# Endelig rapport for kildekodegjennomgang av løsning for digital smittesporing av koronaviruset

18.05.2020

**Ekspertgruppen for kodegjennomgang av løsning for digital smittesporing**

*Jeanine Lilleng (leder av ekspertgruppen)*
*Odd Rune Lykkebø*
*Bjørn Borud*
*Øyvind Indrebø*
*Eivind Andreas Arvesen*
*Aleksander Slater*
*Elina Sande Heimark (sekretær)*

# Innledning

Utbruddet av COVID-19 er et alvorlig utbrudd av smittsom sykdom som kan få alvorlige helsekonsekvenser for mange mennesker. Som et tiltak for å minske smitte av dette viruset har Simula utviklet appen Smittestopp på bestilling fra Folkehelseinstituttet (FHI). Deler av systemet som ligger bak appen blir levert av Norsk Helsenett og FHI selv.

Ekspertgruppen for kodegjennomgang av løsningen for digital smittesporing er satt ned av Helse- og Omsorgsdepartementet (HOD) for gjennomgang av appen Smittestopp og bakenforliggende løsninger. HOD har bedt IKT Norge om å anbefale aktuelle deltakere. Ingen av deltakerne er knyttet til FHI, HOD eller Simula.

Vi viser til mandat[1] som ble publisert av regjeringen 8.4.2020.

Mandatet sier at det av sikkerhetsmessige grunner "*er besluttet at koden i prosjektet ikke skal gjøres åpent tilgjengelig. For å øke tilliten til systemet, vil en gruppe uavhengige personer få full innsikt i alle sider av koden, installasjonen av den og bruken av koden.*"

Vi har valgt å beskrive løsningen relativt detaljert, slik at det vil være mulig for brukere, tilsynsmyndigheter, folkevalgte og myndighetene å få en god forståelse av hvordan denne løsningen virker, hva slags data som samles inn, hvordan de behandles, hvordan de brukes og hvordan de deles. Vi beskriver delene av løsningen med enkle ord først, og deretter grundigere, for å favne et så bredt publikum som mulig.

Gruppen skal ifølge mandatet levere følgende:

1. En åpen rapport til Helse- og omsorgsdepartementet med en overordnet vurdering av om *sikkerhet* og *personvern* er forsvarlig ivaretatt.

2. En rapport unntatt offentlighet til Simula og FHI med kopi til HOD om eventuelle identifiserte svakheter som må rettes.

[1] "Mandat ekspertgruppe.pdf - Regjeringen.no." https://www.regjeringen.no/contentassets/82254fd2dd5f431cb98f57ac28ca1510/mandat-ekspertgruppe.pdf. Aksessert 21 Apr. 2020.

# Avgrensninger

09.04.2020 ble ekspertgruppens midlertidige rapport[2] publisert, med våre beskrivelser, vurderinger og anbefalinger på dette tidspunktet.

FHI har nå hatt tid til å få på plass store deler av løsningen. I denne rapporten har vi sett på alle delene som inngår i appen Smittestopp og systemet rundt denne. Programvare blir sjelden helt ferdig, og man må alltid forvente at funksjonalitet fortsetter å forandre seg. Det er ikke mulig å uttale seg om framtidige endringer i funksjonalitet eller bruk av data. Alle komponentene som inngår i løsningen er i skrivende stund tatt i bruk. Det er to unntak: Automatisk SMS utsending ved kontaktsporing og modulen for avidentifisert og aggregerte data. Våre evalueringer og beskrivelser er basert på slik systemet ser ut i dag, 18.05.2020.

Data fra systemet vil også bli brukt til å lage datasett som eksporteres. Vi gjør en analyse av avidentifisering og aggregering av dataen som er samlet og datasettene som lages for bruk etter 30 dager, men ikke hva datasettene brukes til etterpå. Systemhendelser i Azure eksporteres til mnemonic til deres Security Operations Center (SOC). Vi beskriver data som deles med dette formålet, men vi har ikke gjort noen sikkerhetsvurdering av dette systemet.

Som ekspertgruppe har vi primært hatt fokus på å beskrive løsningen for å gjøre det mulig for andre å ta et valg om å bruke, eller ikke bruke appen basert på fakta. Samtidig ser vi at det for legfolk kan være krevende å konkludere basert på disse beskrivelsene, og vil derfor også diskutere funnene våre.

Vi har ikke hatt tilgang til fullstendig konfigurasjon av skyløsningen, ettersom det er flere ting som er opprettet ikke-programmatisk. Vi har heller ikke hatt direkte tilgang til byggesystem eller produksjonsmiljø, og kan dermed for eksempel ikke verifisere om konfidensialitet og integritet er ivaretatt med tanke på utro tjenere.

Tilsyn og revisjon av bruk av data etter lansering er utenfor omfanget av vårt mandat. Vi anbefaler at det opprettes en uavhengig part som har ansvar for tilsyn med og revisjon av tilgang til dataene.

Det må også poengteres at en ekspertgruppe, selv med lengre tid til disposisjon enn vi har hatt, ikke vil kunne finne *alle* feil i en slik løsning. Denne rapporten må regnes som et øyeblikksbilde, og vil ikke nødvendigvis reflektere løsningens fremtidige tilstand.

---

2 https://www.regjeringen.no/globalassets/departementene/hod/fellesdok/rapporter/200409_forelppig_rapport_ekspertgruppe_sporingsapp.pdf

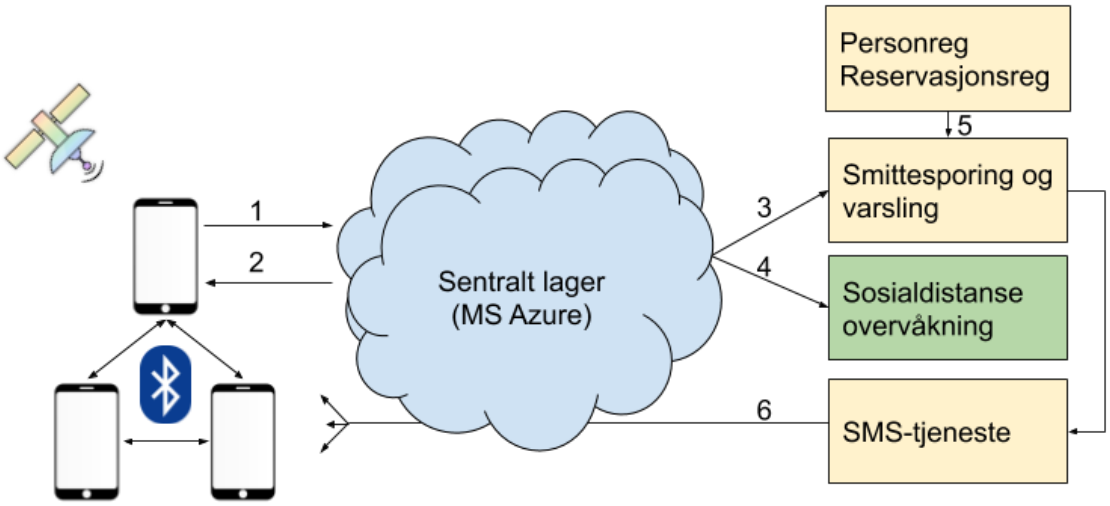
# Fremgangsmåte

Ekspertgruppen ble etablert lørdag 4. april og fikk tilgang til de første delene av kildekoden og rapporter søndag 5. april. Vi har fått direkte tilgang til alle Git-repositories der utvikling skjer. Det er tatt utgangspunkt i kildekode, arkitekturskisser og samtaler med leverandørene for å forstå helheten i systemet. Deretter har vi gjort dypdykk i koden og gjennomført kodeanalyse for å identifisere mangler. Det er brukt over 1000 timer til dette arbeidet, så ekspertgruppen har satt seg godt inn i systemet, og har opparbeidet seg en god helhetsforståelse på tvers av alle komponentene.

I store deler av perioden har vi hatt daglige møter med Simula og FHI hvor vi har presentert våre funn. Vi har også brukt disse møtene for å innhente mer informasjon når dette har vært nødvendig. Simula, Folkehelseinstituttet (FHI) og Norsk Helsenett (NHN) har gjort endringer basert på våre tilbakemeldinger.

Deler av skyløsningen er satt opp av Microsoft, og mye her er konfigurert direkte på serverne gjennom interaktive grensesnitt snarere enn som konfigurasjon sjekket inn i versjonskontroll. For å få oversikt over disse delene av systemet har vi hatt møter med de som har satt opp dette hos Microsoft og brukt f.eks. deling av skjerm, slik at vi har kunnet se hvordan ting er satt opp selv om vi ikke selv har hatt direkte tilgang til løsningen.

# Overordnet beskrivelse av løsningen




Figur 1. 1: Autentisering og lokasjon/nærkontakt-data. 2: Dine innsynsdata, samt eventuell ny identifikator. 3: Liste over hvem som har vært i kontakt med smittet person, og detaljer rundt disse personene. 4: Aggregerte og avidentifiserte data. 5: Telefonnummer fra person og reservasjonsregisteret brukes for å identifisere personer. 6: SMS-tjeneste sørger for å varsle eventuelle personer.

Smittesporingsløsningen består i grove trekk av tre deler:

- Appen (Android og iOS) som brukere installerer på mobiltelefonen sin
- En skyløsning som er levert av NHN med støtte av Microsoft (Azure)
- Webapplikasjoner fra FHI og NHN.

Appene kommuniserer kryptert med skyløsningen. I skyløsningen lagres data både i en SQL-database og i råformat i Azure Data Lake. Når man har en “smittet” bruker, kan FHI innhente en liste over kontakter som denne brukeren har vært i nærheten av fra en analyse-applikasjon som kjører i skyløsningen.

Den samlede lokasjonsdata blir også brukt til å produsere 5 datasett med aggregert data, samt varianter av disse. Disse datasettene vil bli brukt av FHI for å følge med på effekten av smitteverntiltakene samt å gjøre forskning.

## Smittestopp-appen

Appen benytter brukerens telefonnummer for å identifisere og autentisere brukeren. Ved hjelp av GPS blir posisjonen din lagret på mobiltelefonen. Samtidig benytter appen Bluetooth Low Energy (BLE) til å lete etter andre enheter i nærheten som også har appen installert og estimerer distansen til disse.

Dine lokasjonsdata, samt avstanden til andre mobiltelefoner i nærheten, lagres så i en sentral skyløsning. Dataene her er ikke tilgjengelige for andre enn FHI, som kan bruke informasjonen til å finne ut hvem som har vært i nærheten av en person som er syk. De lagrede data skal også brukes til forskning, men da er dataen avidentifisert og aggregert slik at den ikke kan brukes for å identifisere deg.

Brukeren blir ledet gjennom en introduksjonssekvens, hvor personvernerklæringen må aksepteres og fødselsdato fylles inn. Forespørsler om tillatelse til å bruke telefonens steds-, Bluetooth- og varslingstjenester blir også gjennomført. Brukerens fødselsdato blir lagret på telefonen, og er brukt for å verifisere at brukeren er over 16 år. Deretter registrerer brukeren sitt telefonnummer mot Microsoft Active Directory / B2C. Det genereres så en unik identifikator (enhets-ID) som bindes til telefonnummeret som blir brukt i all videre kommunikasjon mot skytjenestene.

Appen registrerer dine bevegelser via lokasjons-API-er på din mobiltelefon, dvs GPS. GPS brukes aktivt og vi forventer at dette vil ha noe effekt på batterilevetiden. Dette er forsøkt å begrenses ved å skru ned presisjonen når enheten holder seg i ro. Lokasjonsdata som breddegrad, lengdegrad, nøyaktighet, hastighet, høyde og nøyaktighet på høyde blir periodisk lagret i en kryptert lokal database på mobiltelefonen.

Bluetooth-enheter i nærheten som også har installert appen annonserer sin tilstedeværelse (eller i tilfelle for iOS: at de er en iOS-enhet). Dette fører til en datautveksling med disse annonserende enhetene via Generic Attribute Profile (GATT) (appendiks F). Enhets-IDer fra de oppdagede og annonserende enhetene sendes over Bluetooth, og blir sammen med signalstyrken og mottakerens sendestyrke lagret kryptert i en lokal database på mobiltelefonen. Det finnes funksjonalitet for å håndtere utfordringer ved at en iOS-enhet vil lete etter andre enheter sjeldnere når den er i bakgrunnsmodus eller skjermen er slått av.

Innsamlet lokasjonsdata og Bluetooth-data blir sendt til Microsoft Azure IoT-Hub over HTTPS, og blir deretter slettet lokalt. Data som er lagret lokalt, men ikke lastet opp enda, blir slettet hvis man logger ut, dersom man blir av-autentisert eller hvis den har mislykkes å laste opp data 40 ganger. Appen har kun skrivetilgang til IoT-Hub.

Appen sender såkalte “heartbeats” til Azure minst en gang i døgnet, etter samme mønster som lokasjons- og Bluetooth-data. Dette brukes for å avgjøre om en bruker fortsatt har appen

installert eller ei. Tilhørende data i skyen blir slettet dersom en bruker ikke lenger anses som aktive. Det sendes i også informasjon om hvorvidt GPS/Bluetooth er skrudd på i appen.

I tillegg sendes bruksanalyse- og telemetri-data løpende fra app til AppCenter.

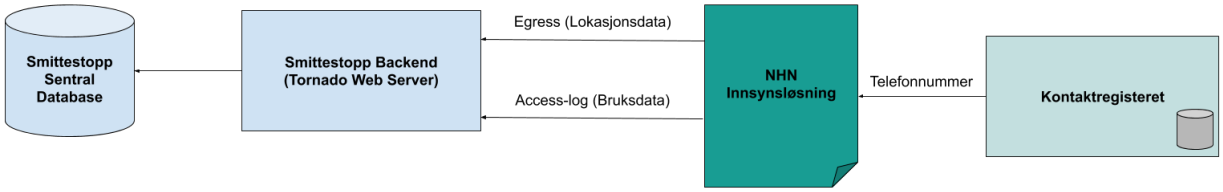
## Innsynsløsning

Innsynsløsningen er en del av Helsenorge sin “Min Helse”-portal. Brukere logger inn med MinID, og får deretter tilsendt en kode på SMS som må tastes inn. Her kan brukere se dataen sin fra bestemte tidspunkt basert på GPS lokasjon. De spesifikke dataene er følgende: nøyaktighet for posisjon, høyde (m.o.h.) og nøyaktighet for høyde.

Brukere kan også se en oversikt over hvem som har vært inne og sett på eller behandlet dataene egen data, og hvilket formål dette har vært i forbindelse med.

Man logger inn på normalt vis med BankID eller lignende nivå 4 elektronisk ID, på lik linje med de fleste statlige systemer i Norge. Her gjenbrukes en løsning som allerede har vært i bruk i mange år. Deretter valideres mobiltelefonnummeret som er koblet opp mot personnummeret. Dette er en ny komponent som er laget for å validere koblingen mot telefonnummeret før man får innsyn i data for dette telefonnummeret.

De fleste har allerede mobiltelefonnummeret sitt registrert i Kontakt- og reservasjonsregisteret. Hvis telefonnummeret ikke er registrert eller det registrert nummer er feil, må dette oppdateres i kontaktregisteret før man kan logge på. Det vil med andre ord si at innsynsløsningen gjør et oppslag i kontaktregisteret, og telefonnummeret bekreftes via en engangskode generert av en kryptografisk slumptallsgenerator som blir sendt per SMS til mobiltelefonen. Brukeren har da 3 minutter på å skrive inn koden.




Etter mobiltelefonnummeret er bekreftet gjør innsynsløsningen 2 API kall mot Smittestoppserveren. Denne er sikret med både sertifikat og IP-restriksjoner, og krever den påloggede brukeren sitt token. Det vil med andre ord si at dette kun kan bli brukt av den påloggede brukeren via NHN sine servere. GPS-data og aksess-logg blir lastet via predefinerte SQL-prosedyrer for den påloggede brukeren og vist. Denne dataen blir ikke lagret hos NHN.

# Varslingsløsning

Ved et bekreftet smittetilfelle er kommunen pålagt å gjøre smittesporing. Det innebærer å finne ut hvilke personer en bekreftet smittet (kalt en indekspasient) har vært i nærkontakt med. De som har vært i nærkontakt med en indekspasient vil så få påbud om å oppgi hvem de har vært i kontakt med og deretter gå i karantene. Hvis indekspasienten har installert appen Smittestopp vil FHI kunne hente ut nærkontakter til andre brukere av appen. Disse kan så enten varsles manuelt via kommunen, eller automatisk med SMS. Om brukeren får en SMS vil det være mulig å verifisere at denne kom fra et legitimt sted ved å sjekke i innsynsløsningen.

I den første fasen vil FHI og Simula verifisere om kontaktdefinisjonen gir forventede resultater, eller om den må justeres. I denne perioden vil ikke automatisk SMS-varsling være aktivert.

Mer detaljert; Kommunelegen deler den smittedes telefonnummer og tidsrommet indekspasienten kan ha smittet andre til FHI. Standardverdi for smittsom periode er 7 dager før positiv test. Smittejegerene i FHI sender så et REST-API-kall til Simula sin backend. Dette starter en asynkron jobb som produserer en rapport på nærkontakter for denne indekspasienten. APIet returnerer en unik URL for forespørselen, der rapporten er tilgjengelig når den er ferdig.

## Analyse av kontaktdata, analysejobben

Analysejobben benytter rådata lagret i SQL-basen fra den smittedes mobiltelefon. Disse rådataene blir prosessert med ulike “forretningsregler” som produserer en graf, hvor nodene er UUIDer representerer mobiltelefoner. Kantene fra den smittede sin node til andre noder representerer kontakter de har hatt, med tilhørende attributter. Det kan således være mange kanter mellom to noder. Det produseres to slike grafer:

- En graf basert på kun GPS-data.
- En graf basert på Bluetooth-data, augmentert med GPS-spor.

Kontaktene (kantene i grafen) er av ulik type, basert på hvilken graf som produseres. Felles for begge typene er at de inneholder varighet og en risikoscore. Kantene i GPS-grafen er basert utelukkende på kryssende baner (“trajectories”).

For Bluetooth-grafen finnes kontakter ved å søke opp alle Bluetooth-eventer som er innenfor tidsperioden og som inneholder den smittede sin UUID. Bluetooth-data om kontakt mellom to telefoner er lagret som en tidsserie med ulike signalstyrker (RSSI). Først segmenteres tidsserien i ulike intervaller, hvor et tidsopphold på 150 sekunder definere slutten på foregående segment og starten på nytt. Ved å benytte ulike terskelverdier for henholdsvis iOS og Android blir RSSI-styrken klassifisert som enten veldig nær (1 m), nær (2 m) og relativt nær (5 m). Hvis

et segment har en hovedvekt (ikke tidsvektet) av veldig nære signalstyrker settes hele segmentet til veldig nær. Tilsvarende for nær og relativt nær.

For å bøte på utfordringer med Bluetooth-teknologi i bakgrunnmodus henter man også ut alle veldig nære kontakter til den smittedes veldig nære kontakter innenfor det aktuelle tidsintervallet ved å se på GPS-baner. Man vil på den måten fange opp mulige ekstra kontakter som ikke ble oppdaget direkte av den smittedes mobiltelefon. For disse veldig nære kontaktene til de som er veldig nær en smittet blir klassifisert som nær (2 m) til den smittede. Risikoscoren vil så bli regnet ut som en vektet sum hvor ulike tidsintervall blir vektet med $1/m^2$ hvor $m$ er den klassifiserte avstanden.

Risikoscoren blir da integralet $\int 1/x(t)^2 \, dt$ hvor $x(t)$ er avstanden mellom smittet og kontakt.

Integralet løper fra kontakten startet til den slutter.

De to grafene blir så satt sammen til én graf. Det er summen av alle risikoscorene for alle kantene (kontaktene) som indikerer hvorvidt en kontakt av en smittet bør isoleres.

Etter valideringsperioden vil smittejegerne i FHI kun få tilgang til en listen med personnummer og telefonnummer til kontakter. I valideringsperioden er det behov for mye mer informasjon for å kunne validere kontaktalgoritmen. Denne utvidede rapporten er beskrevet i neste avsnitt.

Basert på dette lager man en rapport for indekspasienten (den som er smittet), som for hver av kontaktpersonene lister opp:

- Relatert indekspasient (for å kunne sammenligne med kommuneleger, manuell sporing)
- Telefonnummer
- Interessepunkter (Points Of Interest/POIs) fra OpenStreetMap
- Antall kontakter
- Søylediagram (oppsummering av all kontakt)
- Interessepunkter kontaktpersonen har vært på (Bluetooth)
- Totalt antall kontakter kontaktpersonen har hatt (Bluetooth)
- Hvilken risikokategori man plasserer kontaktpersonen i (Bluetooth)
    - En terskelbasert kategorisering av risiko-score.
- Risikoscore (Bluetooth)
- Akkumulert risikoscore (Bluetooth)
- Og akkumulert varighet basert (Bluetooth)
- Akkumulert risikoscore (GPS)
- Varighet basert (GPS)

- For hver dag i den etterspurte perioden:
    - For hver kontakt mellom indekspasienten og kontaktpersonen:
        - Dato

- HTML-kart (oppsummerer all kontakt mellom de to)
    - Dette er et 'interactive folium map': kart over bevegelsene til indekspasienten og kontakten tidsbegrenset til "kontaktperiode".
- Akkumulert risiko (Bluetooth)
- Akkumulert varighet (Bluetooth)
- Medianavstand (Bluetooth)
- Interessepunkter (Bluetooth)
- Akkumulert risiko (GPS)
- Akkumulert varighet (GPS)
- Medianavstand (GPS)
- Interessepunkter (GPS)

Denne informasjonen vil være til nytte i testperioden, da FHIs smittejegere også vil være bidragsytere til å validere automatisk kontaktsporing. Treff i Smittestopp sammenlignes med kontakter funnet ved manuell kontaktsporing. Valideringen av løsnigen består i å vurdere om kriteriegrunnlaget for varslene er gode, og om man finner flere kontakter ved bruk av Smittestopp appen. Man ønsker også å sammenligne gruppene: kontakt som er sporet manuelt, men ikke automatisk; kontakt som ikke er sporet manuelt, men automatisk; og kontakt som er sporet både manuelt og automatisk. Når funksjonaliteten er validert ønsker man å automatisere prosessen med kontaktsporing via Smittestopp.

Etter valideringsperioden er det planlagt å sende varsler til kontakter automatisk. Det er noe risiko for at man kan forstå hvem som har smitten en med denne type varsling. Dette er nevnt i appens personvernerklæring.

## Aggregerte data til analyse og forskning

Smittestopp har i hovedsak to formål: Kontaktsporing og overvåkning av kontakt og bevegelsesmønstre. Smittesporing er beskrevet ovenfor i varslingsløsningen. Overvåkningen av kontakt- og bevegelsesmønstre skal i hovedsak brukes for å

1) Identifisere hvordan ulike smitteverntiltak fra myndighetenes side påvirker sosial distanse
2) Input til FHI sine SEIR[3]-modeller som i hovedsak fremskriver ulike epidemiscenarier, og som er et viktig verktøy i beslutninger rundt åpning og stenging av samfunnet.
3) Og input til utvikling av agentbaserte smittespredningsmodeller og forskning for å kunne stille mer forberedt til neste pandemi.

For å tilfredsstille behovene over har FHI laget 5 aggregerte datasett med noen varianter. Hver av disse datasettene gir en telling på aggregert nivå av de underliggende dataene. Her er noen eksempler på felter i datasettene:

[3] https://www.fhi.no/en/id/infectious-diseases/coronavirus/coronavirus-modelling-at-the-niph-fhi/

- Antall aktive brukere av appen i bydel Østensjø i Oslo.
- Antall personer med alder 16 -19 som tar buss i 10 til 20 minutter 17. Mai i Oslo.
- Antall personer med alder 60 - 65 som har 10 eller flere nærkontakter 18. Mai i Åsane.

Detaljert informasjon om disse datasettene finnes i avsnittet om lagring, bruk og sletting av data.

For å bygge de ulike datasettene vil man først dele rådataene opp i ulike geografiske områder (polygoner) og tidsintervaller. Avhengig av hvilke hendelser man ønsker å telle vil man aggregere rådataene videre inn i ulike kategorier. I denne aggregeringen setter man på et tilfeldig filter som inkluderer en persons data med 90% sannsynlighet. Hvis et felt med tellinger er mindre enn $k$ (hvor $k$ stort sett er 5 eller 10) vil dataene ikke bli inkludert.

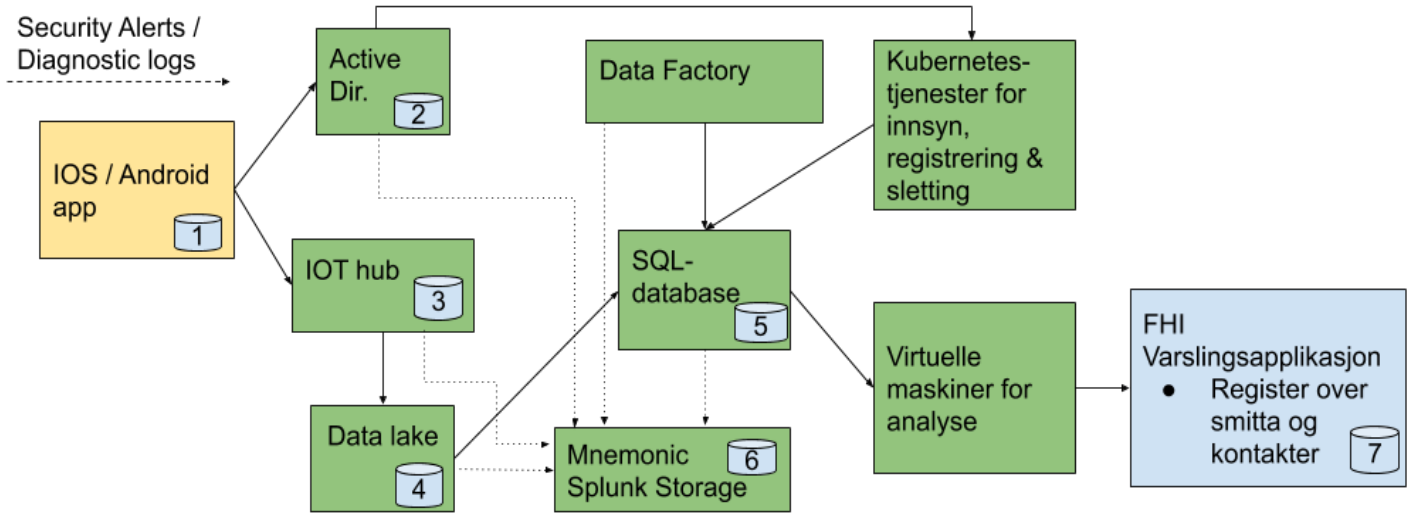
# Lagring, bruk og sletting av data

Enkelt forklart lagres data som samles fra GPS og Bluetooth først på brukerens mobiltelefon. Her lagres den ca en time før den sendes videre til skyløsningen. Hvis mobiltelefonen er slått av eller er uten internett vil den prøve å laste dette opp senere. Ved opplasting til skyløsningen blir data slettet fra brukerens mobiltelefon. I skyløsningen lagres data i 30 dager, før den slettes. Før 30 dager er gått, blir dataen aggregert, og lagres for FHIs forskning.

Brukeren kan når som helst slette dataene sine via appen. Hvis brukerens mobiltelefon ikke har lastet opp heartbeat på 7 dager, anses brukeren som inaktiv og dataene vurderes slettet. Dataen blir kun slettet hvis det ikke er lastet opp data fra en annen mobiltelefon, som er knyttet til samme telefonnummer i perioden. Brukeren vurderes som aktiv dersom det er lastet opp data fra minst en mobiltelefon koblet til brukerens telefonnummer de siste syv dagene.

Data som samles inn fra mobiltelefonen (GPS, Bluetooth, WiFi, akselerasjon, etc.) blir først mellomlagret på mobiltelefonen, før de blir sendt videre til “skyen”, som i praksis er et eksternt lager eid av Microsoft, som aksesseres via internett. Skyløsningen består av flere komponenter, men sentralt ligger en database som tillater strukturert lagring og oppslag av sensordata (herunder Bluetooth-kontakter og lokasjon).

Denne databasen brukes til smittesporingen, og for å lage de aggregerte datasettene.




**Figur 2.** Overordnet oversikt over Azure-tjenester i bruk, og hvor data lagres. Bokser med en nummerert sylinder representerer et datalager hvor data kan potensielt ligge i lang tid. Pilene indikerer en dataflyt fra-til. Stiplede linjer indikerer diagnose og sikkerhetsmeldinger fra individuelle tjenester.
**Lager:** 1: Data lagret lokalt hos brukeren. 2: Telefonnummer og assosiert bruker-ID. 3: Innkommende lokasjonsdata fra brukerne. 4: En midlertidig buffer for lagring av data i Azure Data Lake før det overføres til SQL-basen. 5: En SQL-database med lokasjonsdata og Bluetoothdata. 6: Et lager for Azure diagnose og sikkerhetslogger som blir lest av mnemonic og verktøyet Splunk via en Azure plugin. 7: FHIs sitt lokale datalager som benyttes i forbindelse med smittesporing og potensielt lager for forskningsdata.

Hvor, og hvem som har tilgang på data er et sentralt spørsmål. I korte trekk blir data lagret på brukernes mobiltelefon (1), i skyen (2-6) og hos FHI (7).

## Lagring på mobiltelefonen (1)

I figur 2 er data lagret på mobiltelefonen markert med “1”. Her lagres enhetens GPS-spor og Bluetooth-kontakter i en kryptert database. Disse blir periodisk sendt til “skyen”.

## Lagring i skyen (Azure - 2, 3 og 4)

Skyløsningen (grønne bokser i figur 2) utgjør appens sentrale datalager. Her havner alle data som mobiltelefonen innhenter og sender videre.

Skyløsningen består av flere orkestrerte komponenter, som figur 2 viser. Noen av disse komponentene lagrer data i ulike tidsperioder, enten som del av en pipeline eller som et endelig lager. Data som finnes i skyløsningen på et tidspunkt det gjennomføres backup på vil ende opp i backup-sett.

Active Directory (AD) (2) forbinder telefonnummer med en intern, unik identifikator (enhets-ID) som benyttes i videre datainnsamling. Denne forbindelsen er lagret her.

IoT Hub tar imot meldinger (med potensielt mange “hendelser”: lokasjoner, parede Bluetooth-enheter) fra mobiltelefoner, spesifikt GPS- og Bluetooth-data. Meldingene blir satt inn i en kø (3), før de sendes videre til Azure Data Lake (ADL) (4) der de blir lagret i en katalogstruktur angitt ved <år>/<måned>/<dag>.

En periodisk jobb kjørt i Azure Data Factory (ADF) kjører en lagret prosedyre på SQL-serveren (5), som henter rådata fra ADL i form av JSON-filer, og setter disse inn i en forberedelses-tabell i SQL-serveren (“staging”). Staging-tabellen inneholder også metadata om den mottatte meldingen, som telefontype, appversjon, OS-versjon også videre. Den samme ADF-jobben gjør også oppdeling av meldinger til individuelle hendelser, og innsetting fra staging til separate tabeller for Bluetooth og GPS-hendelser. Tilkoblingen til ADL fra SQL-serveren autoriseres med en SAS-token lagret i SQL (‘stored credential’). *Detaljert informasjon om SQL-tabellene finner du i appendiks D.*

Generelt benyttes det i de ulike datalagrene i figur 2 rollebasert tilgangsstyring, som definerer hvilke personer (eller maskiner) som kan lese forskjellige data basert på tildelte roller. Kort oppsummert er løsningen administrert av NHN, og data kan dermed leses og skrives av administratorer hos NHN. Simula er tildelt roller som bidragsytere og organisasjonen har representanter med tilgang til å endre og lese i store deler av løsningen, men ikke innhold i databasen.

Hele databasen er kryptert "at rest" i Azure med en symmetrisk kryptering.

Databasen er låst ned spesielt med egne brukere, der hver bruker gis tilgang til å kjøre kun spesifikke forhåndsdefinerte spørringer for uthenting av data. I produksjon er det bare brukere tilknyttet den automatiserte varslingstjenesten skal ha tilgang til å utføre spørringer. Administratorer av databasen er unntatt fra disse begrensningene.

Detaljer rundt de ulike rollene og hvilke roller som har tilgang til hvilke tjenester finnes i tabellene i appendiks D.

## Data til mnemonic (6)

Teknisk-diagnostisk data og sikkerhetsvarsler lagres her, og tilgjengeliggjøres for mnemonic i deres SOC.

## Lagring i sikker sone (FHI 7)

FHI lagrer oppslagsdata for indekspasienter som ikke finnes i smittesporinsgsystemet. Dersom et oppslag mot Simula med en indekspasient gir positivt resultat lagres følgende i en lokal database hos FHI (detaljer rundt disse attributtene finner du i beskrivelse av varslingsløsningen over):

For alle kontakter en indekspasient har hatt:

- Indekspasienten
- Telefonnummer til indekspasienten (kryptert)
- Interessepunkter kontakten har vært i nærheten av
- Antall kontakter
- Risikokategori
- Søylediagram
- Bluetooth Akkumulert Risikoscore
- Bluetooth Akkumulert Varighet
- GPS Akkumulert Risikoscore.
- GPS Akkumulert Varighet
- Oppsummert Plot med Detaljer i Html
- Per dag i forespurt periode:
    - Dato
    - GPS Akkumulert Risiko
    - GPS Akkumulert Varighet
    - GPS Medianavstand
    - GPS Interessepunkter
    - Bluetooth Akkumulert Risiko
    - Bluetooth Akkumulert Varighet
    - Bluetooth Medianavstand
    - Bluetooth Interessepunkter

## Beskrivelse av slettemekanismene

Det er tre mekanismer som vil utløse ulike automatisk sletting:

1) Bruker ber om at sine data slettes
2) Dataene som er lagret er mer en 30 dager gammel
3) En bruker har ikke sendt heartbeat i løpet av  syv dager

Lokale data på telefonen blir slettet så rask de er lastet opp til skyen via Azure IoT-hub, appen blir avinstallert, eller brukeren ber om å få slette sine data. Man vil også slette lokale app-data hvis en opplasting har mislykkes 40 ganger.

Data lagret i IoT-hub-ens kø blir fjernet etterhvert som meldinger blir konsumert og lagret i Azure Data Lake.

Hvis brukeren ber om å slette sine data vil brukeren bli markert for sletting, og vil bli slettet i den påfølgende nattlige slette-jobben.

Dette inkluderer:

- Enhets-ID fra IoT-hub, som medfører at denne ikke lenger kunne motta data for denne brukeren.

- Oppføring i Active Directory: telefonnummer og enhets-ID, og brukeren som lagrer telefonnummeret.
- Data fra SQL-tabellene.
- Data fra Data Lake.

Lagrede indekspasienter i FHI sin lokale database (7) slettes etter 30 dager.

## Avidentifisering og aggregering av data

Koden for å produsere datasett er per nå ikke ferdig. Gruppen er kjent med følgende om de planlagte datasettene:
Det er planlagt å produsere 5 primære datasett ("tabeller") ved å benytte spørringer mot den sentrale databasen som er beskrevet tidligere. Det skal produseres flere varianter av disse med ulike parametere for aggregering. Datasettene skal produseres ved å gjøre tilfeldige trekk fra sentraldatabasen. Det er et mål om å oppnå k-anonymitet i datasettene.

I denne sammenhengen betyr k-anonymitet at hvis en telling er mindre enn k, vil feltet bli slettet. Hvis f.eks. kun 2 personer i et alderssegment har vært på bussholdeplasser i Bodø den 28. mai 2020, så kastes radene. Litt avhengig av hvordan k-anonymiseringsalgoritmen implementeres vil man kunne oppnå "differential privacy" ved at man først tar et tilfeldig uttrekk etterfulgt av k-anonymisering. Differential privacy kan gi brukeren en garanti for at brukerens data bidrag ikke vil ha en konsekvens for brukeren. Denne garantien er selvfølgelig avhengig av hvor stor grad de aggregerte dataene er differentially private.

Det finst 3 ulike geografiske regioner som de ulike datasettene skal aggregeres på:
- Regioner med flere enn 200 brukere.
- Regioner med flere enn 2000 brukere.
- Regioner med flere enn 50.000 brukere.

Brukerne vil bli tildelt en alderskategorier, slik at det vil være mulig å gjøre tellinger langs denne dimensjonen også. Mekanismen for å hente ut alder som brukeren registrerer ved onboarding og lagres lokalt på telefoner er enda ikke kjent.

Data aggregeres også på tid:
- Timer
- 3-timersintervall
- Dager

Det er forsøkt å oppnå en balanse mellom tid og geografisk aggregering for å ivareta en grad av anonymitet, samtidig som datasettene beholder sin praktiske nytteverdi. Det er videre lagt opp til 2 varianter av hvert datasett, ett som generelt har lavere granularitet (flere detaljer), og ett som har høyere (færre detaljer). Disse to datasettene skal ha ulike tidsfrister for sletting og ulik tilgangskontroll.

Datasettene som skal produseres, og deres formål:

1. Hjemme: Antall brukere som sover i ulike geografiske områder per dag.
2. Stedskategorier/POIs og transportmidler: Hvor lang tid som brukes i ulike stedskategorier og transportmidler. Typiske stedskategorier er dagligvarebutikker, skoler og restauranter. Typiske transportmidler er å gå, sykler eller kjøre.
3. Tetthet i offentlige rom: antall mennesker i ulike typer offentlige rom.
4. Kontakttyper: Detaljer rundt kontakt-typer: hvor de skjer og hvor lenge de varte.
5. Sosial distansering: Hvor mange kontakter som skjer innenfor ulike geografiske regioner og i ulike tidsintervall.
    a. Antall kontakter i store regioner per time.
    b. Antall kontakter gruppert på kontaktlengde per dag.
    c. Antall kontakter gruppert på total reiselengde per dag eller time.
    d. Antall kontakter gruppert på ulike reiser per dag.

## Logging av tilgang til data

Smittejegere har anledning til å se kontaktsporingsdata i listevisning (telefonnummer, personnummer og antall kontakter per indekspasient). Visning i lister blir ikke logget før smittejegeren velger å hente ut dekryptert telefonnummer eller fødselsnummer for kontakten.

Det som da blir logget er brukernavn for personen som gjør oppslaget, hva som ble dekryptert, og en forklaring på hvorfor. I skrivende stund er sistnevnte forklaring hardkodet, men det er ikke avgjort om denne bør endres til å oppgis av brukeren i det konkrete tilfellet eller oppgis av applikasjonslogikk avhengig av kontekst.

FHI vil sannsynligvis identifisere sine brukere i innsynslogger som “Saksbehandler #N” eller tilsvarende, da dette er i tråd med praksis for å ivareta personvernet til deres ansatte i eksisterende innsynsløsninger.

Tilgangsloggene er lagret i SQL-databasen i Azure og aksesseres via API-kall. Auditlogging er påskrudd i SQL-basen, men uthenting av disse loggene er ikke automatisert.
Teknisk-diagnostisk data og sikkerhetsvarsler fra de ulike Azure-tjenestene blir skrevet til et datalager som aksesseres av mnemonic.

# Data og Personvern

## Formål

Den norske appen avviker fra andre løsninger for å adressere COVID19-pandemien ved at den har to formål: smittesporing og evaluering av effekt av smitteverntiltak i forbindelse med COVID-19. Dette betyr at man ikke direkte kan sammenligne hva denne appen samler inn med andre apper som bare dekker ett av formålene. De to forskjellige formålene har ikke nødvendigvis behov for samme type data. For å kunne diskutere dette i en personvernkontekst må vi se på disse hver for seg.

## Kontaktsporing

Hensikten med kontaktsporing er å finne personer som har vært utsatt for smitte fordi de har oppholdt seg i nærheten av personer som blir identifisert som smittet. For å få til dette brukes GPS-koordinater og Bluetooth-kontakt mellom mobiltelefoner som har installert appen.

**Dataminimering**
Jo flere begrensninger man legger i både innsamling og lagring av data, jo lavere risiko utgjør det for brukerens personvern.

Teoretisk sett, kan man oppnå kontaktsporing ved at man lagrer Bluetooth-ID'en til alle mobiltelefoner man har vært i nærheten av, lokalt på mobilen. Ved registrert smitte vil man så trenge å dele informasjon om hvilken enhets-ID er smittet. Dette kan enten gjøres ved at den smittede laster opp sin enhets-ID eller ved at den smittede laster opp alle enhets-IDer brukeren har møtt.

Praktisk sett, vil verken Bluetooth eller GPS gi perfekte data. Bluetooth kan ha problemer med manglende kontakt pga sovende mobiltelefoner og varierende signalstyrke. GPS har som regel en nøyaktighet på 3 - 10 meter og fungerer best utendørs. Med sentral lagring av disse dataene kan man kompensere noe for disse svakhetene, ved å korrelere forskjellige datakilder.

**Kunne dette vært løst på en annen måte med mindre inngripen i personvernet?**
Ved prosjektstart var det få gode alternativ til sentral lagring, hvor man samtidig ville oppnå god nok kvalitet på data for kontaktsporing. I ettertid har Google og Apple kunngjort at de vil lansere utvidelser av Bluetooth-APIer for kontaktsporing, som kan redusere behovet for sentral lagring. Dermed kan lagring av data lokalt på mobiltelefonen kan være et realistisk alternativ, men det er per i dag ukjent hvor gode resultater man kan få, uten å gjøre kompensering med sentralt lagrede data.

**Bør man vurdere å samle inn data fra andre datakilder?**
Man kan forestille seg at man kunne ha brukt annen type datakilder. Et eksempel på dette kan være sensorer som temperatur, lys, trykk og NFC. En løsning kunne ha brukt dette for å supplere den informasjonen som allerede samles inn. Det finnes ikke store undersøkelser om effekten av å bruke GPS og Bluetooth til smittesporing enda. Dermed er det vanskelig å vurdere om det er behov for flere datakilder.

## Analyse av hvordan smitteverntiltak påvirker bevegelsesmønsteret i befolkningen.

FHI ønsker å kunne se hvordan smitteverntiltak mot spredning av COVID-19 påvirker bevegelsesmønsteret i befolkningen. Det ikke behov for like nøyaktig data til aggregerte datasett som ved kontaktsporing.

**Dataminimering**
Det er behov for GPS-data for å kunne vite hvordan befolkningen beveger seg. Bluetooth-data blir brukt for å telle antall nærkontakter mellom personer på ulike stedskategorier.

**Kunne dette vært løst på en annen måte med mindre inngripen i personvernet?**
For å kunne svare på disse spørsmålene må man vite hvor personer oppholder seg. Per i dag samles all lokasjonsdata og lastes opp. Det ville vært mulig å aggregere GPS og Bluetooth-data på mobiltelefonen og legge til støy med visse statistiske egenskaper, som ledd i en "differential-privacy"-tilnærming, før de blir lastet opp til en sentral server uten at dette vil forringe kvaliteten til datasettet.

**Er dataen som samles inn tilstrekkelig?**
Det hentes per i dag ikke inn data om brukerens fødselsår eller aldersgruppe. Hvis vi ser bort fra denne mangelen, vil dataen som samles inn dekke behovet for å kunne lage de datasettene FHI ønsker.

## Andre betraktninger om personvern

Informasjonen som samles via Smittestopp-appen kan oppleves som et vesentlig inngrep i personvernet. Lagring av personlig data skal alltid ledsages av en vurdering av inngrepet i personvernet, holdt opp mot forventet effekt. Her vil det i et forskningsperspektiv være bedre med mer informasjon, mens det i et personvernsperspektiv vil være motsatt.

Smittestopp appen er frivillig å bruke i Norge. All behandling av persondata i Norge er basert på GDPR. I tillegg finnes det for appen Smittestopp *Forskrift om digital smittesporing og epidemikontroll i anledning utbrudd av Covid-19*[4], som ble fremmet av HOD og har ikrafttredelse

[4] https://lovdata.no/dokument/SF/forskrift/2020-03-27-475

27.03.2020 – 01.12.2020. Forskriften er hjemlet i *Lov om vern mot smittsomme sykdommer [smittevernloven]*. Denne forskriften sier blant annet at all data skal slettes etter 30 dager.

Mulighet for deling av lokasjonsdata blir begrenset i forskriften. I og med at Bluetooth er en kommunikasjonsprotokoll for bruk mellom bevegelige enheter, anser vi ikke Bluetooth-data som lokasjonsdata. Dermed er det vanskelig å se at forskriftens begrensninger mot å dele lokasjonsdata også gjelder Bluetooth-data.

# Evaluering

## Funn

### App

Appen starter innhenting av bruksanalyse- og telemetri-data første gang appen åpnes. Dette sendes til AppCenter ved de fleste interaksjoner med appen. Data inkluderer bla. App-versjon, teleoperatør, språkinnstillinger, telefonmodell og -produsent, skjermstørrelse, OS-type og -versjon, tidssone, samt en unik identifikator. Dette er ikke nevnt i personvernerklæringen, og behandlingsgrunnlaget er dermed uklart. Brukeren burde ha mulighet til å velge vekk denne datainnsamlingen via onboarding-prosessen.

I vår forrige rapport presenterte vi et funn: Statisk ID som ble brukt ved informasjonsutveksling mellom Bluetooth enheter. Dette er noe som Simula jobber med å utbedre. Planen er å kryptere ID og timestamp med en offentlig nøkkel, deretter dekryptere dem med en privat nøkkel i skyen. Denne løsningen var ikke ferdig per 18.05. 2020, men er planlagt produksjonssatt i løpet av uke 21.

### Backend

Skyløsningens eneste form for sessionshåndtering skjer ved hjelp av en en evigvarende connection string som brukes mellom enhet og IoT-Hub. Det betyr at det er mulig å laste opp data på vegne av andre brukere dersom man har deres connection string. Ettersom denne er statisk krever dette kun at én legitim overføring mellom enhet og server snappes opp av noen andre.

Funksjonaliteten som brukes til å knytte telefonnummer mot enhets-ID i skyen er en såkalt *“preview feature”*, som leverandøren selv sier[5] man *ikke skal* bruke til å prosessere personopplysninger eller annen data som holdes til høyere compliance-krav.

### Innsynsløsning

Sletting av opplastede data fra mobiltelefonen, sletter også lagret data om hvem som har sett på informasjonen til brukeren. Sletting av denne dataen burde vært frikoblet og gjort direkte i innsynsløsningen.

Brukere kan ikke se egne Bluetooth treff i innsynsløsningen.

[5] https://azure.microsoft.com/is-is/support/legal/preview-supplemental-terms/

Innsynsløsningen mangler støtte for at brukerne kan se FHIs tilgangslogg. Informasjonen blir lagret i systemet, men ikke gjort tilgjengelig for brukeren. NHN har snakket om å implementere dette.

### Varslingsløsning

Koden som finner kontakter har dårlig kodekvalitet. I enkelte tilfeller er den vanskelig å lese, og er såpass komplisert og uoversiktlig at den ikke kan anses som enkelt vedlikeholdbar. Spesielt er dette fremtredende i definisjonen av en Bluetooth nærkontakt. Her er det svakheter både i implementasjon (uleselig og ikke vedlikeholdbar kode) og i metodikk. Overordnet kan man si at metoden for å avgjøre om en kontakt er en nærkontakt eller ikke introduserer unødvendige feilkilder. Å konvertere en tidsrekke med signalstyrker til en eller flere nærkontakter er ikke rett frem. Men det virker mot sin hensikt å ikke ta med sendt signalstyrke (tx_power) og glatte ut tidspunktene målingen blir gjort innenfor en avgrenset kontakt.

Metoden for å oversette signalstyrke til avstand er også problematisk. Her har man valgt å kategorisere hvert signal til enten å være veldig nær (1 m), nær (2 m) eller relativt nær (5 m). Andelen veldig nære, nære og relativt nære avgjør om hele kontakten klassifiseres som veldig nær, nær eller relativt nær. Kategoriseringen av hele kontakten brukes så inn i beregningen av en risikoscore. Ettersom risikoscoren ikke er lineær med tanke på avstand (signalstyrke) vil denne måten å regne på kunne føre til inkonsistente klassifiseringer. Ved å i stedet beregne en riskoscore basert på tidsrekken med signalstyrken først, for så å klassifisere avstand på signalene, vil man antageligvis få færre “inkonsistente” klassifiseringer.

SMS er ikke en sikker kommunikasjonskanal, og det er relativt enkelt å forfalske disse. En måte å avhjelpe dette på kunne vært å bruke en annen kommunikasjonskanal, f.eks. DigiPost.

## Diskusjon

Ulike valg tatt under utviklingen speiler prioriteringene som har blitt foretatt under utvikling. Herunder finner vi en rekke valg-akser, som utviklingstid, sikkerhet, behov, brukervennlighet og personvern. Å fatte valg som prioriterer en av dimensjonene, kan gjøre det vanskelig å oppnå andre moment.

Det er viktig å ta med seg at dette ikke er “ja/nei”-dimensjoner, men at det finnes grader. For eksempel har utviklingstid sterk innvirkning på valg som har med personvern å gjøre; Mer avanserte former for personvern krever ofte en lenger utvikling- og verifiseringsperiode og vice versa. Samtidig kan det være langt enklere å ivareta sikkerheten i mindre kompleks kode.

### Flere formål

Ved å kombinere de to formålene *smittesporing* og *vurdering av effekt av smitteverntiltak* i samme app forventes det at flere vil laste ned appen, enn det som ville lastet ned en app som

bare deler data for forskning/analyse. Man risikerer likevel at potensielle brukere som ville ha lastet ned en ren smittesporingsapp, ikke er komfortable med å laste ned den kombinerte appen.

Andelen av befolkningen som må laste ned appen er veldig forskjellig for de to formålene. For å oppnå god effekt ved smittesporing må godt over halvparten av befolkningen ha appen. For å få gode datasett for forskning/analyse kan så lite som 10% utbredelse være nok.

Å oppnå 60% utbredelse av kontaktsporing er et ambisiøst mål. Hvis man lot brukerene selv bestemme om de vil dele data til forskning/analyse, ville nok flere valgt å laste ned appen. I og med at utbredelseskravene for å oppnå formålet er så mye mindre for forskning/analyse, tror vi heller ikke at dette formålet ville blitt negativt påvirket med et slikt valg.

## Sentral vs. lokal lagring

Samling av dataen sentralt muliggjør kontaktsporing på data med høyere kvalitet. Måten man oppnår dette er å bruke data lastet opp av bruker A for å få et mer komplett bilde av bruker B. Dette kan kompensere for svakheter i måten Bluetooth er implementert. I en test- og valideringsfase vil det være lettere å justere algoritmen for smittesporing i en sentral løsning. System med sentral lagring er enklere å få til å virke som forventet.

I en situasjon hvor man har som mål å tilgjengeliggjøre løsningen så fort som mulig, vil det å gå for en sentralisert løsning redusere risiko for feil og time-to-market. Samtidig trenger ikke dette å bety at man ikke over tid kan gå over til en mer distribuert løsning. Ved å være mer selektiv i forhold til hva man laster opp, og ved å gjøre det mulig å skille smittesporing fra datainnsamling til andre formål, vil man kunne redusere nødvendig data innsamlet for de som kun tillater smittesporing, og også redusere risikoen knyttet til sentral lagring.

Med sentral lagring vil det ha en stor negativ konsekvens dersom det forekommer et datainnbrudd eller en annen form for misbruk (som lekkasje, utro tjenere eller feil). Så mye persondata på et sted vil i seg selv vil være et yndet mål for trusselaktører, som APTer (Advanced Persistent Threats), og disse jobber som kjent[6] målrettet mot aktører og systemer involvert i COVID-19-respons.

## Innsamling av lokasjonsdata (GPS)

Innsamling av lokasjonsdata gjør det mulig å lage datasett for vurdering av effekt av smitteverntiltak. Det gjør det også mulig å bruke stedskategori og transportmiddel som faktorer når man regner på kontaktrisiko. Lokasjonsdata gir også flere muligheter for å kompensere for implementasjonssvakheter i Bluetooth.

[6] https://www.us-cert.gov/ncas/alerts/AA20126A

Selv om man har åpenbar nytte av lokasjonsdata, vil innsamling av lokasjonsdata være en større inngripen i personvernet enn å bare samle data fra Bluetooth. Dette påvirker graden av inngripen i personvernet og kan potensielt også påvirke adopsjonen av appen.

## Å bruke Google/Apples nye API

De nye APIene fra Google og Apple for å gjøre smittesporing forventes å gi en betydelig forbedring i hvor ofte Bluetoothkontakt mellom to mobiltelefoner som er i nærheten av hverandre registreres. Man vil dermed kunne få mer komplette data over enheter man har møtt på. Dette vil gjøre at mulighetene for effektiv kontaktsporing kun basert på Bluetooth vil øke betraktelig.

Google/Apples API krever både distribuert løsning og at appen ikke skal laste opp lokasjonsdata. Dette gjør at en slik endret app ikke vil kunne produsere data til de ønskede datasettene. Man vil heller ikke kunne forbedre smittesporingen ved å bruke lokasjonsdata.

Google/Apples løsning er ikke kompatibel med eksisterende app og vil således enten kreve en betydelig nedskalering av ambisjonsnivået rundt datainnhentingen i eksisterende app eller en ekstra app som kun gjør kontaktsporing.

## K-anonymitet vs. differential privacy i produksjon av “anonyme” aggregerte data

Det er ikke mulig å skape verdifulle datasett om personer uten at eksistensen av datasettet innebærer en personvernkonsekvens. Det finnes flere ulike tiltak for å redusere personvernrisiko til de registrerte. Det finnes effektive metoder for å hente ut presis statistikk om et datasett på en slik måte at statistikken i seg selv har minimal personvernrisiko. Den eneste kjente metoden som kan garantere en øvre skranke for tap av personvern knyttet til skapelsen/offentliggjøringen av de aggregerte datasettene er *differential privacy*. $k$-anonymitet, på den andre siden, garanterer kun at èn brukers bidrag inn i statistikken/datasettet ikke kan skilles fra mindre enn $k-1$ andre individer. Det stilles ingen garantier for at det ikke er mulig å redusere $k$ betraktelig med tilleggsinformasjon og på denne måten utlede personlig informasjon om enkeltindivider.

**Eksisterende oppsett**
I byggingen av de aggregerte datasettene brukes i dag tre sikkerhetsmekanismer:

- Tilfeldig uttrekk (x% sjanse for at en registrert blir telt)
- k-anonymitet (datasettet inkluderer ikke tellinger på under k-individer)
- Tilgangskontroll (datasett med lav $k$ vil være tilgjengelig for færre mennesker over en kortere periode)

Fordelen med $k$-anonymitet er at det er enkelt å implementere. Tilfeldig uttrekk er et lite steg opp i vanskelighetsgrad, men ikke nevneverdig vanskeligere. Tilfeldig uttrekk gitt at

tilfeldigheten er høy nok vil gi høy beskyttelse, men felles for både $k$-anonymitet og tilfeldig uttrekk er at man sletter data og dermed forringer kvaliteten. I en differential privacy løsning vil man legge til støy på individnivå. Dette gjør det like vanskelig å bevise at et individ er registrert, som med tilfeldig uttrekk, men siden man ikke faktisk sletter noe vil det være mulig å beholde mer av verdien i datasettet.

De aggregerte datasettene som FHI ønsker er veldig godt egnet for et differential privacy oppsett. Dette vil kreve en litt mer avansert implementasjon, men siden personvernet og ikke minst verdien av dataene vil bli bedre burde man absolutt vurdere en slik løsning.

Det er også mulig å lage differentially private design hvor rådata (data uten støy) ligger på mobiltelefonen slik at det kun er de differentially private datasettene som ligger sentralt lagret.

## Kontaktsporing eller direkte prediksjon av smitte

Et av formålene med Smittestopp er å kunne gjøre mer presise sosial distanse-, og selvisoleringstiltak. Uten Smittestopp er man i hovedsak begrenset til manuell smittesporing, omfattende karanteneregler og aktivitets- og næringsforbud som tiltak for å redusere smitte. Med en gang man velger å samle inn lokasjonsdata for å gjøre tiltakene mer presise, har man også mulighet til å tenke nytt. Å prøve å predikere hvem som er smittet er et slikt eksempel.

Formålet er det samme; å sette riktig person i karantene, å hindre smittespredning med lavest mulig samfunnskostnad. Nøyaktigheten til disse prediksjonene vil i starten være dårlige, men fordi man vil ha en klart definert måte å teste om man gjettet rett eller ei (positiv/negativt resultat ved testing), vil man kunne trene opp systemet. Eksempel på signal som kunne være spennende å bruke er:

- minutter brukt på kollektivtransport,
- opphold på offentlige steder,
- nærkontakter med ukjent smittestatus

Ved å rangere brukere etter smitterisiko vil man kunne teste mer effektivt, samtidig som man samler data som kan brukes til å øke presisjonen.

En slik fremgangsmåte vil kreve at man bygger historiske datasett på de som blir testet og vil ikke være mulig å gjøre under eksisterende forskrift. De historiske datasettene kan bygges ved eksplisitt samtykke eller en form for datadonasjon. Det er fullt mulig å se for seg et mindre inngripende design enn eksisterende løsning, men et slikt system vil nødvendigvis være mer inngripende enn en maksimalt distribuert kontaktsporingsløsning tilsvarende PACT/MIT, Google/Apple, o.l.

## Åpen kildekode

Ved å ha åpen kildekode (open source) gir det andre anledning til å bidra, for eksempel i form av å oppdage og eventuelt utbedre feil i koden. Dette vil i et langsiktig tidsperspektiv kunne bidra til å oppnå sikrere programvare. For prosjekt som behandler personopplysninger tilbyr det også brukerne direkte innsikt i prosessering av disse. I et kortere tidsperspektiv vil åpen kildekode potensielt medføre en risiko for at sårbarheter blir oppdaget og utnyttet før open source miljøet rekker å oppdage og fikse disse.

Ettersom ekspertgruppen ikke vil eksistere etter 20. mai, vil åpen kildekode gjøre det mulig for open source miljøet å fortsette den jobben ekspertgruppen har gjort til nå.

## Utviklingsløpet

Grunnleggende finnes det to forskjellige typer utviklingsprosjekt: Prosjekt der man lager noe man allerede vet er mulig, og prosjekt der man underveis finner ut om noe er mulig. I den siste typen prosjekt er det vanskelig å vite om det er mulig å oppnå det man prøver på, før man har begynt å samle inn data. Med appen Smittestopp ønsker man å løse to forskjellige problem i samme app.

Det ene problemet er å samle data for å kunne vurdere hvordan myndighetenes smitteverntiltak påvirker bevegelsesmønsteret og kontakthyppigheten til befolkningen. Å laste opp lokasjonsdata fra en app har vært løst før. Det er lav risiko ved en slik implementasjon, da det er få ukjente faktorer.

Det å bruke mobiltelefon for å oppdage at noen har vært utsatt for smitte, er derimot en helt ny problemstilling. Det finnes artikler og teoretiske betraktninger om hvordan dette kan gjøres, men det har ikke dette vært implementert og testet i stor skala. Når dette i tillegg er basert på teknologi (Bluetooth og GPS) som verken er laget for formålet eller er helt nøyaktig, så er det uklart om man vil kunne oppnå praktisk fungerende kontaktsporing.

Det har vært stor fokus på kort leveringstid fordi koronasituasjonen var uavklart i prosjektets tidlige faser. Samtidig er det ikke nødvendigvis slik at det man forsøker å oppnå er teknisk eller praktisk mulig. Sett i lys av dette kan det være naturlig å holde mulighetene åpne: F. eks ved å bruke sentraliserte databaser, innhente data fra alle sensorer, etc. Når man nå har bygd “noe som fungerer” er det på tide å vurdere hvordan man kan oppnå ikke-funksjonelle mål.

For Smittestopp er de viktigste ikke-funksjonelle målene sikkerhet og personvern. Utviklingen går så over i en fase der det handler om å velge løsninger som er “sikker nok”, “rask nok”, “funksjonell nok” og “personvernsbevarende nok”, der “nok” vil ha ulik valør blant ulike mennesker og i ulike miljøer.

For å ivareta personvernet er det viktig at det strammes inn på innsamling og bruk av personlig data så tidlig som mulig i utviklingsprosessen. Man kan argumentere for at man fremdeles er så tidlig i utviklingsløpet at man ikke har kommet til et punkt der det er naturlig å starte innsstrammingen. I så fall ville man forventet at man bare samler inn data som hjelper med å forbedre appen. Man burde ikke åpne for å laste ned appen utenfor testkommuner før man er ferdig med valideringsfasen.

Utvikling av appen Smittestopp foregår i en situasjon som forandrer seg fort. Man kan se for seg at det kan komme løsninger eller resultat fra andre initiativ som man kan benytte seg av. Smitteutbredningen i Norge forandrer seg også. Det er naturlig at det man så på som den beste løsningen ved prosjektstart, ikke lenger er den beste løsningen når situasjonen forandrer seg.

Når man driver med utvikling av et nytt produkt eller konsept er det viktig at man regelmessig evaluerer om man faktisk vil nå det målet man har satt seg med tilnærmingen man har valgt. Det er viktig å ha milepæler med testkriterier som hjelper å oppdage så tidlig som mulig om man trenger å justere kursen.

## Oppsummering

I en test- og utviklingsfase hvor man hurtigst mulig ønsker å validere hvordan mobiltelefonbasert kontaktsporing kan implementeres, kan man bruke testdata som er frivillig donert av beta brukere. Når man har funnet en funksjonell løsning, må man sørge for å ta hensyn til ikke-funksjonelle krav som sikkert og personvern, før man starter på neste fase av utviklingsløpet.

At ikke-funksjonelle krav ikke enda var ivaretatt var grunnen til at vi anbefalte å ikke ta appen ut av testfasen i vår midlertidige rapport. Vi mente at den ikke var klar i forhold til sikkerhet. De tekniske og arkitektoniske valgene i løsningen slik den foreligger i dag har betydelige personvernskonsekvenser. Vi kan ikke se at disse er tilstrekkelig hensyntatt i utformingen av løsningen. Det er fremdeles viktige betraktninger rundt personvern som bør hensyntas før lansering.

# Konklusjon

Ekspertgruppens mandat er å levere en åpen rapport med en overordnet vurdering av om sikkerhet og personvern er forsvarlig ivaretatt.

Vurderingen av appen Smittestopp er basert på hvordan selve appen og systemene rundt ser ut per 18. mai 2020. Utvikling av programvare er en levende prosess og våre betraktninger vil ikke nødvendigvis lenger være gjeldende dersom det gjøres endringer av større karakter eller hvis man ikke klarer å oppnå ønsket effekt i kontaktsporingen.

**Sikkerhet**
Vi mener at sikkerheten kan bli forsvarlig ivaretatt, men dette forutsetter at den planlagte krypteringen av Bluetooth IDer blir satt i produksjon. Dermed er sikkerheten ikke forsvarlig ivaretatt per i dag.

**Personvern**
Som teknologer kan vi vurdere personvernkonsekvenser når vi sammenligner funksjonelt ekvivalente løsninger. Vi kan derimot ikke gjøre en avveining mellom løsninger med forskjellig nytteverdi. Vi mener det finnes mange muligheter for å redusere mengden data som lagres i eksisterende løsning eller ved å bytte ut deler eller hele systemet med andre funksjonelt ekvivalente løsninger. Derfor mener vi at personvernet ikke er forsvarlig ivaretatt per i dag.

# Anbefalinger

**Forskrift og anonymisering**
Vår tolkning av forskriften som den foreligger i dag er at det ikke vil være mulig å lagre de ønskede datasettene utover 30 dager. Vi antar at datasettene som produseres er innenfor opprinnelige intensjoner bak forskriften, men fordi det finnes en sletteklausul, og ingen referanse til avidentifiserte aggregerte data eller “differential private” data med lengre lagringstid enn 30 dager tror vi ikke det vil være mulig å gjøre analyse på effekt av virkemidler via data fra Smittestopp uten å gå på akkord med forskriften slik den foreligger i skrivende stund. Vi anbefaler at forskriften blir oppdatert i henhold til hensikten. Dvs at man referer til lagring av avidentifiserte aggregerte data utover 30 dager, i tillegg til den eksisterende slettingen av rådata.

**Dele opp formålene og gjøre det mulig å velge å delta kun for det ene**
Ved å gi innbyggerne muligheten til å akseptere kun ett av formålene, vil man gi brukerne større grad av autonomi. Sannsynligvis er flere innbyggere villige til å delta i kontaktsporing enn å delta med data til analyse/forskning. I og med at utbredelsen av deltakere i kontaktsporing er så viktig, kan det å dele opp formålene være et godt virkemiddel for å øke utbredelsen av appen og samtidig ivareta brukernes interesser.

**Fjern all data man ikke trenger**
En periodisk jobb som analyserer og fjerner data som ikke er nødvendig for kontaktsporing, herunder sletting av alle GPS-data eldre enn 15-16 dager og aggressiv beskjæring av databasen for GPS-baner som ikke krysser andre GPS-baner vil øke dataminimeringen uten å gå på akkord med funksjon.

**Implementere differential privacy der det hentes ut data til datasett**
Ved å benytte differential privacy når man henter ut data til datasettene oppnår man både at man har mer komplette data, og at man kan kvantifisere hvor god “anonymiseringen” av data er. Sampling etterfulgt av “trygge” k-anonymiserings algoritmer tilfredsstiller differential privacy.

Samtidig er det relevant å påpeke at det også er mulig å redusere personvernkonsekvensen, og samtidig øke nøyaktigheten i datasettet, med å velge en ren differential privacy-algoritme.

**Når valideringsfasen er over (når det finnes en stabil algoritme for kontaktsporing) bør man vurdere en mer distribuert løsning**

Når man har en stabil algoritme for kontaktsporing, bør man vurdere om en mer distribuert løsning er mulig. En helt eller delvis distribuert implementasjon vil kunne være mindre inngripende i personvernet og samtidig øke utbredelsen av kontaktsporingsappen.

**Gå over til en distribuert modell for innsamling av data**

I en modell der man gjør kontaktsporing uten å laste opp lokasjonsdata og Bluetooth-data sentralt, kan man også produsere datasettene uten å laste opp rådata. I stedet kan man lage datasettene ved å gjøre opplasting av data med bruk av differential privacy-filtre. Dette vil bidra til å ytterligere redusere inngrepet i personvernet.

**Tilgjengeliggjør så mye kode som mulig som åpen kildekode**

Vi ser ingen grunn til at koden som generer datasettene ikke kan bli tilgjengeliggjort umiddelbart. Dette er for å gi brukerne reell og direkte innsikt i hvordan sine data behandles. App-koden og backend-koden bør også bli gjort tilgjengelig så fort som mulig, etter å ha fjernet sensitive API nøkler og passord fra koden og Git historien. Koden som brukes til kontaktsporingen bør også åpnes opp, med mindre det her er kommersielle hensyn som gjør dette vanskelig.

**Sørg for regelmessig evaluering av løsning, formål og effekt**

Smittesporingsappen er nybrottsarbeid og rammebetingelsene kan endre seg fort. COVID-19 forekomsten vil endre seg og ny teknologi kan bli tilgjengelig, derfor er det viktig å ha regelmessige evalueringer for å sikre at de valgte løsningene fremdeles er egnet og at problemet man prøver å løse fremdeles er relevant. I tillegg bør det gjøres vurderinger om man har oppnådd en god nok kvalitet på kontaktsporingen.

# Appendiks A: Ordforklaringer

**Avidentifiserte og aggregerte data**

Nytteverdien av et datasett er tett knyttet til personvernkonsekvensene av det samme datasettet. Man kan gjøre tiltak som å avidentifisere, aggregere og legge på tilfeldig støy for å redusere personvernkonsekvensen men dette vil også gå utover verdien til datasettet. Personvernforordningen er vag rundt hvilke krav som stilles for at et datasett er anonymt. Vi ønsker ikke å begi oss inn på en tolkning av den juridiske definisjonen og mener at seriøse aktører som har gode formål og behandlingsgrunnlag ikke bør påberope datasett anonymitet.

**BLE**

Bluetooth Low Energy

**Bluetooth**

Alle steder det refereres til Bluetooth for denne løsningen mener vi Bluetooth Low Energy (BLE).

**Differential privacy**

Med differential privacy mener vi at mekanismen som aggregerer data fra rådatasettet er $(\varepsilon, \delta)$-differential private med $\delta < 1$. Ved liten $\varepsilon$ og mikroskopisk $\delta$ vil en slik mekanisme ha en veldig sterk garanti for at en brukers bidrag ikke vil ha en personvernskonsekvens for brukeren, gitt at de underliggende dataene er trygge.

**Indekspasient**

I kontekst av denne rapporten brukes benevnelsen *indekspasient* om en person som har fått påvist smitte av COVID-19 og som en smittesporingsaktivitet tar utgangspunkt i. Personene som har vært i kontakt med en indekspasient refereres gjerne til som *kontakter*.

**K-anonymitet**

Ett datasett sies å være k-anonymt hvis det for alle rader i datasettet finnes minst k-1 helt like rader.

**Kontaktsporing**

Å finne noen som potensielt har vært i kontakt med noen som er smittet.

**Lokasjonsdata**

Brukes om GPS-data.

# Appendiks B: Parter og roller

## Helse og Omsorgsdepartementet (HOD)

Fra Wikipedia:

Helse- og omsorgsdepartementet (forkortet HOD) er et norsk departement som har ansvaret for helsepolitikk, folkehelse, alkohol- og narkotikapolitikk, helsetjenester og helselovgivning i Norge. Departementet har som mål å sørge for at befolkningen får gode og likeverdige helsetjenester uavhengig av bosted og økonomi, og bidra til å fremme god folkehelse.

Nettside: https://www.regjeringen.no/en/dep/hod/id421/

**Rolle i Smittestopp**

- Helse- og omsorgsdepartementet har det overordnede ansvaret for at befolkningen får gode og likeverdige helse- og omsorgstjenester.
- HOD har ansvar for rammebetingelsene for virksomhetene i sektoren. For SmitteStopp betyr det lovgrunnlag og finansiering.
- HOD er ansvarlig for oppnevnelse av ekspertgruppen for gjennomgang av kildekoden til Smittestopp.

## Folkehelseinstituttet (FHI)

Hentet fra Wikipedia:

Folkehelseinstituttet (FHI) er et norsk statlig forvaltningsorgan underlagt Helse- og omsorgsdepartementet. FHI er en nasjonal kompetanseinstitusjon innen folkehelse i bred forstand og gir råd til myndigheter, helsetjeneste, politikere, media og publikum. FHI har hovedkontor i Oslo.

Instituttets hovedvirkeområder er helseovervåking, forskning og forebygging. Instituttets faglige virksomhet foregår i fagdivisjoner (betegnet områder): Område for psykisk og fysisk helse, område for smittevern, miljø og helse, område for helsedata og digitalisering samt område for helsetjenester.

Viktige fokusområder er beredskap innen smittevern og miljømedisin, psykisk helse, rusmiddelforskning, helsestatistikk, befolkningsundersøkelser, laboratoriebasert forskning,

livsstil og helse, sosiale helseforskjeller, helseovervåking og registre, samt internasjonale helseproblemer. FHI har et omfattende samarbeid med Verdens helseorganisasjon.
FHI ble dannet i 2001 ved fusjon av Statens institutt for folkehelse og andre institusjoner og het offisielt Nasjonalt folkehelseinstitutt frem til 2016.

Nettside: https://www.fhi.no/

**Rolle i smittestopp**
Bestiller.

## Norsk Helsenett SF

Norsk Helsenett SF (NHN) er et norsk statsforetak etablert og eid av Helse- og omsorgsdepartementet. NHNs oppdrag er å levere og videreutvikle nasjonal IKT-infrastruktur for sikker samhandling mellom aktørene i helse- og omsorgssektoren.

Nettside: https://www.nhn.no/

**Rolle i smittestopp**
Norsk Helsenett er ansvarlig for innsynsløsningen og nettstedet helsenorge.no.

## Simula Research Laboratory (Simula)

Fra Wikipedia:

Simula Research Laboratory AS er et norsk forsknings- og innovasjonsselskap som inngår i Universitetet i Oslos "randsone". Det ble etablert som et prosjekt under Universitetet i Oslo i 2001, og ble stiftet som aksjeselskap i 2002.

Selskapet skal drive grunnleggende langsiktig forskning på utvalgte områder innen programvare- og kommunikasjonsteknologi, og gjennom dette bidra til nyskaping og innovasjon i næringslivet. Statlig eierskap i selskapet skal bidra til å sikre et høyt internasjonalt nivå på forskningen og utdanning av høyt kvalifiserte forskere.

Selskapet har ikke erverv til formål, og skal ikke dele ut utbytte til sine eiere. Simula Research Laboratory er et ideelt og allmennyttig foretak. Selskapet har tre hovedoppgaver; forskning på høyt internasjonalt nivå, utdanning i samarbeid med norske universiteter og nyskaping basert på forskningen i senteret. Senteret driver forskning innen tre fagområder; Networks and Distributed Systems, Scientific Computing og Software Engineering. Simula Research Laboratory har sitt navn fra programmeringsspråket Simula som ble utviklet på 1960-tallet.

**Rolle i Smittestopp**

Simula står for hovedvekten av utvikling, og koordinering av utvikling, i Smittestopp. De er hovedansvarlige for utvikling av iOS-og Android-applikasjonen, og utvikling av programvare for analyse av data.

## Shortcut

Shortcut er et norsk selskap som leverer utviklingstjenester, hovedsaklig med fokus på app-utvikling på smarttelefoner. Nettside: https://shortcut.no/

**Rolle i Smittestopp**
Shortcut har bidratt til utvikling av iOS- og Android-app for Smittestopp i samarbeid med Simula Research Laboratory.

## mnemonic

mnemonic er et selskap som tilbyr sikkerhetstjenester.

Nettside: https://www.mnemonic.no/

**Rolle i smittestopp**
mnemonic er ansvarlige for å ta imot loggdata fra backend-systemene til Smittestopp og følge opp disse i sin SOC.

# Appendiks C - Data som er synlig for NHN

GPS Data fra Egress API

| **Feltnavn** | **Forklaring** |
|---|---|
| time_from | Tidspunkt for forrige GPS punkt |
| time_to | Tidspunkt for dette GPS punktet |
| latitude | Breddegrad |
| longitude | Lengdegrad |
| accuracy | Nøyaktighet på breddegrad og lengdegrad |
| speed | Fart |
| speed_accuracy | Nøyaktighet på fart |
| altitude | Høyde over havet |
| altitude_accuracy | Nøyaktighet på høyde over havet |

Tilgangslogg Data fra Access-Log API

| **Feltnavn** | **Forklaring** |
|---|---|
| timestamp | Tidspunkt på siste oppslag |
| phone_number | Pålogget bruker sitt telefonnummer |
| person_name | Navn på person som har gjort oppslag |
| person_organization | Organisasjons navn på person som har gjort oppslag |
| person_id | Id nummer for person som har gjort oppslag |
| technical_organization | Organisasjons navn på person som har gjort oppslag |
| legal_means | Formålet med oppslaget |
| count | Hvor mange ganger oppslaget er blitt gjort |

# Appendiks D - Azure, roller og tilganger

## Mer detaljerte tabeller over roller og tilganger i Azure (“skyen”)

| **Rolle** | **Beskrivelse** | **Organisasjon** |
|---|---|---|
| Eier - subscription | Administrere hele løsningen utenom databaseinnhold | FHI eier, NHN administrerer |
| Bidragsyter - subscription | Endre hele løsningen utenom databaseinnhold | Simula |
| Leser - subscription | Lese hele løsningen utenom databaseinnhold | Simula |
| Analytiker | Lese “anonymiserte” data fra virtuelle maskiner | Simula |
| SQL dba | Administrere DB med pseudonymiserte data | Simula, Microsoft |
| Admin - B2C | Lese kobling mellom mobilnr og device ID | Simula, NHN |

Relativt til de individuelle datalagrene i Azure betyr det følgende tabell.

| **Navn** | **Formål** | **Tilgang for roller** |
|---|---|---|
| B2C/AD | Autentisering, registrering. | ● Eier<br>● Bidragsyter<br>● Leser<br>● Admin B2C |
| IOTHub | Meldingskø, mottak og midlertidig lagring av sporingsdata. | ● Eier<br>● Bidragsyter<br>● Leser |
| Data Lake | Backup av mottatte sporingsdata. | ● Eier<br>● Bidragsyter<br>● Leser |
| SQL-Db | Hovedlager av sporingsdata | ● Eier<br>● Bidragsyter (Read only)<br>● SQL Dba |

| | | |
|---|---|---|
| Virtuelle maskiner | Mellomlagring for generering av “anonymiserte” kontaktlister. | ● Eier<br>● Bidragsyter<br>● Analytiker |
| Logg-lagring | Lagre diagnoselogger, for Splunk (mnemonic) | ● Shared Access Key |

Tabeller i SQL (kolonner angitt vertikalt per tabell)

| **BTEvent** | **GPSEvent** | **Staging** |
|---|---|---|
| id | id | uuid |
| pairedtime | timefrom | platform |
| rssi | timeto | osversion |
| txpower | lat | appversion |
| daypart | long | model |
| pairedid | accuracy | ev |
| | speed | day |
| | speedaccuracy | hour |
| | altitude | filepath |
| | altitudeaccuracy | isgps |
| | daypart | |
| | mps | |

I tillegg er det definert noen nyttetabeller, som konverterer fra uuid til id og tabell over id sin siste aktivitet. En uuid kan altså ha mange id-er i event-tabellene.

# Appendiks E : Eksisterende standarder og løsninger

Denne seksjonen av rapporten tar ikke sikte på å gi noen uttømmende oversikt over andre løsninger for smittesporing, ettersom det ville være for plasskrevende. I stedet vil vi se litt på andre tilnærminger til smittesporing.

## PACT: Private Automated Contact Tracing

Private Automated Contact Tracing[7] (PACT) er er en protokoll laget av MIT. Protokollen baserer seg på kringkasting av tidsbegrensede 224 bits (28 bytes) tall som man kaller "chirps". Disse tallene genereres ved at mobil-app'en genererer et tilfeldig 256 bits tall hver time ("seed"), som bare er kjent for app'en. Dette tallet brukes så for å produsere en sekvens av chirps som kan kringkastes. Et chirp $\mathbf{r_t}$ produseres ved ved å regne ut:

> $\mathbf{r_t = PRF(s,t)}$, der **t** er tid angitt i minutter, **s** er seed-verdien for inneværende time og PRFer en pseudorandom funksjon.

Disse chirp-verdiene kringkastes så over Bluetooth Low Energy sammen med sendestyrke, og kan så observeres av andre parter innenfor rekkevidde av BLE-sendingen. En observatør (typisk andre brukere av app'en) vil kunne large chirp-verdi, sendestyrke[8], observert signalstyrke og tid, men også berike dataene med kontekstinformasjon slik som lokasjon, og lagre disse i en lokal database på mobiltelefonen.

Kombinasjonen av sendestyrke og observert signalstyrke kan brukes for å estimere avstand.

Det er ikke mulig ut ifra chirp-verdiene alene å enkelt avgjøre at et sett med chirps kommer fra samme mobil. For å gjøre dette må man ha tilgang på seed-verdien som bare finnes på avsenderens mobil. (Det bør dog bemerkes at man kan avgjøre at en serie chirp-verdier kommer fra en gitt mobiltelefon basert på lokale observasjonsforhold).

Når en person er smittet vil denne motta en autorisasjonskode som setter vedkommende i stand til å laste opp sine seed-verdier til en sentral database. En bruker kan selv velge hvilke tidsrom hun vil laste opp seed verdier for. Om brukeren ikke ønsker å laste opp seed-verdiene for et gitt tidsrom vil app'en kunne erstatte disse seed'ene med nye tilfeldige verdier, og det er ikke mulig ut ifra dataene alene å fastslå hvilke tidsrom brukeren holder tilbake data fra.

Seed-verdiene fra infiserte brukere lastes daglig ned av alle brukere av app'en. Disse kan så brukes for å generere alle chirp-verdier den infiserte brukeren kan ha kringkastet i de periodene som er dekket av de opplastede dataene. Dette skjer lokalt på mobiltelefonen og hver

[7] https://pact.mit.edu/

[8] BLE-standarden spesifiserer at avsender kan inkludere data om valgt sendestyrke som metadata.

chirp-verdi slås opp i den lokale databasen på mobiltelefonen. Man kan så bruke eventuelle treff i databasen for å kalkulere risiko for infeksjon og beslutte om brukeren skal varsles.

Brukeren har autonomi i valg av deling av data og kan velge å utelate deler av dataene ved opplasting ved å erstatte seed-verdier.

## DP-3T: Decentralised Privacy-Preserving Proximity Tracing

Decentralised Privacy-Preserving Proximity Tracing[9] (DP-3T) er en frittstående, åpen protokoll for kontaktsporing i forbindelse med COVID-19. Den er opprinnelig produsert av forskere og akademikere på tvers av Europa, siden har mange bidratt til å forbedre denne.

Medlemmer fra DP-3T har deltatt i det løse samarbeidet rundt Pan-European Privacy-Preserving Proximity Tracing[10] (PEPP-PT), men er ikke eneste protokoll under denne paraplyen. Protokollen er implementert i mobilapplikasjoner, tilgjengelig som åpen kildekode[11].

Systemet baserer seg på distribuert, kryptert lokal lagring av kontakt (definert som BLE-treff mellom enheter med applikasjon installert), og bruk av en sentral server som kommunikasjonskanal for smittede identifikatorer. Mobilapplikasjonen sletter løpende foreldet data, oppgitt som 14 dager.

Dersom en bruker diagnostiseres med COVID-19, kan denne brukerens logg over treff siste 14 dager lastes opp til en sentral server, dog kun i koordinasjon med helsemyndighetene ved at brukeren får en autorisasjonskode som kan brukes til å autorisere opplasting av data dersom brukeren gir eksplisitt tillatelse til dette.

DP-3T har to alternative design; Felles for dem er prinsippene om at all beregning skjer lokalt – den sentrale server lagrer kun en kompakt representasjon av midlertidige IDer fra bekreftet smittede (innen smittsom tidsrom). Mobilapplikasjonene laster ned disse, og rekonstruerer IDer i en sjekk mot egen historikk.

DP-3T lar også brukere ta stilling til om de ønsker å dele anonym data med epidemiologer ved installasjon. Opplasting er begrenset til tilfeller der det er påvist kontakt med en smittet person, og inneholder kun anonym data om kontakthendelser brukeren har hatt med bekreftet smittede personer. Brukeren må bekrefte at de ønsker å laste opp data dersom det viser seg at hen har hatt kontakt med noen som er smittet.
For hver smittede person brukeren har hatt kontakt med lastes det opp status på brukerens smitte, konktaktnøkkel og metadata som antall treff de to har hatt og tidsinformasjon per treff

[9] https://github.com/DP-3T/documents
[10] https://www.pepp-pt.org/
[11] https://github.com/DP-3T

relativ til når symptomer startet(for å si noe om hvilken periode hvilken fase av den smittsomme perioden kontakt inntraff. Ingen lokasjon eller presis tidspunkt deles.
Dersom brukeren ikke har hatt kontakt med noen som er smittet lastes dummy-data opp for å hindre traffikkanalyse.

Dokumentet poengterer at alle løsninger basert på Bluetooth vil være sårbare for wardriving, samlere, antenner, osv. og å kunne kombinere dette med metadata som lokasjon, timing eller AV-opptak. Forfatterne hypotetiserer k-out-of-n deling, pluss eventuell bruk av RSSI, som løsning. Det poengterer at ingen *trenger* å få vite den globale interaksjonsgrafen.
Infiserte IDer vil kunne skaffes av angripere. En ond server kan identifisere brukere. Falsk kontakt kan ikke utelukkes. Jamming kan skje.

## Apple|Google

Apple[12] og Google[13] kunngjorde 10. April et samarbeid om en protokoll for kontaktsporing i sammenheng med COVID-19. Resultatet er ikke en app som sådan, men dokumentasjonsutkast til spesifikasjoner for rammeverk API, kryptografi og Bluetooth. Spesifikasjonen baserer seg på å bruke BLE Beacon-teknologi til å utveksle unike, roterende, device-spesifikke tokens.

Ved registrering i app vil hver enhet generere en unik *tracing key* (32 bytes fra en kryptografisk random number generator). For hvert døgn vil enheten utlede en *daily tracing key* (HMAC-basert nøkkelderiveringsfunksjon basert på *tracing key* og dag siden start, trunkert til 16 bytes). Ved jevne mellomrom, når Bluetooth MAC randomiseres (hvert 15. min) deriveres en ny *rolling tracing key* (16 bytes trunkert hash av *daily tracing key* og tidspunktet BLE MAC ble endret).

Når en bruker tester positivt på COVID-19 regner enheten ut *daily tracing keys* for dagene da brukeren kunne vært smittsom basert på *tracing key*. Disse, i tillegg til tilhørende dag nummer (dag siden start) sendes så til diagnoseserver, som aggregerer nøkler for alle brukere som er diagnostiserte og distribuerer dem til alle klienter som bruker kontaktsporing. Dermed kan alle klienter hente ned nøkler assosiert med smitte, utlede *rolling tracing key* basert på disse og sammenligne mot egen kontakthistorikk.

[12] https://www.apple.com/covid19/contacttracing
[13] https://blog.google/inside-google/company-announcements/apple-and-google-partner-covid-19-contact-tracing-technology/

## BlueTrace

Singapore lanserte tidlig sin kontaktsporingsapp (TraceTogether[14]). Kort tid etter beskrev de denne i protokollen BlueTrace[15], og offentliggjorde en referanseimplementasjon som åpen kildekode kalt OpenTrace[16].

BlueTrace baserer seg på å logge treff via BLE mellom deltagende enheter ved ikke identifiserbare beskjeder med midlertidige identifikatorer. Identifikatorene roterer hyppig for å unngå sporing fra tredjeparter. Kontakthistorie lagres lokalt, og denne dataen kan ikke aksesseres direkte av helsemyndigheter.
Dersom en bruker smittes eller er indekspasient bes denne om å dele sin kontakthistorikk med relevante helsemyndigheter ved å bruke en autorisasjonskode. Kun helsemyndigheter har muligheten til å dekryptere den delte kontakthistorikken for å lese personlig identifiserbar informasjon og utføre videre smittesporing.

Når en bruker registrerer seg med telefonnummer tildeler en sentral tjeneste en unik ID til brukerens enhet. Ved hver kontakt med en deltagende BLE-enhet utveksles en roterende midlertidig identifikator (anbefalt levetid: 15 min.), som inneholder unik bruker-ID, sammen med opprettelsestidspunkt og utløpstid for denne kryptert symmetrisk med en nøkkel kun myndighetene har. Dette Base64-enkodes sammen med en initialiseringsvektor og en integritetstag, som også brukes som krypteringsparametere.

Disse midlertidige identifikatorene sendes i forover-daterte batcher med fra server til enheter med jevne mellomrom. Protokollen støtter også federering.
Spesifikasjonen nevner også at BlueTrace-enheter bør implementere blacklisting av nylig sette enheter, og ikke forsøke å tilkoble disse mens de er svartelistet for å unngå å lagre duplikattreff.

Protokollen innebærer også å utveksligssignalstyrkeindikator (RSSI), enhetsmodell (på grunn av variasjon mellom modeller selv med samme RSSI), protokollversjon og aktuell helseautoritet.
En av grunnene protokollens skapere oppgir til å ikke bruke GPS er at det er for unøyaktig, og at de anser det som for dårlig til innendørs bruk. De mistenker også at færre kan ville bli med på lokasjonssporing enn på ren kontaktsporing.

På grunn av begrensninger på bakgrunnsfunksjonalitet i Bluetooth på iOS oppfordrer de brukere med iPhone til å holde appen i forgrunnen og ikke låse skjermen – men å snu mobiltelefonen opp ned for å aktivere en egenimplementert strømsparemodus som demper skjermens belysning.

[14] https://www.tracetogether.gov.sg/
[15] https://bluetrace.io/
[16] https://github.com/opentrace-community

## Smittestop (Danmark)

En offisiell Dansk app som skal hjelpe til å begrense smitte, kalt Smittestop, slippes angivelig[17] i løpet av kort tid. Appen så opprinnelig[18] ut til å være mer eller mindre identisk med Norske Smittestopp hva angikk implementasjonsdetaljer, og skulle også ha flere formål – for eksempel for å kunne vurdere myndighetenes smitteverntiltak, eller å aggregere data til forskning. Det er nå gjort en pivotering[19]. Nå er planen at appen skal basere seg på Apple og Google sine oppdaterte APIer for kontaktsporing via Bluetooth.

## NHS COVID-19 (UK)

Britiske National Health Service (NHS) har lansert[20] (i begrenset omfang, under test) en app for bruk i arbeidet med å sakke utbredelsen av COVID-19. Deres app registrerer kontakt ved hjelp av BLE (inkl. senderstyrke, signalstyrke, dato og lengde). De bruker daglig roterende offentlige identifikatorer, som løpende sendes til et lokalt lager. Samtidig innhentes detaljer om mobiltelefonen, som NHS sier trengs for å avgjøre avstand mellom to mobiltelefoner som har kontakt nøyaktig.

For å registrere seg oppgir brukerne første del av sin postkode (slik at NHS kan planlegge respons), men oppgir ingen annen personlig informasjon. Dermed finnes det i utgangspunktet ingen kobling mellom data og person i systemet. Det opprettes en “installasjons-ID” som er spesifikk for enheten, som er koblet mot all data – som må anses som pseudonymisering, og dermed personopplysninger.

Brukere av appen varsles (via notification) dersom de har vært i kontakt med en annen bruker som oppgir at de har opplevd symptomer på viruset via selvrapportering. Dermed kan kontaktene isolere seg selv inntil de eventuelt kan testes av helsemyndighetene.

NHS har uttalt at de også kommer til å bruke data som samles inn til forskningsformål relatert til folkehelse.

## StopCovid

Frankrike utvikler[21] en app basert på ROBERT[22]-protokollen (som er et forslag for PEPP-PT), som de tidligere utarbeidet i samarbeid med tyske myndigheter. De forespeiler å kun bruke BLE for å registrere kontakt, og å kommunisere “midlertidige pseudonymer” mellom enheter.

---

[17] https://fagbladet3f.dk/artikel/dansk-corona-app-skal-bygges-paa-google-og-apples-teknologi
[18] https://www.dr.dk/nyheder/penge/dansk-corona-app-efter-norsk-model-det-kan-du-forvente
[19] https://www.altinget.dk/artikel/heunickes-corona-app-skifter-kurs
[20] https://www.ncsc.gov.uk/information/nhs-covid-19-app-explainer
[21] https://techcrunch.com/2020/04/28/france-postpones-parliament-debate-on-contact-tracing-app/
[22] https://github.com/ROBERT-proximity-tracing/documents

Disse “midlertidige synonymene” deles ut fra en sentral server ved registrering av app, og kan kobles mot en bruker av brukeren selv og autoritete. Hver enhet vil laste opp alle sine data løpende til sentralt lager. Helsemyndighetene vil vurdere risikofaktor på de ulike treffene og dersom en gitt bruker diagnostiseres kan brukeren velge å sende inn “midlertidige pseudonymer” for aktuelt tidsrom, merket som med risiko.

## TBD (Tyskland)

Tyskland hadde opprinnelig tenkt til å lage en løsning som var forespeilet lik som Frankrikes planlagte løsning, men har nå droppet[23] sine planer om en app basert på ROBERT-protokollen etter at Apple og Google kunngjorde sitt samarbeid om nye platform-APIer, og ser nå ut til å arbeide mot en løsning med desentralisert arkitektur som vil fungere med disse.

## Rakning C-19

Islands løsning[24] bruker kun lokasjonsdata, som lagres lokalt på brukernes mobil i 14 dager. Brukere som diagnostiseres med COVID-19 bes dele sine lokasjonsdata med helsemyndighetene. De bes i samme forbindelse om å oppgi sitt nasjonale identifikasjonsnummer for å knytte data opp mot en person.

[23] https://techcrunch.com/2020/04/27/germany-ditches-centralized-approach-to-app-for-covid-19-contacts-tracing/

[24] https://www.covid.is/app/en

# Appendiks F: Bluetooth

## Generell beskrivelse av BLE

Bluetooth Low Energy (BLE) er en trådløs nettverksteknologi som brukes for energieffektiv kommunikasjon mellom, blant annet, mobiltelefoner og tilleggsutstyr som pulsmålere, smartklokker, tastatur mm. BLE sameksisterer på smarttelefoner med Bluetooth Classic som er teknologien som vanligvis brukes i trådløse headset og lignende.

For mer detaljert informasjon om Bluetooth anbefaler vi å besøke https://www.bluetooth.com/. Det finnes også en mer lettlest beskrivelse av Bluetooth teknologier på Wikipedia: https://en.wikipedia.org/wiki/Bluetooth. Mens man leser spesifikasjoner kan det være nyttig å være klar over at ikke alle enheter bruker Bluetooth helt som beskrevet i spesifikasjonen.

De delene vi er mest i kontekst av kontaktsporing er Generic Access Profile (GAP) og Generic Attribute Profile (GATT).

### Generic Access Profile (GAP)

GAP beskriver hvilke roller ulike BLE-enheter spiller i et system der flere enheter deltar, og hvordan disse kan kommunisere med hverandre. GAP beskriver to roller enheter kan ha: *Peripheral* og *Central*. Peripheral vil som oftest brukes om tilbehør som pulsmålere, smartklokker etc. Central er som oftest en mer ressursrik enhet med mer minne og prosessorkraft enn en *Peripheral*, som f.eks en smarttelefon eller en PC.

Merk at samme enhet kan opptre både som *Central* og *Peripheral* på samme tid.

### Generic Attribute Profile(GATT)

GATT definerer et klient/tjener grensesnitt og definisjon av *Services* og *Characteristics*. Man kan, grovt forenklet, tenke på *Characteristics* som datafelter som kan leses, skrives og observeres, og *Services* som en gruppering av *Characteristics* som hører sammen.

GATT er organisert i et hierarki som man kan lese mer om i avsnitt 6.5 av *Bluetooth Core Specification*[25].

## Beskrivelse av virkemåte i Smittestopp

Smittestopp definerer en Service med en universelt unik ID (UUID) med verdien **e45c1747-a0a4-44ab-8c06-a956df58d93a**. Det gjør at når en enhet oppdager en service med

[25] https://www.bluetooth.com/specifications/bluetooth-core-specification/

denne Service ID'en så vet app'en at den har funnet en annen enhet med Smittestopp appen –eller i det minste har den funnet en enhet som *utgir* seg for å være Smittestopp-appen.

Når Smittestopp finner en annen instans av appen forsøker den å koble seg opp til denne for og be om å få lese en *Characteristic* som er identifisert ved ID'en **64b81e3c-d60c-4f08-8396-9351b04f7591**. Dersom dette lykkes vil appen få en *EnhetsID* i retur.

*EnhetsID* er en verdi som unikt identifiserer en deltager i Smittesporing. Denne er per nå *statisk, men en løsning der dette krypteres sammen med timestamp er under utvikling.* Man kan i *teorien* ikke avlede telefonnummer eller annen informasjon direkte av denne IDen, men den identifiserer brukeren unikt i systemet og dette brukes i backend-systemene for å identifisere brukere.

## Begrensninger på iOS og Android

Smarttelefoner som støtter BLE har ofte en del begrensninger i hvordan applikasjoner kan gjøre bruk av BLE.  Årsaken til dette er både at man ønsker å ivareta brukerens personvern, men også for å begrense strømforbruk.

Disse begrensningene gjør at det er utfordringer forbundet med å bruke BLE for kontaktsporing.