button: **Rediger meldeskjema** (highlighted with a red circle)
- Text Input Field:
- Footer Note: Merk: Meldingen vil bli synlig for din institusjon og alle prosjektet er delt med.
- Bottom Button: [Send melding](#)

You cannot reuse your data for another purpose; you must ask for new consent

MECS
Multimodal Elderly Care Systems



Samtykkeerklæring om deltakelse i workshop

Bakgrunn og formål

Vi ønsker å hente idéer på hvordan en robot hjemme skal se ut og hva slags materialer en robot bør bestå av. Vi er også interessert i hva slags krav og behov eldre har for å bo trygt og selvstendig hjemme. Dette vil vi bruke for å lage prototyper videre.

Om MECS

MECS er et pågående forskningsprosjekt ved Universitetet i Oslo, Institutt for Informatikk. Prosjektets formål er å undersøke bruk av informasjons- og kommunikasjonsteknologier (IKT) i sammenheng med det å bo trygt hjemme. I dette ligger det å forstå beboeres behov gjennom brukersentrert design, utvikle effektiv sansing av hverdagsaktiviteter, samt utvikle læringsmetoder for å forutse uønskede hendelser. Prosjektet vil på denne måten demonstrere mulighet for modellering og prediksjon for økt sikkerhet og personvern til hjemmet.

Hva innebærer deltakelse i studien

Målgruppen for studien er eldre mennesker som bor hjemme og som er villige til å delta i studien fordi du passer i denne beskrivelsen.

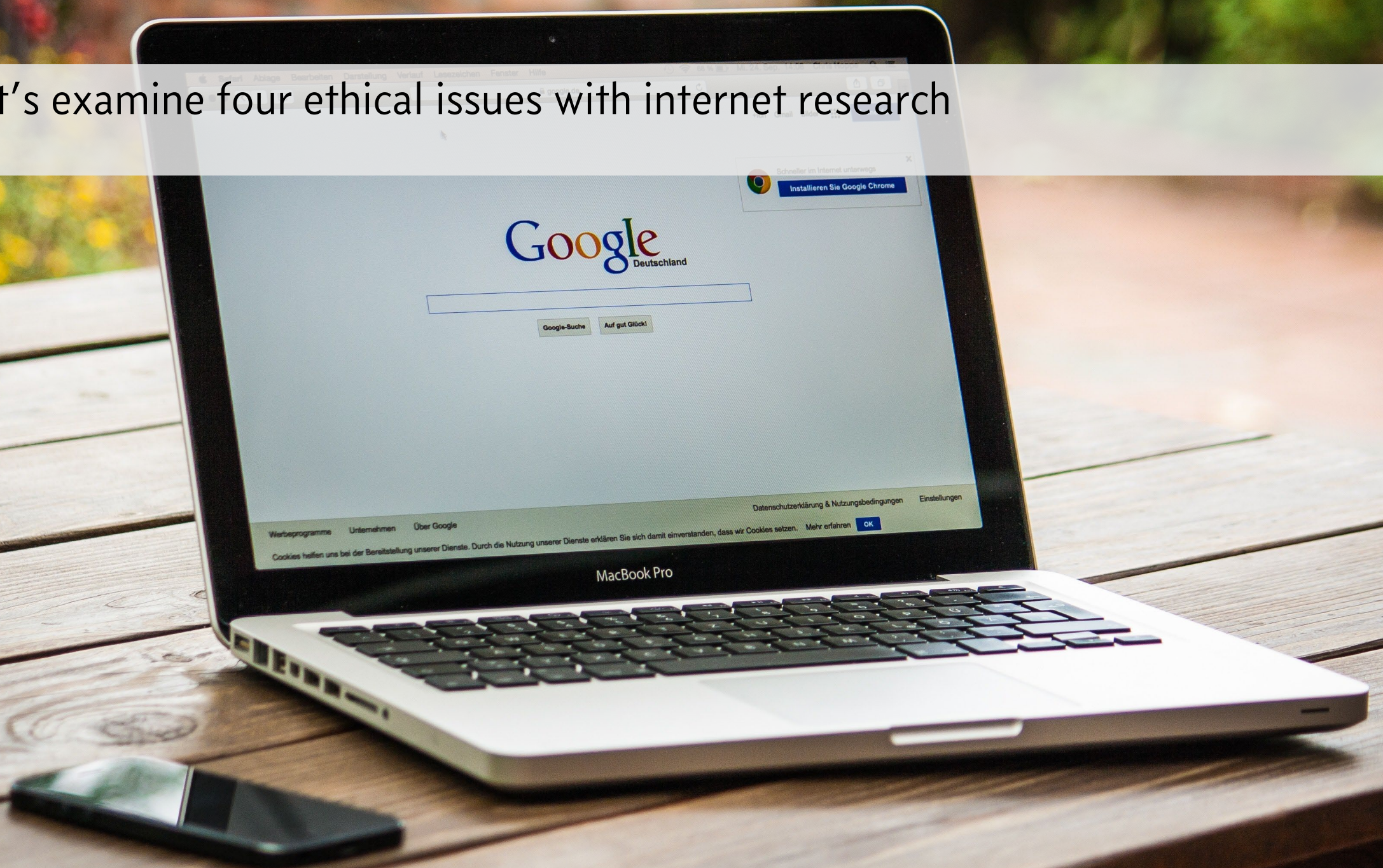
Du vil delta i en workshop hvor du vil bli introdusert til MECS og presentert noen av de teknologier som er presentert i denne erklæringen.

Unless you are working with health data you *probably do not need* approval from REK

REK (*Regional Komiteer for Medisinsk og Helesfaglig Forskningsetikk*) looks at all research projects that involve medicine, health, or research biobanks.

REK's approval only looks at the health and medical aspects of the research, you still need to fill a form with NSD if you are processing personal data.

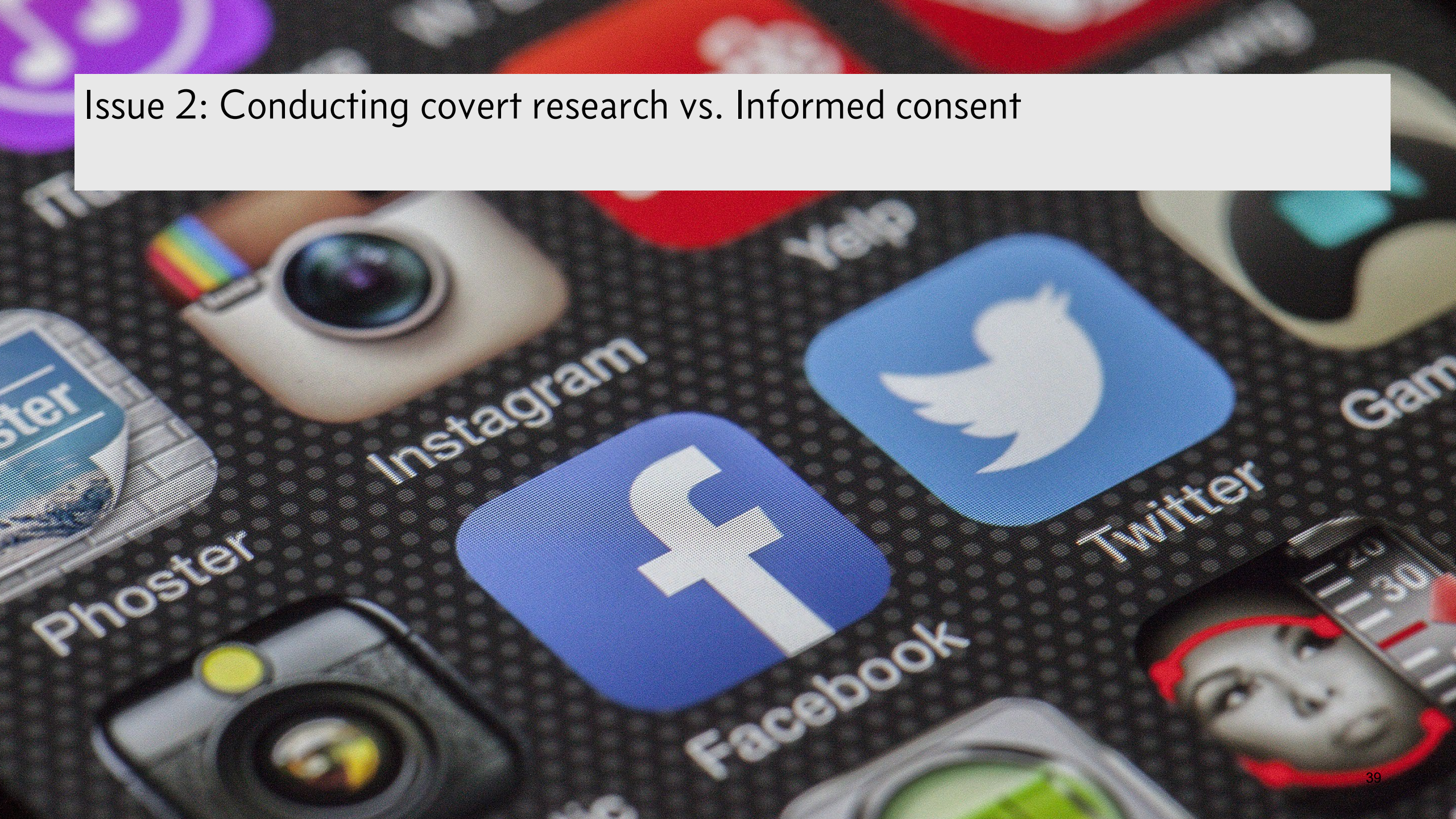
Let's examine four ethical issues with internet research



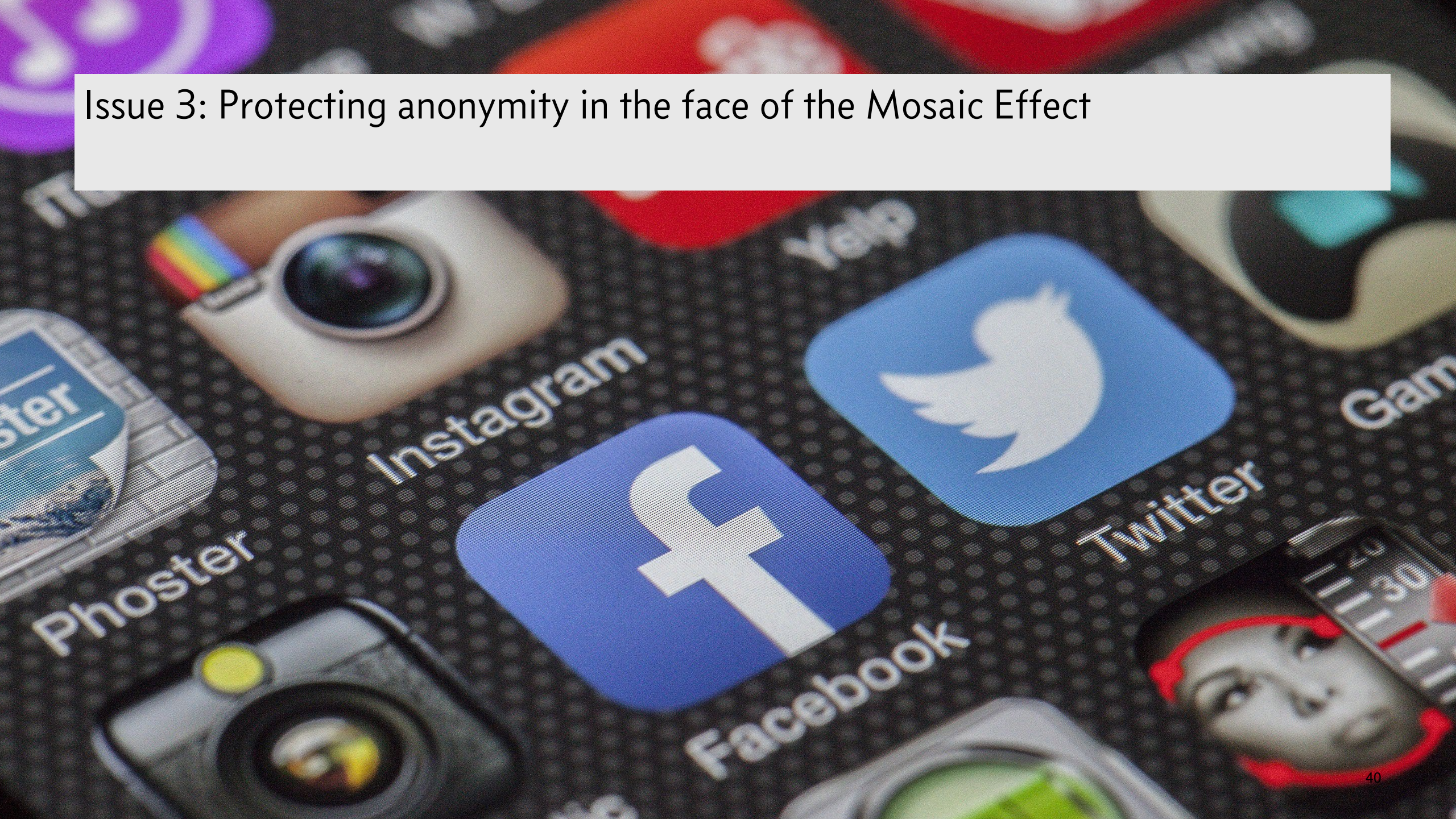
Issue 1: Is online interpersonal media (social media) considered public or private information?



Issue 2: Conducting covert research vs. Informed consent



Issue 3: Protecting anonymity in the face of the Mosaic Effect



Issue 4: Handling the raw data from internet research



Internet Research Ethic 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>
<http://aoir.org/reports/ethics2.pdf>
<https://aoir.org/reports/ethics3.pdf>

Issue 1: Is social media information private or public + Issue 2 covert research versus informed consent: The *Gaydar* Study

Jernigan, C., & Mistree, B. F. T. (2009). Gaydar: Facebook friendships expose sexual orientation. *First Monday*, 14(10). <https://doi.org/10.5210/fm.v14i10.2611>

Jernigan & Mistree's look at the relations between people on Facebook to find a person's orientation; from the abstract

Public information about one's coworkers, friends, family, and acquaintances, as well as one's associations with them, implicitly reveals private information... After analyzing 4,080 Facebook profiles from the MIT network, we determined that the percentage of a given user's friends who self-identify as gay male is strongly correlated with the sexual orientation of that user, and we developed a logistic regression classifier with strong predictive power.

Jernigan & Mistree collected the information without contacting the user; later in the article, emphasis added

“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.”

Discussion point

What are your opinions about how the data was collected?

Do you think it was necessary to gather informed consent?

Does the GDPR provide any sort of guidance in this issue?

AOIR has suggestions about when informed consent is *not necessary*

- Data is collected from the public sphere with no intervention from 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

Issue 3: Protecting anonymity: Researchers have a duty to protect the anonymity of the people in the data

Researchers 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*

The Mosaic Effect: your anonymous data may reveal more information when combined with other information

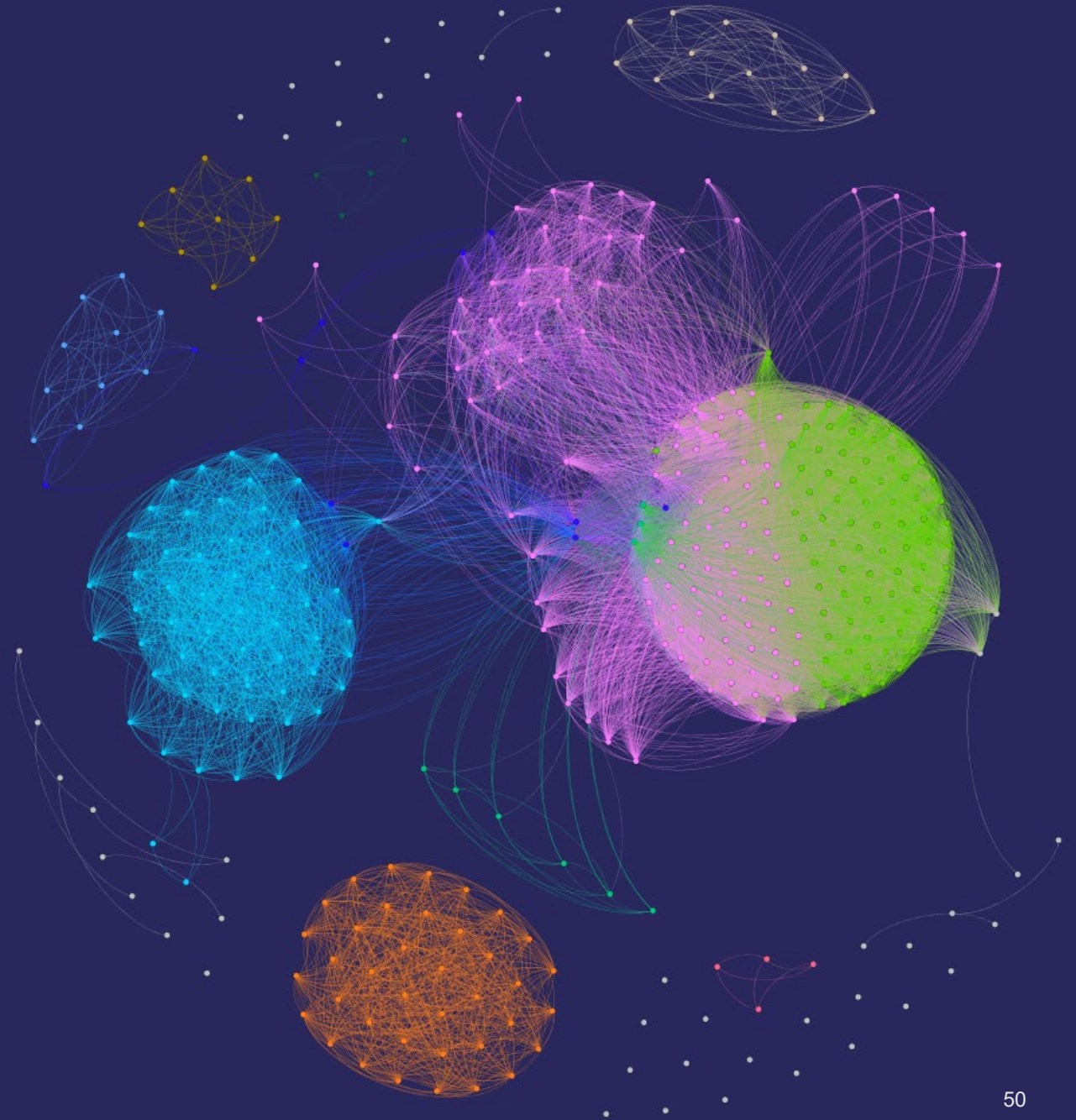
“The Mosaic Effect occurs when the information in an individual dataset, in isolation, may not pose a risk of identifying an individual (or threatening some other important interest such as security), but when combined with other available information, could pose such risk. Before disclosing potential 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.”

—Open Data Policy-Managing Information as an Asset
(<http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>)

An example of the mosaic effect in humanitarian datasets: an edge indicates common column headers. Over 400 datasets were analyzed, over 90% (377) of datasets shared information with at least one other dataset.

Source: *Exploring the Mosaic Effect on HDX Datasets*

(<https://centre.humdata.org/exploring-the-mosaic-effect-on-hdx-datasets/>)



Mosaic Effect Example 1: One master's student tries to protect a source

Espen Munch: *En antropologisk analyse av elektronisk nettkommunikasjon*, 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.”

An attempt at 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.

An attempt at pseudonymizing a direct quote...

From: [redacted]
Subject: [redacted]
Newsgroup: [redacted]
Date: [redacted]
Was Accepted: [redacted]

Why do [redacted]
historians [redacted]
scholars [redacted]
falsehood [redacted]

Don't look this up in
a search engine!!!

In 2006, AOL released “anonymized” search data for research purposes...

AOL Proudly Releases Massive Amounts of Private Data

Michael Arrington @arrington?lang=en / 15 years

Yet Another Update: [AOL: “This was a screw up”](#)

Further Update: Sometime after 7 pm the download link went down as well, but there is at least one [mirror site](#).

AOL is in damage control mode – the fact that they took the data down shows that someone there had the sense to realize how destructive this was, but it is also an admission of wrongdoing of sorts. Either way, the data is now out there for anyone that wants to use (or abuse) it.

Update: Sometime around 7 pm PST on Sunday, the [AOL](#)

Source:

<https://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/>

The User ID in the data could be linked up with other information

The New York Times

A Face Is Exposed for AOL Searcher No. 4417749



By Michael Barbaro and Tom Zeller Jr.

Aug. 9, 2006

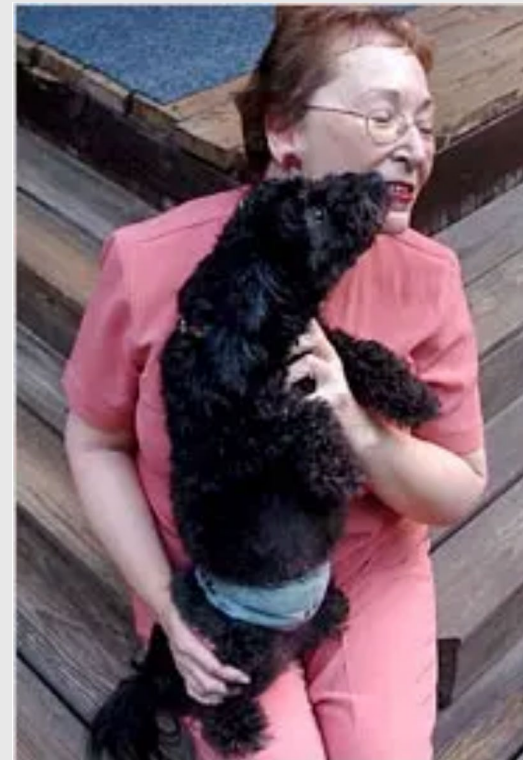
Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from “numb fingers” to “60 single men” to “dog that urinates on everything.”

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for “landscapers in Lilburn, Ga.” several people with the last name

Source:

<https://www.nytimes.com/2006/08/09/technology/09aol.html>



Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem. Erik S. Lesser for The New York Times

Issue 4: Protecting raw data for good research practice

- To assist in peer-review and a possibility for helping in replication a study, raw data should be available on request.
- Keep the data, but pseudonymize the records using different numbers of real IDs. Keep raw data access restricted

In summary, researchers must protect the data they collect about people, but there is no universal solution

- The greater the vulnerability of the data subject, the larger the moral obligation of the researcher to protect the data subject from harm
- Harm depends on context and researchers need to have a good judgement about what can cause harm

“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”

—Gisle Hannemyr