



DECEMBER 2019

Nabotjek af udenlandske erfaringer med kommerciel brug af lokaliseringsdata for mobiltelefoner

Erhvervsstyrelsen – afrapportering



Indholdsfortegnelse

1. Ledelsesresumé
2. Generel datarejse
3. Tre cases
4. Minicases
5. Bilag 1. Identificerede cases
6. Bilag 2. Metoder til anonymisering af lokaliseringsdata

Vigtige begreber

Lokaliseringsdata: lokaliseringsdata fra mobiltelefoner, som teleselskaberne indsamler gennem telemaster. Lokaliseringsdata er én form for positionsdata.

LI-leverandører: leverandører af location intelligence; firmaer med særlig ekspertise i bearbejdning af positionsdata. De samarbejder ofte med teleselskaber om at omdanne lokaliseringsdata til platforme og koncepter for anvendelse af mobildata.

Aftagere: private virksomheder eller offentlige organisationer, der anvender lokaliseringsdataprodukter i forbindelse med udvikling eller optimering af egne produkter og services.

Identifiers: personoplysninger såsom navn og cpr-nummer, der kan identificere et individ.

Pseudoidentifiers: personoplysninger såsom alder og køn, som tilsammen kan anvendes til at udlede, hvilket individ der er tale om.

Pseudonymisering: udskiftning af identifiers i datasættet med en ident, id-nummer eller #hashfunktion. Samtidig beholder man en mappingliste, som efterfølgende kan anvendes til at genskabe identifiers. Pseudonymisering kan være et første skridt i anonymiseringen af data.

Anonymisering: anvendelse af særlige teknikker, der har til formål at fjerne eller ændre personoplysninger i en sådan grad, at man ikke længere kan udlede eller udskille et individ på baggrund af data. Data kan være anonymiseret i forskellige grader, men der er ikke noget fast mål for, hvornår data er fuldt anonymiseret.

Aggregering: en særlig anonymiseringsteknik, hvor man ændrer nøjagtigheden/granulariteten af data i en sådan grad, at det ikke længere er muligt at udskille, sammenkoble eller udlede de oprindelige værdier (se bilag 2).

Differential privacy: en særlig anonymiseringsteknik, hvor støj tilføjes systematisk til data, så data sløres på individniveau, men på aggregeret niveau stadig viser de samme mønstre og tendenser (se bilag 2).

Nabotjek af udenlandske erfaringer med kommerciel brug af lokaliseringsdata for mobiltelefoner

Formål og fremgangsmåde

Formål

Teleselskaberne indsamler via deres telemaster lokaliseringsdata fra de enkelte mobileenheder, der er indenfor mastens rækkevidde – data, som benyttes til at dirigere trafikken på telenettet og til at debitere selskabernes abonnenter. Kommerciel anvendelse af anonymiseret lokaliseringsdata rummer et stort potentiale både i forhold til udvikling af nye produkter, services og forretningsmodeller og til udvikling af kunstig intelligens.

I forbindelse med den nationale strategi for kunstig intelligens blev initiativet *Bedre brug af lokaliseringsdata for mobiltelefoner* igangsat med en løbetid på tre år (fra 2019-2022).

Som led i dette initiativ er formålet med denne rapport at udarbejde et nabotjek af, hvordan lokaliseringsdata fra mobiltelefoner er blevet brugt til kommercielle formål indenfor eksisterende juridiske rammer. Der er i rapporten fokus på, hvordan de enkelte projekter indenfor brug af lokaliseringsdata er blevet vurderet og godkendt indenfor de gældende juridiske rammer, herunder hvilke eventuelle tilpasninger der er blevet foretaget til projektet i forhold til at leve op til de juridiske rammer. Endvidere er der afdækket, hvordan projekterne teknisk har været sat op, blandt andet i forhold til aggregering og anonymisering af data samt styring af adgang til data.

Fremgangsmåde

Som led i gennemførelsen af nabotjekket er data indsamlet dels gennem desk research, dels gennem interviews. Desk research er primært anvendt i den indledende afdækning af cases, der har afdækket brancher og aktører involveret i arbejdet med lokaliseringsdata. Desk researchen har bestået i gennemgang af myndighedsrapporter, lovgivning samt teleselskabers og LI-leverandørers hjemmesider.

På baggrund af den indledende desk research blev en række cases udvalgt til deep dives. Disse deep dive-cases er blevet gennemført via interviews med teleselskaber, LI-leverandører og aftagere af lokaliseringsdataprodukter. I alt er der gennemført 13 interviews på tværs af syv lande, der tilsammen afdækker forløbet i tre cases og to minicases og viser, hvordan aktørerne indenfor feltet generelt indsamler, behandler og anvender lokaliseringsdata.

Nabotjek af udenlandske erfaringer med kommerciel brug af lokaliseringsdata for mobiltelefoner

Overordnede konklusioner

Overordnede konklusioner

Nabotjekket afdækkede 26 cases, hvor lokaliseringsdata er blevet anvendt i kommercielt øjemed med følgende overordnede konklusioner:

- Der er primært tale om analyser af trends og mønstre på baggrund af aggregeret data, herunder analyser af bevægelsesmønstre, mobilitet eller lignende indenfor brancherne transport, turisme, detailhandel og marketing.
- Der er ikke fundet eksempler på engrossalg af data.
- Teleselskaber, LI-leverandører og aftagere følger i store træk den samme fremgangsmåde, når de arbejder med lokaliseringsdata.
- Lokaliseringsdata alene har ikke stor værdi, men værdien findes, når lokaliseringsdata kombineres med andre data.
- Størstedelen af de adspurgte aktører oplever ikke nævneværdige juridiske barrierer for deres arbejde med lokaliseringsdata, da GDPR og e-databeskyttelsesdirektivet sætter klare rammer for, hvad man må og ikke må. Dog ses der nationale variationer, hvor Tyskland stiller de skrappeste krav til anonymiseringen af data.

- Aktørerne opsøger ikke af egen drift, med enkelte undtagelser, kontakten med myndighederne, men forholder sig alene til de juridiske rammer i henholdsvis national lovgivning og GDPR. Det betyder også, at der blandt aktørerne er en del usikkerhed om, hvornår data er tilstrækkeligt anonymiseret.
- Teleselskaberne er generelt tilbageholdende med at give adgang til anonymiseret data.
- Der er en overvægt af enkeltstående projekter, hvilket peger på, at markedet for anvendelse af lokaliseringsdata stadig er relativt umodent. Dog er der identificeret enkelte projekter, hvor data leveres kontinuerligt og indgår i et fast produkt.
- Lokaliseringsdata er mindre præcise end f.eks. GPS data. Til gengæld har lokaliseringsdata en høj grad af repræsentativitet.
- Enkelte LI-leverandører oplever, at teleselskaberne har en forsigtig tilgang til anvendelse af lokaliseringsdata.

Nabotjek af udenlandske erfaringer med kommerciel brug af lokaliseringsdata for mobiltelefoner

Overordnede konklusioner

Overordnede konklusioner fortsat

- I Europa er det primært offentlige organisationer, der efterspørger løsninger, der anvender lokaliseringsdata, hvilket kunne tyde på, at markedet endnu ikke har fået øjnene op for potentialet i at anvende anonymiseret lokaliseringsdata kommercielt. Én tese er, at teleselskaberne er mere åbne for at give adgang til anonymiseret data, når det er en offentlig myndighed, der efterspørger adgangen, end når det er en privat aktør.
- Teleselskaberne efterspørger selv analyser af lokaliseringsdata i forbindelse med for eksempel placeringen af master til det nye 5G-netværk.
- De tre cases giver desuden anledning til at konkludere, 1) at kombinationen af lokaliseringsdata og andre datakilder forøger værdien af at anvende lokaliseringsdata, 2) at et tæt samarbejde med forskningsinstitutioner medvirker til at modne projekter til kontinuerlige produkter, og 3) at det er nødvendigt at indgå i tæt dialog og samarbejde med teleselskaberne for at understøtte adgangen til anonymiseret data.

Generel datarejse

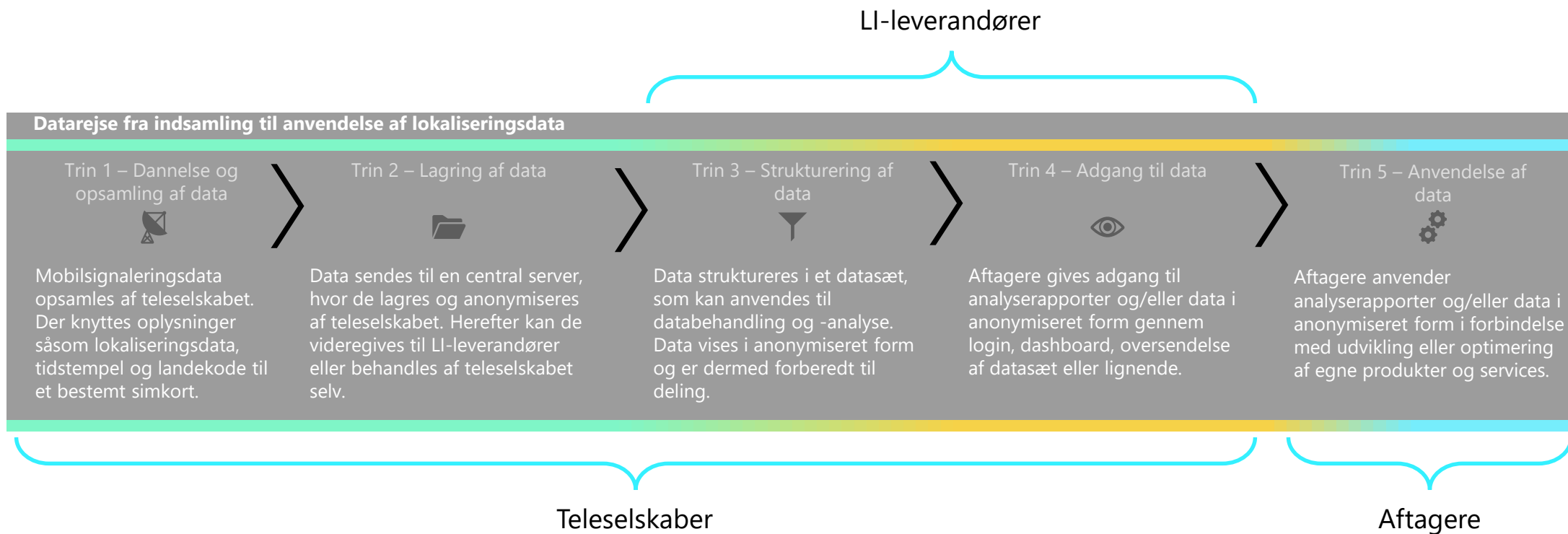
De tværgående konklusioner præsenteres i en generel datarejse, som illustrerer, hvilken spændvidde der er i teleselskabers, LI-leverandørers, private virksomheders og offentlige organisationers indsamling, behandling og anvendelse af lokaliseringsdata.

Gennemgang af cases

Introduktion

Fem trin i datarejsen for lokaliseringsdata

Både beskrivelsen af den generelle datarejse og de tre cases er struktureret omkring nedenstående proces. Processen indeholder fem trin og beskriver trinene, fra mobildata indsamles, til de anvendes til at opnå et konkret, kommercielt formål.








Generel datarejse

Hovedpointer

Fra indsamling til anvendelse af data

Teleselskaber, LI-leverandører og aftagere følger i store træk den samme fremgangsmåde, når de arbejder med lokaliseringsdata. Nedenfor er gengivet de vigtigste pointer for hvert trin i datarejsen:

-  Trin 1 – Dannelse og opsamling af data: Præcisionen af lokaliseringsdata er generelt lavere sammenlignet med andre lokaliseringsdata. Til gengæld har lokaliseringsdata en høj grad af repræsentativitet.
-  Trin 2 – Lagring af data: LI-leverandører får aldrig overdraget data af teleselskabet, men kan tilgå data gennem teleselskabets cloudbaserede platform. Inden LI-leverandører gives adgang til data, pseudonymiserer teleselskaberne altid data. I de fleste tilfælde anonymiserer teleselskaberne ligeledes data, inden de giver adgang til LI-leverandører.
-  Trin 3 – Strukturering af data: Lokaliseringsdata i sig selv besidder sjældent en kommerciel værdi, før de bliver kombineret med andre datakilder. LI-leverandører står oftest for denne kombineret og strukturering af data, da de både har kompetencerne og er tættere på kundernes behov end teleselskaberne.
-  Trin 4 – Adgang til data: Overleveringen af data til aftageren afspejler projektets karakter samt aftagernes kompetencer og behov i forhold til selv at kunne foretage videre analyser af data.
-  Trin 5 – Anvendelse af data: Lokaliseringsdata anvendes stort set udelukkende til analyser af mønstre og trends på tværs af større grupper af individer. Projekter, der fokuserer på enkeltstående individer, er sjældne.

Juridiske rammer og myndighedskontakt

De fleste aktører, der arbejder med lokaliseringsdata, oplever kun enkelte, mindre juridiske barrierer i deres interaktion med EU-lovgivning og national lovgivning. Disse relaterer sig til usikkerhed om, i hvor høj grad data skal aggregeres, inden de er tilstrækkeligt anonymiserede, til at de må anvendes til kommercielle formål.

Derudover er teleselskaber, LI-leverandører og aftagere sjældent i kontakt med myndigheder om anvendelse af lokaliseringsdata, på trods af at det primært er offentlige organisationer, der efterspørger projekter indenfor lokaliseringsdata.

Barrierer

Enkelte LI-leverandører oplever, at teleselskaberne har en forsigtig tilgang til anvendelse af lokaliseringsdata. Derudover er der enkelte eksempler på, at teleselskaberne overestimerer prisen på deres data.

Dannelse og opsamling af data 1/2

Pointe: En høj grad af repræsentativitet i indsamlede lokaliseringsdata opvejer en relativt lavere grad af præcision sammenlignet med andre positionsdata. Præcisionen af lokaliseringsdata er generelt højest i byområder, hvor tætheden af telemaster er størst.

Opsamling gennem telemaster

Teleselskaberne indsamler to typer lokaliseringsdata:

- Passive lokaliseringsdata
- Aktive lokaliseringsdata.

De passive data består af CDR'er (call detail records). CDR'er er oplysninger, som teleselskaber i lang tid har registreret, da de oprindeligt og stadig anvendes til at opgøre forbruget blandt teleselskabernes kunder. CDR'er registrerer og dokumenterer handlinger foretaget i et telenetværk såsom telefonopkald og tekstbeskeder. CDR'erne indeholder forskellige attributter tilknyttet telefonopkaldene og tekstbeskederne, for eksempel i form af opkaldsvarighed, telefonnummer på initiativtageren til og modtageren af opkaldet/beskeden, tidstempel, og hvilke telemaster opkaldet/beskeden forbindes til.

Indenfor de seneste år er teleselskaberne ligeledes begyndt at indsamle aktive lokaliseringsdata. Aktive lokaliseringsdata har udelukkende til formål at dokumentere mobiltelefonbrugernes lokalisering. De aktive data indsamles ved, at mobiltelefoner cirka hvert femte minut automatisk forbindes til nærmeste telemast, hvorefter data vedrørende mobiltelefonens lokalisering registreres. I modsætning til CDR'er, hvor lokaliseringsdata kun registreres, når mobiltelefonbrugeren foretager et opkald, sender en besked eller lignende, registreres de aktive data automatisk med et fast tidsinterval.

Repræsentativitet

Blandt de interviewede teleselskaber og LI-leverandører fremhæves repræsentativitet som en af de primære kommercielle fordele ved at anvende lokaliseringsdata.

Teleselskaber med en stor markedsandel indenfor det geografiske område, hvori de ønsker at indsamle data, kan indsamle data fra et stort og repræsentativt udsnit af mobiltelefonbrugere. Det skyldes blandt andet, at teleselskaberne ikke behøver at indhente samtykke fra deres kunder til at anvende anonymiseret data, der er indsamlet gennem telemaster.

Andre positionsdata såsom gps-data vil i de fleste tilfælde være mindre repræsentative end lokaliseringsdata, da den enkelte mobiltelefonbruger skal give samtykke til, at der må indsamles gps-data gennem telefonen. Det vil derfor kun være et udsnit af mobiltelefonbrugere, der giver tilladelse til, at firmaer indsamler deres lokaliseringsdata, og ofte vil der være en skævhed i, hvilke typer personer der giver samtykke. Denne risiko er ikke til stede i analyser af lokaliseringsdata, da teleselskaberne indsamler data fra alle mobiltelefonbrugere.



Dannelse og opsamling af data 2/2

Pointe: En høj grad af repræsentativitet i indsamlede lokaliseringsdata opvejer en relativt lavere grad af præcision sammenlignet med andre positionsdata. Præcisionen af lokaliseringsdata er generelt højest i byområder, hvor tætheden af telemaster er størst.

Præcision

Præcisionen i lokaliseringsdata er generelt dårligere sammenlignet med anden positionsdata såsom gps-data. Præcisionen vil som hovedregel være bedst i byområder og dårligst i mere tyndt befolkede områder. Samtidig vil det ofte være nødvendigt at kombinere lokaliseringsdata med andre datakilder eller benytte algoritmer, hvis der er behov for at opnå en præcision, der er bedre end 50 meter.

Præcisionen af lokaliseringsdata varierer indenfor intervallet 50 meter til 20 km. Præcisionen afhænger af tætheden af telemaster i det område, hvor data indsamles af teleselskabet. Typisk vil der være en større koncentration af telemaster i byområder, hvorfor præcisionen oftest er bedre i byer sammenlignet med landet og andre tyndt befolkede områder.

Faktorer, der kan forværre præcisionen, er såkaldte pingpongeffekter. Disse effekter opstår, når en person befinder sig lige mellem to telemaster. I dette tilfælde vil personens mobiltelefon skiftevis forbinde sig til de to telemaster. Hvis de to telemaster befinder sig langt fra hinanden, vil det i de indsamlede data derfor fremstå, som om personen har bevæget sig urealistisk hurtigt over en stor distance.

Pingpongeffekten er et eksempel på, hvordan lokaliseringsdata kan have vanskeligt ved at opfange glidende bevægelsesmønstre. Da mobiltelefoner ikke konstant opretter forbindelse til telemaster, er teleselskaberne og LI-leverandørerne nødt til at benytte algoritmer til at beregne de sandsynlige bevægelser, der foretages imellem de datapunkter, hvor mobiltelefonerne sender data til telemasterne.

Udfyldning af datahuller

Teleselskabet O2 i Storbritannien anvender 20 forskellige algoritmer i forbindelse med deres analyser af transportmønstre for blandt andet at kunne identificere, hvor mobiltelefonbrugerne befinder sig, når deres mobiltelefon ikke opretter forbindelse til en telemast.



Lagring af data 1/1

Pointe: For at have kontrol over data opbevarer teleselskaberne data i cloudbaserede platforme, hvorfra teleselskaberne kan styre LI-leverandørernes adgang til pseudonymiseret data.

Pseudonymisering

Første skridt i at gøre de indsamlede lokaliseringsdata klar til videre databehandling er, at teleselskabet foretager en pseudonymisering af data. Processen for pseudonymisering følger overordnet de samme skridt på tværs af lande og teleselskaber med undtagelse af Tyskland.

Data, der er pseudonymiseret, har fået fjernet alle direkte personhenførbare oplysninger (identifiers) såsom navn og cpr-nummer, som er blevet erstattet med et tilfældigt id-nummer eller #hash; dermed kan personerne i datasættet ikke direkte identificeres. Det vil dog stadig være muligt at genskabe de oprindelige data ved hjælp af en nøgle, herunder knytte data til en person.

Det er et datasikkerhedscenter i de enkelte teleselskaber, der har ansvaret for at foretage pseudonymiseringen af data. Det er dermed kun enkelte personer i teleselskaberne, der har adgang til at se data, inden identifiers fjernes. Selve processen med at fjerne identifiers og tilføje tilfældige id-numre foretages ikke manuelt af de enkelte personer, men maskinelt.

I Tyskland skal teleselskaber følge særligt restriktive krav til pseudonymiseringen; her sker der en rehashing af data hvert døgn. Det vil sige, at det enkelte datapunkt tildes et nyt id-nummer hvert døgn, så det bliver sværere at tilbagespore, hvilke personer der er repræsenteret i de pseudonymiseret data.

Opbevaring af data

På tværs af de undersøgte lande er der variation i graden af databehandling, som lokaliseringsdata har gennemgået, før teleselskaberne giver eksterne aktører såsom LI-leverandører adgang til data.

I enkelte tilfælde kan LI-leverandører få adgang til data, der udelukkende har fået fjernet identifiers. I de fleste tilfælde er data dog allerede aggregeret, når LI-leverandøren får adgang til dem. Når data er aggregeret, ændres detaljeringsgraden, så data gøres mere upræcise, for eksempel ved at præsentere personers alder i et interval i stedet for den præcise alder (se gennemgang af trin 3 for uddybning af aggregering). Juridisk er data mere sikkert, når LI-leverandøren først får adgang til dem efter aggregering. Derimod begrænser det mulighederne for at kombinere lokaliseringsdata med andre datakilder, når data allerede er aggregeret. Det kommercielle potentiale i udnyttelsen af data bliver derfor reduceret.

Teleselskaberne lagrer data i cloudservere som for eksempel AWS (Amazon Web Services) og ikke i on-premise-servere, som er fysisk placeret hos teleselskaberne. Teleselskaberne får, i samarbejde med cloudleverandøren, opsat deres egen platform i cloudserveren, hvor alle lokaliseringsdata placeres.

Teleselskaberne har adgang til platformen gennem en loginløsning og kan også give LI-leverandører loginadgang til data i platformen. LI-leverandørerne kan ikke eksportere data, der udelukkende har fået fjernet identifiers fra teleselskabernes platforme, og de pseudonymiseret data forlader derfor ikke teleselskabernes platforme.



Strukturering af data 1/2

Pointe: Lokaliseringsdata bliver oftest anvendt til analyser af mønstre i data, hvorfor aggregering er den mest benyttede anonymiseringsteknik. Ofte forestår LI-leverandørerne disse analyser af datamønstre, da de er tættere på aftagerne end teleselskaberne og besidder de nødvendige specialkompetencer indenfor databehandling.

Anonymisering

Overordnet anvender teleselskaberne og LI-leverandørerne to teknikker til anonymisering af data:

- Aggregering
- Differential privacy

Aggregering er den mest benyttede anonymiseringsteknik blandt interviewpersonerne i denne undersøgelse, mens anonymiseringsteknikken differential privacy også anvendes i enkelte tilfælde. De to teknikker er præsenteret mere dybdegående i bilag 2.

Der kan være tale om aggregering af flere forskellige typer data og ved hjælp af forskellige metoder:

- Geografisk aggregering: Angivelsen af mobiltelefonbrugernes lokalisering gøres mindre præcis. Eksempelvis grupperes de indsamlede specifikke lokaliseringer i postnumre i stedet for at vise, præcis hvilken lokalisering data stammer fra.
- Aggregering af baggrundsoplysninger: Angivelsen af mobiltelefonbrugernes personoplysninger gøres mindre præcis, for eksempel ved angivelse af personers alder i intervaller i modsætning til specifik alder.

Differential privacy er et ekstra anonymiseringslag, som lægges på data, inden det aggregeres. Ved denne anonymiseringsmetode tilføjes der støj til data på systematisk vis. En måde at anvende differential privacy på er at tilføje falske personer til datasættet eller ændre værdierne konsekvent på data. Eksempelvis kan databehandleren trække fem år fra alderen på nogle individer i datasættet og lægge fem år til alderen på lige så mange individer i datasættet. Data sløres dermed på individniveau, men på aggregeret niveau vil data stadig vise de samme mønstre. Data kan derfor stadig anvendes til analyser af mønstre og trends.

Et minimum antal af observationer

Teleselskaber inddeler ofte områder i zoner, hvori der skal være et minimum antal af observationer i samme zone på samme tidspunkt, før data bliver gemt. På denne måde sikres det, at der ikke indgår individuelle observationer i datasættet, som er lette at knytte til en specifik person, der for eksempel bor i et tyndt befolket område.

Orange i Spanien anvender en minimumsgrænse på 25 individuelle observationer i den samme zone. Hvis antallet af individuelle observationer i zonen er utilstrækkeligt, udvides zonen, indtil minimumsgrænsen overholdes.



Strukturering af data 2/2

Pointe: Lokaliseringsdata bliver oftest anvendt til analyser af mønstre i data, hvorfor aggregering er den mest benyttede anonymiseringsteknik. Ofte forestår LI-leverandørerne disse analyser af datamønstre, da de er tættere på aftagerne end teleselskaberne og besidder de nødvendige specialkompetencer indenfor databehandling.

Kombinering af data

I de fleste tilfælde vil lokaliseringsdata alene ikke have stor værdi for aftagerne. For at skabe relevante analyser er der i nogle tilfælde derfor behov for at kombinere lokaliseringsdata med andre datakilder. Det skyldes blandt andet, at præcisionen af data er lav, og at lokaliseringsdata i sig selv kun indeholder få baggrundsoplysninger om mobiltelefonbrugerne.

Lokaliseringsdata kombineres med tre andre typer data:

- Andre positionsdata
- Socioøkonomiske oplysninger og andre baggrundsoplysninger
- Hjemmeside- og applikationsdata.

For at øge præcisionen af data kombineres lokaliseringsdata med andre typer positionsdata. Teknikken består i at kombinere lokaliseringsdata med andre datakilder, der mere præcist eller oftere registrerer personers lokation. I analyser af offentlig transport fra Storbritannien og Estland suppleres lokaliseringsdata for eksempel med positionsoplysninger fra de to landes svar på rejsekort.

For at øge dybden i data kan lokaliseringsdata kombineres med socioøkonomiske oplysninger; dette ses blandt andet i forbindelse med analyser foretaget for butikcentre. I disse analyser anvendes lokaliseringsdata til at identificere, hvor kunderne bor, hvilket kædes sammen med en række antagelser om deres socioøkonomiske baggrund.

En sidste kombinationsmulighed, som LI-leverandørerne kan anvende, er kobling mellem lokaliseringsdata og hjemmeside- eller applikationsdata. Ved denne kobling bliver det muligt at se, hvilke mobilapplikationer og hjemmesider brugeren af mobiltelefonen anvender på forskellige lokationer. Denne datakombination kan for eksempel anvendes til at identificere, om personer, der befinder sig i nærheden af en reklame for et firma, går ind på firmaets hjemmeside eller benytter deres app på telefonen.



Adgang til data

Pointe: Aftagerne får adgang til data i det format, der passer bedst til deres analysekompetencer og behov. Aftagerne vil oftest få adgang til data med en forsinkelse, hvilket skyldes kvalitetssikrings- og anonymiseringsprocesser samt økonomiske hensyn.

Overlevering af data til aftagere

Når aftagerne skal vælge et overleveringsformat, er det særlig vigtigt at tage hensyn til, om der er tale om et enkeltstående kontra et kontinuerligt projekt, og om aftageren selv ønsker at have muligheden for at kunne analysere data.

Nedenstående fire overleveringsformater efterspørges i stort set lige høj grad:

- Rapporter
- Datasæt (csv-format)
- Dashboards
- API-adgang

Med rapporter ser aftageren blot resultaterne af analysen. Rapporter er derfor særlig anvendelige i forbindelse med enkeltstående projekter, hvor der skal afrapporteres i forhold til en problemstilling, der finder sted indenfor et afgrænset tidspunkt og sted. Dette kan for eksempel være analyser i forbindelse med en bestemt begivenhed eller en kampagne.

Overlevering af datasæt er målrettet de aftagere, der selv gerne vil analysere data. Der vil derfor oftest være tale om aftagere, der har erfaring med dataanalyse. De overleverede data vil dog stadig være anonymiserede, hvilket begrænser aftagernes muligheder for at anvende data på individplan.

Dashboards er relevante for de aftagere, der ikke på samme måde ønsker at dykke ned i data, men i stedet gerne vil have præsenteret de overordnede data i et lettilgængeligt visuelt format. Dashboards kan særligt anvendes i forbindelse med projekter, hvor data opdateres flere gange over en længere periode, eller hvor der er tale om fortløbende analyser. Dashboards kan være dynamiske over tid og er ofte interaktive. Interaktive dashboards giver brugeren mulighed for at zoome ned i data, filtrere og sortere analyseresultater.

Ligesom det er tilfældet med dashboards, er API-adgang mest relevant i projekter, hvor data leveres over en længere periode, eller hvor der er tale om fortløbende analyser. Et API er en softwaregrænseflade, der giver LI-leverandørerne mulighed for at udstille data, så aftageren selv kan tilgå data efter behov. Aftageren kan derfor løbende selv følge med i opdateringerne af de data, der stilles til rådighed. Det er derfor også aftagernes egen opgave at downloade og analysere data, hvorfor aftageren selv skal besidde analysekompetencerne hertil.



Adgang til data 2/2

Pointe: Aftagerne får adgang til data i det format, der passer bedst til deres analysekompetencer og behov. Aftagerne vil oftest få adgang til data med en forsinkelse, hvilket skyldes kvalitetssikrings- og anonymiseringsprocesser samt økonomiske hensyn.

Opdateringsgrad

I hovedparten af de kortlagte projekter rapporteres data ved afslutningen af det pågældende projekt eller med en kadence fra fem dage til en måned. Der vil desuden oftest være en forsinkelse på de data, som aftageren får adgang til, og nærrealtidsløsninger er derfor sjældne. Dette skyldes særligt tre forhold:

- Kvalitetssikring
- Anonymisering
- Omkostninger

Før data bliver frigivet, gennemgår det en kvalitetssikringsproces. Teleselskaberne sikrer sig med andre ord, at der ikke er åbenlyse fejl i data, og at data har tilstrækkelig præcision til at kunne anvendes til det givne formål.

Derudover kan aftagerne grundet juridiske restriktioner ikke få direkte adgang til lokaliseringsdata, der indeholder identifiers eller pseudo-identifiers på mobiltelefonbrugerne. Data skal derfor gennemgå en anonymiseringsproces, inden de kan frigives til aftagerne.

Afslutningsvis er der et økonomisk hensyn at tage for aftagere, når de skal beslutte, hvor ofte data skal opdateres. Jo oftere aftagerne ønsker at få opdateret data, desto dyrere bliver det, da teleselskaberne kan kræve betaling for data alt efter opdateringsfrekvens.

Grundet disse forhold er det i de fleste tilfælde stadig udfordrende at levere nærrealtidsdata til aftagere. Ifølge de interviewede LI-leverandører er det kun i få tilfælde, at aftagerne efterspørger nærrealtidsdata. Nærrealtidsløsninger efterspørges eksempelvis i forbindelse med begivenheder, hvor man ønsker at kunne monitorere koncentrationer af store menneskemængder.

Nærrealtid

Teleselskabet O2 i England har for nylig lanceret et produkt, der anvender nærrealtidsdata med ned til 30 sekunders forsinkelse på data. Produktet er møntet på monitorering af trafik og kan ved hjælp af nærrealtidsdata, kunstig intelligens og maskinlæring opstille scenarier for, hvordan trafik mest optimalt kan omdirigeres, for eksempel i tilfælde af trafikulykker, og omlægges. Produktet er stadig relativt nyt og er ikke taget i anvendelse af nogle aftagere endnu.



Anvendelse af data 1/2

Pointe: Lokaliseringsdata anvendes primært til analyser af mobilitetsmønstre blandt større menneskemængder indenfor brancherne transport, turisme, detailhandel og marketing. Enkeltstående projekter og pilotprojekter forekommer oftere end projekter, hvor der kontinuerligt leveres data til aftageren.

Anvendelsesformål

Lokaliseringsdata bliver primært anvendt til analyser af overordnede trends og mønstre indenfor et specifikt forretningsområde. De anvendte data bliver derfor præsenteret i aggregeret form, hvor det er muligt at drage tværgående konklusioner på tværs af en bred skare af personer over en længere tidsperiode. Det bunder særligt i tre forhold:

- Efterspørgsel
- Samtykke
- Præcision af data.

For det første er det analyser af trends og mønstre, som der ifølge LI-leverandørerne er størst efterspørgsel på blandt aftagere; aftagere ønsker identificere tendenser i adfærd og bevægelsesmønstre blandt deres brugere og kunder fremfor at følge det enkelte individ.

For det andet kræver det samtykke fra de enkelte individer, hvis lokaliseringdata med identifiere eller pseudonymiseret lokaliseringdata skal deles med en tredjepart. Teleselskaberne skal derfor indhente samtykke fra deres kunder, hvis de ønsker at foretage individbaserede analyser, hvorimod teleselskaberne ikke er forpligtet til at indhente samtykke ved analyser af anonymiseret data.

For det tredje vil der med den nuværende tekniske infrastruktur sjældent være tilstrækkelig præcision i lokaliseringdata til at kunne levere brugbar information om individets bevægelser.

Kun i enkelte tilfælde såsom indenfor geomarketing er det relevant at kende det enkelte individs lokation. Ved geomarketing modtager den enkelte person reklamer eller lignende baseret på personens lokation. Data vil dog ofte være for upræcise til at kunne fastlægge, præcis hvor individet befinder sig. Det er derfor kun muligt at målrette marketing efter, om personen er hjemme eller ude, eller om personen befinder sig indenfor et bredt defineret område.



Anvendelse af data 2/2

Pointe: Lokaliseringsdata anvendes primært til analyser af mobilitetsmønstre blandt større menneskemængder indenfor brancherne transport, turisme, detailhandel og marketing. Enkeltstående projekter og pilotprojekter forekommer oftere end projekter, hvor der kontinuerligt leveres data til aftageren.

Brancher

De fleste projekter, hvor der anvendes lokaliseringsdata, foregår indenfor brancherne transport, turisme, detailhandel og marketing. Samtidig er der en overvægt af offentlige organisationer, der efterspørger projekter baseret på lokaliseringsdata.

De offentlige projekter foregår oftest indenfor brancherne turisme og transport. Eksempler på projekter indenfor turisme er offentligt ejede turistorganisationer, der ønsker at måle mængde af, bevægelse blandt og oprindelse af turister, der besøger et område eller en specifik begivenhed. Indenfor transport anvendes lokaliseringsdata særligt til planlægning af offentlig transport gennem analyser af mobilitet.

Private projekter udføres overvejende indenfor brancherne detailhandel og marketing. Det kan blandt andet være indkøbscentre, der ønsker at undersøge mængden af kunder, hvornår de har flest kunder, hvor deres kunder kommer fra, kundernes socioøkonomiske baggrunde eller lignende. Ligeledes kan butiksejere ved hjælp af lokaliseringsdata undersøge, hvor det vil være mest relevant at placere en ny butik baseret på viden om bevægelsesmønstre i forskellige områder. Indenfor marketing kan lokaliseringsdata anvendes til at måle effekterne af marketingkampagner ved eksempelvis at undersøge antallet af besøgende omkring en butik før og efter gennemførelsen af kampagnen.

Herudover efterspørger teleselskaberne også selv analyser af egne lokaliseringsdata. Her er formålet at kortlægge, i hvilke områder der bor flest mennesker, og derigennem beslutte, hvor teleselskaberne skal prioritere at etablere 5G-netværk først.

Projekternes karakter

De fleste teleselskaber og LI-leverandører, der medvirker i denne undersøgelse, udarbejder både analyser i forbindelse med enkeltstående projekter og leverer data kontinuerligt til faste kunder. De enkeltstående projekter og pilotprojekter er dog de hyppigst forekommende. Dette indikerer, at markedet for lokaliseringsdata stadig er under udvikling, da de fleste aftagere har svært ved at se, hvordan data kan indgå som en fast bestanddel af udviklingen af deres produkter og services.

Kontinuerlige projekter

Nationalbanken i Estland er et eksempel på en kunde, der over en længere periode har modtaget analyser af lokaliseringsdata. Positium har siden 2009 leveret data til Nationalbanken på månedlig basis.



Rammer og barrierer 1/2

Pointe: Teleselskaber, LI-leverandører og aftagere oplever sjældent juridiske barrierer for deres arbejde med lokaliseringsdata og er sjældent i dialog med nationale myndigheder om rammerne for anvendelse af lokaliseringsdata.

Juridiske rammer

Teleselskaber, LI-leverandører og aftagere, der arbejder med lokaliseringsdata, er underlagt EU-regulering (GDPR og e-databeskyttelsesdirektivet) samt nationale data- og telekommunikationslovgivninger, når de arbejder med indsamling og behandling af lokaliseringsdata. De fleste aktører indenfor feltet oplever i denne sammenhæng kun enkelte, mindre juridiske barrierer i deres interaktion med EU-lovgivning og national lovgivning.

Generelt betragtes de juridiske rammer som grundvilkår, der skal efterleves. Derfor oplever LI-leverandørerne ikke, at lovgivningen begrænser deres produktudvikling. Ingen af de interviewede LI-leverandører giver udtryk for, at de har forsøgt at udvikle produkter, som de har været nødt til at kassere midt i udviklingsprocessen på grund af lovgivning. Ligeledes har ingen af de interviewede teleselskaber, LI-leverandører eller aftagere oplevet juridiske konsekvenser af deres anvendelse af lokaliseringsdata.

Teleselskabet Orange (Spanien) og LI-leverandøren Positium (Estland) nævner dog, at der i forbindelse med aggregering af data kan være visse lovgivningsmæssige uklarheder, i forhold til hvilket granularitetsniveau data må aggregeres til. Orange oplever, når de er i kontakt med andre europæiske teleselskaber, at der blandt teleselskaberne er en usikkerhed, i forhold til i hvor høj grad data skal aggregeres, før de må anvendes.

Myndighedskontakt

Generelt har teleselskaberne, LI-leverandørerne og aftagerne ingen eller meget sparsom kontakt med nationale myndigheder i forbindelse med deres anvendelse af lokaliseringsdata. Kun enkelte af interviewpersonerne er regelmæssigt i dialog med nationale myndigheder.

Orange er løbende i dialog med det spanske svar på Datatilsynet og sparrer regelmæssigt med dem om, hvordan lokaliseringsdata anvendes lovligt i forbindelse med specifikke projekter. Dette bidrager til en øget tryghed hos teleselskabet, i forhold til at de ikke anvender data ulovligt. Interviewpersonen fra Orange oplever derimod ikke, at kolleger fra teleselskaber i andre lande har den samme adgang til sparring hos deres respektive nationale myndigheder.

I Tyskland kræver det tilladelse fra nationale myndigheder at lagre og behandle lokaliseringsdata. Dette er ikke tilfældet i nogen af de andre lande, der indgår i denne undersøgelse. Tyskland har generelt en mere restriktiv lovgivning på området end andre europæiske lande; som tidligere nævnt indeholder den tyske lovgivning blandt andet restriktive krav til anonymisering af data, for eksempel ved at lokaliseringsdata skal rehashes (ændring af id-nummer) hvert døgn.

Rammer og barrierer 2/2

Pointe: Teleselskaber opleves af nogle LI-leverandører som tilbageholdende i forhold til at engagere sig i projekter, der indeholder lokaliseringsdata.

Forsigtighed blandt teleselskaber

Enkelte LI-leverandører oplever, at teleselskaberne har en forsigtig tilgang til at anvende lokaliseringsdata. Derudover er der enkelte eksempler på, at teleselskaberne overvurderer prisen på deres data.

Teleselskaberne opleves som usikre, i forhold til hvornår data er anonymiseret i tilstrækkelig grad, hvilke former for data de må videregive til LI-leverandører, og hvilke formål lokaliseringsdata må anvendes til. Derfor kan teleselskaberne være tilbøjelige til at afslå projektmuligheder af forsigtighedshensyn eller kræve, at LI-leverandørerne gennemgår omfattende og omkostningstunge due diligence-processer, inden de opstarter samarbejde med LI-leverandøren.

Derudover nævner en enkelt LI-leverandør, at teleselskaberne overvurderer værdien af de data, de ligger inde med, og de analysemuligheder, som lokaliseringsdata rummer i sig selv. LI-leverandøren peger derfor på, at der er behov for større forståelse blandt teleselskaberne for, at lokaliseringsdata skal gennemgå en omfattende struktureringsproces – en proces, som teleselskaberne sjældent selv har de faglige kapaciteter til at gennemføre – før de kan anvendes til analyseformål.

Dialog om aggregering

Teleselskaberne opleves som påpasselige i forhold til at arbejde med lokaliseringsdata, blandt andet fordi de er usikre på, i hvor høj grad data skal aggregeres,. Før de må udstilles. Orange i Spanien efterlyser derfor, at der er mere regelmæssig dialog mellem EU og de europæiske teleselskaber om, hvad det tilstrækkelige niveau for aggregering af data er.

Tre cases

Tre cases præsenterer eksempler på anvendelse af lokaliseringsdata. De tre cases illustrerer, hvilke muligheder og begrænsninger lokaliseringsdata rummer fra indsamling til anvendelse af data.

Tre cases

Hovedpointer



O2 og Transport for London: Project EDMOND

- O2's relativt store markedsandel var en medvirkende faktor i indsamlingen af repræsentative data.
- Ved hjælp af algoritmer og kombination af lokaliseringsdata med andre data, blandt andet fra det engelske rejsekort, formåede O2 at konstruere et fuldstændigt billede af transportmønstre i London.
- Transport for London anvendte de gennemførte analyser af lokaliseringsdata til at træffe beslutning om fremtidige investeringer i infrastruktur.



Positium og nationalbanken i Estland: Beregning af betalingsbalancen

- Positium har gennem en tiårig periode kontinuerligt leveret data og analyser til nationalbanken i Estland. Projektet adskiller sig fra de fleste andre projekter, hvor lokaliseringsdata anvendes, da der er tale om et fortløbende projekt.
- Selvom præcision generelt er vigtigt i forhold til anvendelse af lokaliseringsdata, har præcision mindre betydning for gennemførelsen af dette projekt, da data udelukkende har til formål at identificere, om mobiltelefonbrugere befinder sig indenfor eller udenfor Estlands grænser.



Orange, Kido Dynamics og Cintra: analyse af motorvejstrafik i Alicante

- Projektet illustrerer, hvordan lokaliseringsdata er et billigt og metodisk validt alternativ til andre metoder indenfor indsamling af positionsdata.
- Orange og Kido Dynamics' partnerskab var afgørende for, at teleselskabet Orange lagrede lokaliseringsdata, ved hjælp af Kido Dynamics' specialkompetencer, kunne blive struktureret til anvendelige analyser.

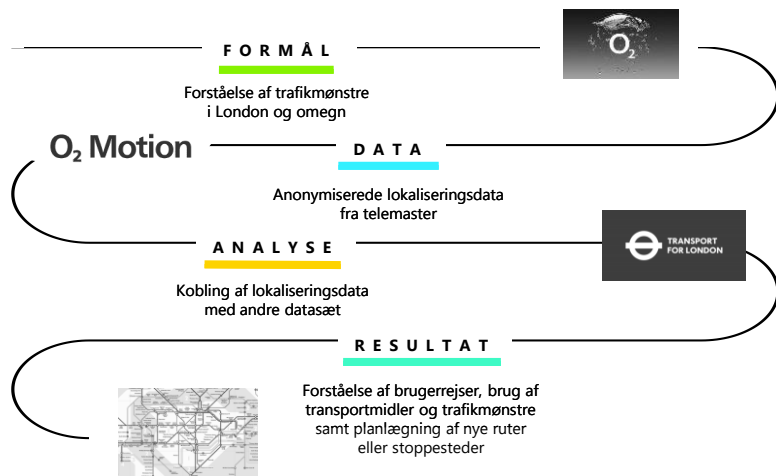
O2 og Transport for London: Project EDMOND

En case, hvor Transport for London brugte O2's lokaliseringsdata til bedre at forstå bevægelsesmønstre i trafikken blandt Londons borgere og besøgende

Baggrund for projekt

Med Project EDMOND (Estimating Demand using Mobile Network Data) ønskede Transport for London at opnå en bedre forståelse af bevægelses- og transportmønstre i Londons trafik med henblik på at kunne styre Londons infrastruktur på et oplyst grundlag. Transport for London indgik derfor i 2016 et samarbejde med O2 (Telefónica), der skulle levere analyser af aggregeret lokaliseringsdata. Analyserne skulle anvendes af Transport for London til at forstå borgernes ruter gennem trafikken og på den måde optimere ruterne ved eksempelvis at ændre busruter eller trafiksignalers indstillinger.

I dag har Transport for London selv udviklet en model til analyse af trafikvaner i London. Modellen baserer sig primært på indsamling af wi-fi-data blandt rejsende i busser og metro.



O2 er en udbyder af telekommunikationstjenester i Storbritannien, der blev grundlagt i 1985. I 2006 blev O2 opkøbt af spanske Telefónica.

Geografisk område

- Har 484 butikker i Storbritannien
- Kundedatabase på cirka 26 millioner
- En markedsandel på lidt over 35 procent af det britiske marked (2019), hvilket gør firmaet til det andenstørste teleselskab i Storbritannien
- Driver cirka 500 telemaster i Storbritannien.

Forretningsmodel

- Startede egen location intelligence-afdeling, O2 Motion, i 2012, som behandler og analyserer data
- Sælger data og analyser videre til både private virksomheder og offentlige instanser
- Leverer data til projekter indenfor transport, detailhandel, medier, marketing og turisme.



Transport for London er den offentlige myndighed, der har ansvaret for transportsystemet i London.

Arbejdsområder

- Implementerer Londons transportpolitik og forvalter byens transporttjenester
- Har ansvaret for Londons busser, undergrundsbane, sporvogne og flodbåde; derudover driver Transport for London også cirka 580 kilometer hovedveje, alle byens 4.600 trafiklys og taxiindustrien.

Erfaring med lokaliseringsdata

- Gik fra at bruge sideroad surveys til at benytte lokaliseringsdata til optimering og forståelse af Londons trafik og trafikanternes bevægelsesmønstre
- Startede med at bruge lokaliseringsdata i forbindelse med O2-samarbejdet i 2016
- Har sidenhen startet deres egen enhed op, som bruger en række andre datapunkter fra for eksempel wi-fi-netværket i deres transportmidler.

O2 og Transport for London: Project EDMOND



Pointe: De indsamlede data i Project EDMOND havde en præcision på cirka 50 meter. Derfor behøvede det indsamlede data at blive kombineret med andre datakilder. Grundet O2's relativt store markedsandel kunne de indsamlede data karakteriseres som en repræsentativ stikprøve af trafikanters bevægelser i London.

Dannelse og opsamling af data

Opsamling af data

Project EDMOND anvendte passive lokaliseringsdata i form af CDR-data opsamlet i en periode på tre måneder fra september 2016 til november 2016. CDR-data er lagrede data hos O2 som indeholder metadata om, hvordan en bestemt telefon bliver brugt. Det kan for eksempel være data om et bestemt opkald såsom opkaldets længde, lokation, tidspunkt og pris.

Præcision af data

De indsamlede lokaliseringsdata i Project EDMOND havde en præcision på cirka 50 meter, hvilket skyldes den store koncentration af telemaster i London. O2 forklarer dog, at deres lokaliseringsdata ikke kunne vise glidende bevægelser, hvilket skabte såkaldte ping-pong-effekter. Derfor var det for O2 vigtigt at koble deres lokaliseringsdata sammen med andre data for at øge præcisionen af det enkelte individs bevægelser og for at mindske risikoen for fejl.

Repræsentativitet

O2's markedsandel på 35 procent i England gør O2 til det andenstørste teleselskab i landet. Det betyder, at O2 kan indsamle en relativt stor stikprøve af de samlede bevægelser blandt britiske mobiltelefonbrugere. O2's lokaliseringsdataprojekter har hovedsagelig bestået af analyser af mobilitet og trafikmønstre. For at gå fra stikprøvedata til forudsigelser af hele befolkningens bevægelser opdeler O2 landet i forskellige zoner. De mindste zoner, de anvender, indeholder 2.000 husstande. Ifølge O2 bidrager deres zoneopdeling til at opnå et nøjagtighedsniveau på 90-95 procent, når de går fra deres indsamlede stikprøvedata til at udtale sig om hele befolkningens bevægelsesmønstre.



O2 og Transport for London: Project EDMOND



Pointe: O2 lagrede de pseudonymiseret CDR-data på en cloudbaseret server, der sikrede, at O2 havde fuld kontrol over data gennem hele projektet.

Lagring af data

Opbevaring af data

O2 lagrer de indsamlede data på deres egen platform på Amazons cloudbaserede server (AWS). Data forlader på intet tidspunkt platformen, hvorfor O2's kunder skal tilgå data gennem O2's platform, efter den er blevet aggregeret. På denne måde har O2 fuld kontrol over data gennem hele projektfasen. I Project EDMOND stod O2 dog selv for databehandlingen, hvorfor der ikke skulle gives adgang til nogen eksterne LI-leverandører.

Pseudonymisering

Når O2 modtager data fra telemasterne, bliver der som det første tilføjet en hashfunktion. Hashfunktionen fjerner alle personlige informationer, herunder navne, fra data, og i stedet tilføjes et hash for hver person. Det er derfor i denne proces, at den første anonymisering af lokaliseringsdata sker, og det er også her, O2's dataanalytikere begynder at behandle data. Når data har fået fjernet identifiere og tilføjet en hashfunktion, fremstår data pseudonymiseret.



O2 og Transport for London: Project EDMOND



Pointe: O2 anvendte algoritmer til at strukturere data til sammenhængende bevægelser og identificere anvendelsen af forskellige transportmidler. For at forbedre præcisionen af data supplerede O2 ligeledes aggregerede og anonymiserede lokalisingsdata med data fra briternes svar på Rejsekort.

Strukturering af data

Databehandling

O2's databehandling havde overordnet to formål:

1. At strukturere datapunkterne til bevægelsesmønstre blandt mobiltelefonbrugere
2. At identificere mobiltelefonbrugernes rejsemetode/transportmiddel

Første skridt i databehandlingen er at koble de individuelle datapunkter sammen til bevægelser fra A til B. En udfordring, som O2 stødte på i forbindelse med at strukturere data, var, at der på nogle strækninger af mobiltelefonbrugernes rejser ikke blev indsamlet lokalisingsdata. Dette problem opstod blandt andet, hvis en person rejste på togstrækninger, hvor der ikke kunne oprettes forbindelse til telemaster. O2 var i disse situationer nødt til at koble en række algoritmer på datasættene for at kunne udfylde hullerne i data i data og på den måde sammensætte sandsynlige bevægelsesmønstre. Ligeledes anvendte O2 20 forskellige algoritmer til at kategorisere, hvilke transportmidler mobiltelefonbrugere rejste med.

Datafusion

I Project EDMOND anvendte O2 data fra Oyster Card (Oyster-kort), der er briternes svar på Rejsekort. O2 forsynede Oyster-kortet med et simkort, så de kunne tracke folks rejser fra ind- til udcheckning.

Derfor var det muligt at koble data fra Oyster-kort med de lagrede CDR-data og dermed forbedre præcisionen af data. Ifølge O2 var denne datafusion et afgørende element i gennemførelsen af Project EDMOND, da lokalisingsdata uden sammenkobling med andre datakilder ikke ville give et fyldestgørende billede af befolkningens bevægelser gennem trafikken.

Aggregering af data

For at anonymisere data yderligere aggregerede O2 den pseudonymiserede data. Formålet med aggregeringen var at forhindre, at individuelle personer kunne identificeres ved at koble oplysninger om personernes bevægelser sammen.

Aggregeringen i Project EDMOND blev foretaget ved, at de individuelle observationer i datasættet blev grupperet til dataklumper bestående af individer, der befandt sig indenfor den samme lokalisingszone på det samme tidspunkt. Værdierne i O2's datasæt blev derfor generaliseret i en sådan grad, at hver enkelt person havde samme lokalisingsangivelse som resten af personerne indenfor den pågældende lokalisingszone.



O2 og Transport for London: Project EDMOND



Pointe: O2 overleverede både færdige analyser af data og datasæt, der gjorde det muligt for Transport for London selv at analysere de aggregerede data mere dybtgående.

Adgang til data

Overlevering af data til Transport for London

O2's data forlader, som tidligere nævnt, af sikkerhedsmæssige årsager ikke deres egen platform. I stedet blev Transport for London tilbudt forskellige overleveringsformater. I Project EDMOND blev O2's data overleveret til Transport for London i følgende formater:

- CSV-data
- Dashboards
- API

Transport for London fik ikke adgang til rådata og havde kun adgang til aggregeret data. Data vedrørende analyser af data med mindre end ti personer var automatisk frasorteret alle data, som Transport for London fik adgang til.

Opdateringsgrad

I forbindelse med Project EDMOND fik Transport for London overleveret data på én gang ved projektets afslutning, da Transport for London ønskede en generel analyse af folks bevægelsesmønstre i modsætning til en livemonitorering af mobilitet.

O2 kan potentielt levere data i nærrealtid med en forsinkelse på cirka 30 sekunder, hvilket dog ikke var relevant i forbindelse med Project EDMOND.



O2 og Transport for London: Project EDMOND



Pointe: Transport for London fik med Project EDMOND adgang til viden om nuværende rejsevaner i London, der blev anvendt til at træffe beslutninger om fremtidige investeringer i infrastruktur.

Anvendelse af data

Transport for London nævner særligt fire resultater fra Project EDMOND, som de indarbejdede i deres modeller og brugte til at træffe strategiske beslutninger om investeringer i Londons infrastruktur:

1. Kortlægning af brugen af offentlig transport i London: Project EDMOND gav Transport for London en præcis kortlægning af anvendelsen af offentlig transport i forskellige områder af London målt op mod andre transportmidler såsom bil, cykel og gåben. Resultatet viste, ikke overraskende, at en stor procentdel af indbyggerne benyttede offentlig transport i centrum af London, hvorimod bil var det foretrukne transportmiddel udenfor centrum.
2. Rejsevaner for non-Londoners: Transport for London anvendte Project EDMOND til at vise, at en større andel af non-Londoners end forventet transporterer sig selv i bil ind til London. Dette fik Transport for London til at gentænke deres strategi for at nedsætte bilforbruget i London og afdække muligheder for øgede investeringer i offentlig transport.
3. Rejsevaner for internationale turister: Project EDMOND har hjulpet Transport for London med at forstå transportmønstre blandt turister bedre. Med adgang til baggrundsoplysninger om mobiltelefonernes nationalitet gjorde Project EDMOND det muligt at differentiere internationale turisternes bevægelser fra britiske statsborgeres bevægelser i London. Transport for London kunne derfor, på baggrund af O2's analyser, identificere, hvilke bydele turister overnatter i, og hvordan turister kommer fra lufthavnene ind til London.
4. Kohorteanalyser: O2's data gjorde det muligt for Transport for London at segmentere deres brugere baseret på grupperinger af en række karakteristika og adfærdsmønstre. For eksempel foretog Transport for London videre analyser af forbrug af offentlig transport i hverdage kontra weekenderne. Analyserne viste, at pendlerne i mindre grad benytter offentlig transport som transportmiddel i weekenderne.



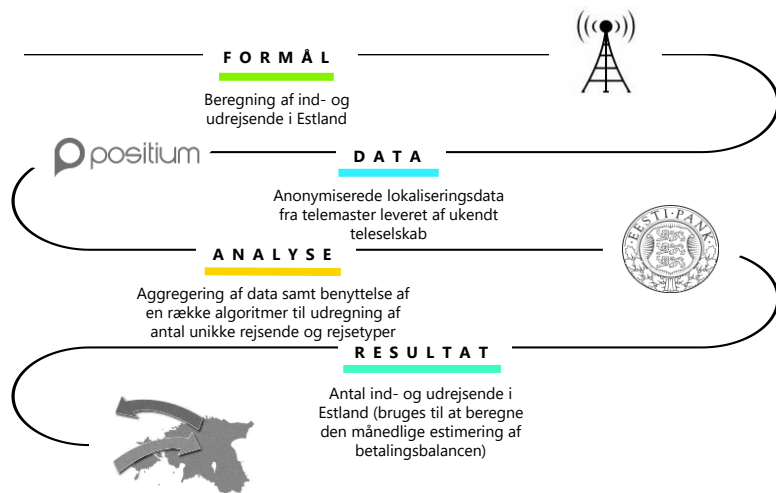
Positium og nationalbanken i Estland: beregning af betalingsbalance

I denne case anvendte nationalbanken i Estland i samarbejde med location intelligence-virksomheden, Positium, lokaliseringsdata fra mobiltelefoner til at beregne turismes indvirkning på den estiske betalingsbalance.

Baggrund for projekt

I 2009 besluttede nationalbanken i Estland at indgå et samarbejde med en af Europas førende location intelligence-virksomheder, Positium, i forbindelse med bankens ønske om at finde en ny, mere præcis måde at beregne antallet af ind- og udrejsende i Estland.

Efter indledende samtaler om mulig anvendelse af lokaliseringsdata til beregning af ind- og udrejsende i Estland blev Positium og nationalbanken i Estland enige om at iværksætte et pilotprojekt i perioden 2009-2010, hvor kun indrejsende skulle beregnes. Dette projekt blev betegnet som en succes, hvorfor udrejsende kom med i beregningen i 2010. Projektet er stadig i drift, og Positium leverer derfor nu på tiende år data til nationalbanken i Estland.



Positium blev etableret i 2006 i Estland af en række forskere fra University of Tartu som en spinoffvirksomhed. Fra universitetet

Geografisk område

- Primært Estland og Finland
- Har siden 2014 arbejdet udenfor Europa på markeder såsom Indonesien og Oman.

Brancher

- Transport, turisme, events.

Forretningsmodel

- Opkøb af pseudonymiseret rådata fra teleselskaber, som anvendes af Positium til at identificere bevægelsesmønstre
- Positiums produkter er enten enkeltstående analyser af bevægelsesmønstre eller kontinuerlige leveringer af lokaliseringsdata til private og offentlige kunder.

Projekter

- Positium har siden 2006 gennemført cirka 300 projekter.
- Fælles for projekterne er, at de hovedsageligt er foretaget for offentlige/offentligt ejede kunder såsom kommuner, turistorganisationer og udbydere af offentlig transport.



Nationalbanken i Estland blev grundlagt i 1919 (året efter Estlands uafhængighed). I 2011 blev Estland medlem af Eurozonen.

Arbejdsområder

- Opretholder stabiliteten i det estiske finansielle system og har i forlængelse heraf ansvar for at indsamle en lang række statistiske oplysninger relateret til Estlands økonomi, herunder oplysninger relateret til den estiske betalingsbalance.

Erfaring med lokaliseringsdata

- I samarbejde med Positium anvender nationalbanken lokaliseringsdata fra mobiltelefoner til at udregne ind- og udrejse i landet. Dataen indgår i beregningen af Estlands betalingsbalance.



Positium og nationalbanken i Estland: beregning af betalingsbalance

Pointe: Præcisionen af de indsamlede CDR-data havde ikke afgørende betydning for gennemførelsen af projektet, idet data udelukkende havde til formål at identificere, om mobiltelefonbrugeren befandt sig indenfor eller udenfor Estland.

Dannelse og opsamling af data

Opsamling af data

I dette projekt leveredes lokaliseringsdata af et estisk teleselskab. Grundet fortrolighedshensyn har det ikke været muligt at få oplyst, hvilket teleselskab der på nuværende tidspunkt leverer lokaliseringsdata. Tidligere har det estiske teleselskab EMT (med moderselskabet TeliaSonera) stået for leveringen af lokaliseringsdata.

Det pågældende teleselskab indsamlede passive lokaliseringsdata, som bestod af CDR'er. CDR'erne indeholdt blandt andet registreringer af:

- Tidspunkt for opkaldsaktivitet
- Tilfældigt genereret ID-nummer
- Telemast-ID med angivelse af koordinator for telemast
- Landet, som den anvendte telefon var registreret i.

Præcision af data

I dette projekt havde præcisionen af data en mindre betydning sammenlignet med andre lokaliseringsprojekter. Det skyldes, at nationalbanken i Estland udelukkende havde behov for at monitorere antallet af ind- og udrejsende i Estland, varigheden på deres rejser mv. Det var derfor ikke nødvendigt at kende mobiltelefonbrugernes eksakte lokation, men blot at vide, om mobiltelefonerne oprettede forbindelse til telemaster indenfor eller udenfor Estland.

I Estland har telemasterne en tæthed, der betyder, at teleselskabet kunne indsamle lokaliseringsdata med en præcision på 100-1000 meter i byområder, mens præcisionen i tyndtbefolkede områder var på 1,5-20 km.





Positium og nationalbanken i Estland: beregning af betalingsbalance

Pointe: Efter teleselskabet havde fjernet identifiere (personfølsomme oplysninger, der kan bruges til at identificere individer), var det stadig muligt at udlede, hvilke individer der indgik i data.

Lagring af data

Opbevaring af data

Da det ikke har været muligt at få oplyst, hvilket teleselskab der stod for indsamling og lagring af data i dette projekt, har det ikke været muligt at identificere, hvordan teleselskabet opbevarede de indsamlede lokaliseringsdata.

Pseudonymisering

Teleselskabet rensede data for alle identifiere. I stedet for, at de enkelte mobiltelefonbrugere fremgik med navn eller cpr-nummer, fremgik den enkelte bruger med sin IMSI (international mobile subscriber identity). IMSI er en unik værdi, som er tilknyttet hver individuel bruger af et mobilnetværk. Det var dog stadig muligt at sammenkæde observationer i datasættet og derigennem identificere individer. Data var med andre ord pseudonymiseret. Når data var blevet pseudonymiseret, blev de overdraget i denne form til Positium, der stod for den videre dataanalyse.





Positium og nationalbanken i Estland: beregning af betalingsbalance

Pointe: Data gennemgik en behandlingsproces, hvori udvalgte data frasorteredes, og data struktureredes efter ind- og udrejser. Først efter datastruktureringen var det muligt at fortolke data. For at sikre en høj datasikkerhed fulgte Positium en fastlagt kontrolprocedure og gav kun et udsnit af medarbejderne i Positium adgang til data, inden disse var blevet aggregeret.

Strukturering af data

Databehandling

For at gøre de individuelle datapunkter analyserbare grupperede Positium datapunkterne til konkrete ind- og udrejser. Denne databehandling foregik ved hjælp af en række algoritmer, som er udviklet af Positium. Ved hjælp af algoritmerne var Positium i stand til at identificere, hvilken type rejse der var tale om. De individuelle indrejser i Estland profileredes, så det blandt andet fremgik, om indrejserne var udtryk for transitrejser, fastboende udlændinge, endagsbesøgende eller flerdagsbesøgende. Derefter frasorterede Positium en række data, der kunne give et misvisende billede af antallet af ind- og udrejsende turister i Estland. Konkret frasorterede Positium nedenstående datapunkter:

- Lokaliseringsdata opfanget af telemaster i Estlands grænseområde
- Transitrejser, hvor den pågældende person opholdt sig kortere end en fastsat tidsperiode i Estland
- Langvarige ophold blandt udenlandske borgere, der havde opholdt sig i Estland minimum 91 dage indenfor de seneste 12 måneder.

Aggregering af data

Da de data, Positium modtog fra teleselskabet, stadig var

personhenførbare, var det kun få ansatte i Positium, der havde lov til at tilgå den pseudonymiserede data. De ansatte skulle følge regler, som er fastsat af Positium og beskriver, hvordan data kan tilgås, hvilket blandt andet betød, at data først måtte deles blandt Positiums resterende ansatte, når data var blevet aggregeret.

Ydermere fulgte Positium en såkaldt statistical disclosure control (SDC), der havde til formål at sikre, at data, der videregives til tredjeparter, er anonymiseret. En SDC er et sæt af regler og principper, der fastlægger, hvornår data er tilstrækkeligt anonymiseret til at kunne videregives til tredjeparter. SDC'en hjælper dermed Positium til at skabe klare linjer på tværs af projekter i forhold til, hvilke data der må videregives.

Det vigtigste element i Positiums SDC er, at data anonymiseres gennem aggregering. Ved aggregering ændres granulariteten af data, så data er mindre præcis. Dette gøres for eksempel ved at angive et interval, som personens alder befinder sig i, i stedet for at angive den præcise alder. Det har desværre ikke været muligt at få oplyst præcis, hvordan SDC'en beskriver Positiums fremgangsmåde i forbindelse med aggregering.





Positium og nationalbanken i Estland: beregning af betalingsbalance

Pointe: Nationalbanken i Estlands adgang til data understøtter dens behov for kvartalsvise beregninger af betalingsbalancen.

Adgang til data

Overlevering af data til nationalbanken i Estland

Positium anvender følgende metoder til at præsentere deres analysearbejde for nationalbanken i Estland:

- Datasæt i Excel
- Dashboards.

Formaterne understøtter, at nationalbanken hurtigt kan danne sig et overblik over udviklingen i Estlands turisme gennem dashboards og selv tilgå de datasæt, der ligger til grund for Positiums analyser.

Positium overleverer udelukkende datasæt indeholdende aggregeret og dermed anonymiseret data til Estlands nationalbank. Idet mobildata er blevet anonymiseret inden overlevering, har Positium ikke nogen restriktioner i forhold til, om data må offentliggøres, eller hvem Estlands nationalbank må dele data med. Nationalbanken i Estland udgiver i denne sammenhæng en rapport omhandlende Positiums analyser, der på kvartalsbasis bliver offentliggjort på nationalbankens hjemmeside.

Opdateringsgrad

Nationalbanken i Estland skal kunne opdatere beregningerne af betalingsbalancen hvert kvartal og har derfor behov for regelmæssige dataleveringer. Nationalbanken i Estland får derfor overleveret datasæt fra Positium samt opdateret deres dashboards én gang om måneden.

Positium tilbyder ikke realtidsmobildata til nationalbanken i Estland eller på andre af sine projekter.





Positium og nationalbanken i Estland: beregning af betalingsbalance

Pointe: Analyserne af lokaliseringsdata understøtter – på en omkostningslet og metodisk valid måde – nationalbankens behov for at kunne beregne turismes indvirkning på betalingsbalancen.

Anvendelse af data

Anvendelse

De data, der bliver leveret til nationalbanken i Estland, indeholder en række forskellige informationer om ind- og udrejsende i Estland. Følgende informationer er inkluderet i de data, Positium leverer til nationalbanken i Estland:

- Antal ind- og udrejsende
- Opholdenes varighed
- Antal unikke besøgende i Estland.

Disse informationer giver tilsammen nationalbanken i Estland et datagrundlag, der kan bruges til at beregne, hvilken indvirkning turisme har på betalingsbalancen. Konkret kobler nationalbanken i Estland oplysningerne om ind- og udrejsende med gennemsnitlige estimater for, hvor mange penge ind- og udrejsende forbruger. På denne måde er nationalbanken i Estland i stand til at estimere, hvor mange penge der går ind og ud af Estland som følge af turisme.

Nationalbanken i Estland ser flere fordele ved at anvende lokaliseringsdata. For det første er mobildata billigere at få adgang til end mere traditionelle måder at optælle ind- og udrejsende på. For det andet vurderer nationalbanken i Estland, at lokaliseringsdata har en høj validitet, og at lokaliseringsdata er repræsentative. Dette vurderer nationalbanken på baggrund af sammenligning med andre statistikker, der måler antallet af ind- og udrejsende, herunder opgørelser fra Talinn lufthavn, Talinn havn og opgørelser af antal krydsninger af Estlands østlige grænse.



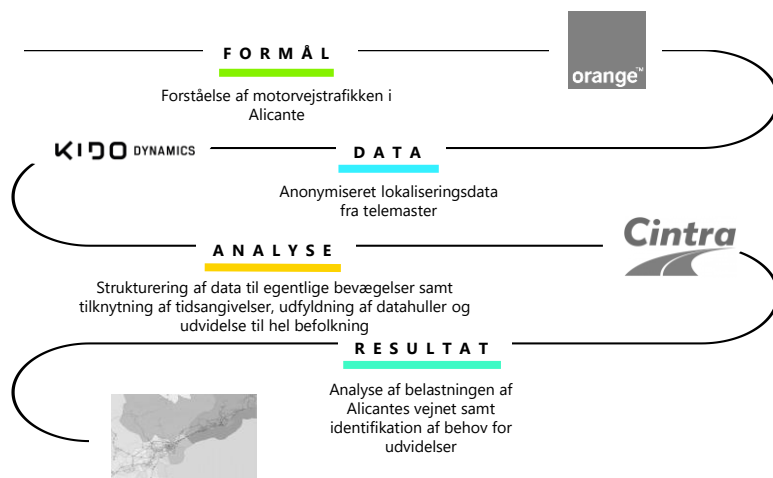
Orange, Kido Dynamics og Cintra: analyse af motorvejstrafik i Alicante

I denne case gennemførte Orange i samarbejde med Kido Dynamics en analyse af trafikken på vejnettet i og omkring Alicante i Spanien.

Baggrund for projektet

I 2017 udarbejdede Orange, i samarbejde med Kido Dynamics, en analyse af motorvejstrafikken i Alicante. De anvendte lokaliseringsdata fra mobiltelefoner til at analysere trafikken over strækningen i en periode på to måneder (august-oktober 2017). Modtageren af analysen var et privat spansk transport- og infrastrukturselskab, Cintra, som ønskede at forstå trafikmønstre for at kunne optimere strækningen med nye ruter eller flere spor.

Orange og Kido Dynamics indgik et partnerskab i forbindelse med projektet, hvor betalingen fra Cintra blev delt imellem de to parter. Cintra-projektet var et enkeltstående projekt, hvor Orange og Kido stod for at levere data og analyser, men ikke var involveret i den videre proces omkring, hvad Cintra besluttede på baggrund af datagrundlaget.



Orange blev grundlagt i 1994 og senere opkøbt af France Télécom, som i 2006 skiftede navn til Orange. I dag er teleselskabet det tredjestørste i Europa.

Geografisk område

- Arbejder i 27 lande med 174 millioner kunder i hele verden
- Orange i Spanien har en markedsandel på lidt over 25 procent af det spanske marked (2019).

Forretningsmodel

- Orange startede med projekter indenfor lokaliseringsdata i 2015.
- Orange behandler og analyserer ikke selv data, men har indgået partnerskaber med cirka ti LI-leverandører.
- Orange leverer primært data til projekter indenfor transport, turisme, marketing og detailhandel.



Kido Dynamics blev grundlagt i år 2018 som en virksomhed med fokus på at gennemføre analyser af bevægelser og mobilitet.

Geografisk område

- Hovedkontor i Schweiz med 14 ansatte
- Gennemfører projekter i otte forskellige lande, herunder Spanien.

Forretningsmodel

- Kido Dynamics arbejder ud fra en *revenue-shared agreement*, hvor de deler deres omsætning for det givne projekt med det teleselskab, de arbejder sammen med.
- Kido Dynamics udarbejder analyser for offentlige og private kunder indenfor transport, smart cities, turisme og sikkerhed.



Cintra blev grundlagt i 1998 i Spanien og er i dag en af verdens førende transport- og infrastrukturvirksomheder.

Arbejdsområder

- Innovative løsninger indenfor optimering af motorvejsudvikling
- Opkrævning af vejafgifter
- Ejer over 300.000 parkeringspladser i Spanien.

Erfaring med lokaliseringsdata

- Cintra har anvendt mobile positioning-data til at monitorere anvendelsen af infrastruktur, herunder for at vurdere det kommercielle potentiale forbundet med at anlægge nye veje, hvor der opkræves vejafgifter.



Orange, Kido Dynamics og Cintra: analyse af motorvejstrafik i Alicante

Pointe: For aktører, der ønsker at kortlægge trafikmønstre hurtigt og effektivt, er lokaliseringsdata et billigere alternativ end andre traditionelle dataindsamlingsmetoder. CDR-data indeholder dog fejkilder, som skal udbedres, før data viser et retvisende billede af mobilitetsmønstrene.

Dannelse og opsamling af data

Opsamling af data

I forbindelse med analysen af motorvejstrafik i Alicante valgte Cintra, i samarbejde med Orange og Kido Dynamics, at anvende lagrede passive lokaliseringsdata i form af CDR-data indsamlet i området omkring Alicante. Gennem CDR-data kunne Orange identificere, hvor mange telefoner der oprettede forbindelse til telemaster omkring hovedveje samt hvornår og derigennem udlede anvendelsen af vejnettet i området.

Tidligere analyser, der har haft til formål at afdække data om den spanske motorvejstrafik, har typisk været gennemført af den spanske stat eller af private infrastrukturvirksomheder såsom Cintra og har primært baseret sig på spørgeskemadata. Denne form for dataindsamling kan koste op mod 2 millioner euro, hvis den skal præsentere et fyldestgørende billede af trafikmønstre.

Præcision af data

Orange har en præcision på sine lokaliseringsdata på ned til 50 meter, som dog varierer alt efter tætheden af telemaster. I forbindelse med dette projekt indeholdt de indsamlede CDR-data dog en række usikkerheder, som Kido var nødt til at håndtere i den senere datastrukturering. De primære usikkerheder omfattede:

- Nedetider for telemaster, så mobiltelefoner ikke oprettede forbindelse til nærmeste telemast
- At hver begivenhed i telenettet ikke blev opfanget i kronologisk rækkefølge og derfor skulle tilbageføres kronologisk
- At de samme begivenheder registreredes mere end én gang og derfor fremstod som flere begivenheder, end de reelt var udtryk for.





Orange, Kido Dynamics og Cintra: analyse af motorvejstrafik i Alicante

Pointe: For at sikre høj datasikkerhed videregiver Orange aldrig rådata til tredjeparter. Derudover fjerner Orange identifiere og opdeler lokationer i zoner, før tredjeparter får adgang til data gennem Oranges cloudbaserede platform.

Lagring af data

Opbevaring af data

Orange opbevarer sine opsamlede lokaliseringsdata i egen Orange-platform på Amazon Cloud. Den pseudonymiseret data på individplan forlader ikke platformen, hvorfor Kido skulle tilgå Oranges platform for at få adgang til data.

Pseudonymisering

Orange grupperer typisk deres data i geografiske zoner. Af anonymiseringshensyn arbejder Orange ud fra en regel om, at minimum 25 personer skal være indenfor den givne zone på det samme tidspunkt for, at data lagres. Hvis dette ikke er tilfældet, er Orange nødsaget til at udvide zonen, indtil minimumsreglen om 25 personer er opfyldt.

For at fjerne alle personfølsomme oplysninger fjerner Orange alle identifiere gennem pseudonymisering af data. Efter denne anonymisering er data rensset for navne, cpr-numre og lignende. Det er dog stadig muligt for Orange og andre aktører, som har adgang til data, at udlede, hvilke individer der er inkluderet i data. Efter pseudonymiseringen er Oranges data klar til at blive tilgået af Kido, som strukturerer data yderligere.





Orange, Kido Dynamics og Cintra: analyse af motorvejstrafik i Alicante

Pointe: Orange overdrager selve databehandlingen til Kido Dynamics, der både har de rette faglige ressourcer og viden om kundernes behov til at strukturere data til brugbare analyser af trafikmønstre. For at mindske risikoen for at kunne identificere individer i datasættet anvender Kido Dynamics de to metoder *differential privacy* og aggregering.

Strukturering af data

Databehandling

Når Orange har lagret data, overtager Kido den videre databehandling. Kidos databehandlingsproces består af fem trin:

1. Filtrering og strukturering af data: Først analyserer Kido sammenhængen mellem de forskellige, individuelle datapunkter og strukturerer dem til egentlige bevægelser fra punkt A til B.
2. Genopbygning af huller i bevægelsesmønstre: Kido anvender maskinlæring til at udfylde de huller, der er i registreringen af de enkelte mobiltelefonbrugeres bevægelser. Kido har udarbejdet en rekonstruktionsalgoritme, som beregner den mest sandsynlige, fuldstændige rute for hver bruger på baggrund af de data, der er til rådighed.
3. Udvidelse til den samlede befolkning: Med viden om dækningsgrad og ved hjælp af statistiske data vedrørende demografien i området tildeler Kido de forskellige observationer i datasættet en vægtning, så data opnår en højere grad af repræsentativitet.
4. Simulering af bevægelsesmønstre i det rigtige vejnet: Simuleringen giver oplysninger om, i hvor høj grad forskellige ruter anvendes, hastigheden af

trafikanternes bevægelser på de forskellige ruter samt en tidsplan for, hvornår ruterne er henholdsvis mest og mindst belastede.

5. Kalibrering af data: Til sidst sammenlignes simuleringerne med andre tilgængelige data, der måler intensiteten af anvendelsen af forskellige trafikruter. Simuleringerne kalibreres, så diskrepansen mellem simuleringerne og de supplerende statistiske kilder mindskes.

Differential privacy og aggregering

Som en del af Kidos databehandlingsproces går data fra pseudonymisering til anonymisering. Anonymiseringen sker gennem en kombination af differential privacy og aggregering. Ved differential privacy-metoden tilføjes støj systematisk til data. Eksempelvis tilføjer Kido falske personer til datasættet, og køn samt alder ændres for nogle af mobiltelefonbrugerne. Data ændres dermed på individniveau, men viser stadig de samme overordnede bevægelsesmønstre.

Dernæst aggregerer Kido både individernes baggrundsoplysninger og foretager en geografisk aggregering. Dermed grupperes eksempelvis mobiltelefonbrugernes aldre i intervaller, og der vises udelukkende, hvilke zoner brugerne befinder sig i - ikke deres præcise lokation.





Orange, Kido Dynamics og Cintra: analyse af motorvejstrafik i Alicante

Pointe: Cintra får adgang til dynamiske analyser af data samt datasæt, virksomheden selv kan foretage videre analyser af.

Adgang til data

Overlevering af data til Cintra

Kido overleverer data til Cintra. Disse data er rensset for identifiere og har gennemgået en differential privacy- og aggregeringsproces. I dette projekt blev anonymiseret data overleveret til Cintra i følgende formater:

- CSV-filer
- GIS-data (SHP-filer)
- Interaktive visualiseringer af GIS-data.

Med CSV-filerne og SHP-filerne kan Cintra selv tilgå de data, der ligger til grund for Kidos analyser, og foretage videre analyser af dem. CSV-filerne viser de grundlæggende data, efter de er blevet samlet til egentlige bevægelser i trin 2 af Kidos databehandlingsproces. GIS-data viser de færdigbehandlede data i et tabelformat.

For at illustrere hovedkonklusionerne i Kidos analyser overleverer Kido ligeledes interaktive visualiseringer af, hvordan vejnettet i Alicante anvendes. Visualiseringerne har et dashboardlignende format, og det er muligt for Cintra at klikke sig ind på de enkelte veje og få præsenteret informationer om deres anvendelse, brug af tilfartsveje mv.

Opdateringsgrad

Projektet for Cintra var et enkeltstående projekt, hvor data blev indsamlet over en periode på to måneder og præsenteret i afslutningen af projektperioden. Da formålet med projektet var at foretage en generel analyse af det daværende trafiknet, havde Cintra ikke behov for at få adgang til realtidsdata eller løbende overleveringer af data.





Orange, Kido Dynamics og Cintra: analyse af motorvejstrafik i Alicante

Pointe: Cintras valg af lokaliseringsdata var medvirkende til, at Cintra på en effektiv og fleksibel måde kunne kortlægge behovet for udvidelse af trafiknettet med afsæt i repræsentative og valide data.

Anvendelse af data

Anvendelse

Cintras primære formål med at gennemføre en analyse af trafikmønstrene i Alicante-området var at undersøge den trafikale belastning af vejnettet og derigennem identificere, om der var behov for en udvidelse af det eksisterende vejnet og i så fald hvor. Ligeledes kunne Cintra anvende materialet til at prioritere, hvor der først skulle ske en udvidelse af vejnettet.

Det er her værd at bemærke, at Orange og Kido ikke var involveret i den videre beslutningsproces omkring opdatering af vejnettet, men blot stillede videngrundlaget til rådighed for Cintra.

Orange og Kido peger selv på en række operative, metodiske og strategiske fordele ved at anvende lokaliseringsdata i gennemførelsen af projektet:

- Gennem lokaliseringsdata er det muligt, på en relativt omkostningslet og hurtig måde, at indsamle informationer, der giver et fyldestgørende billede af trafikanters bevægelser.
- Data kan enkeltsegmenteres på variable såsom køn, oprindelsespunkt, destination, hastighed, frekvens og tidspunkt.
- Det er ikke nødvendigt at definere krav til output, inden projektet starter. Det kan besluttes undervejs i projektet, hvilke analyser der er mest relevante at udføre.
- Beslutningstagere får et mere repræsentativt og validt datagrundlag end med andre metoder til kortlægning af mobilitet.



Minicases

To cases, der giver et hurtigt indblik i anvendelsen af lokaliseringsdata til kommercielle formål

Vodafone og Carto: mobilitetsanalyse

Nedenfor præsenteres en case, hvor Vodafone i samarbejde med location intelligence-virksomheden Carto udarbejdede en mobilitetsplatform i Barcelona til brug for virksomheder og offentlige organisationer.

I et partnerskab mellem Vodafone og Carto blev en end-to-end-platform udarbejdet med det formål at vise nuværende mobilitetsmønstre og forudsige fremtidige bevægelsesmønstre i Barcelona.

Vodafone er Spaniens tredjestørste teleselskab med omkring 14 millioner kunder, hvilket svarer til cirka 26 procent af det spanske marked. Teleselskabet kunne med en betydelig markedsandel derfor indsamle repræsentative data, der blev overleveret til Carto i anonymiseret form, klar til videre databehandling.

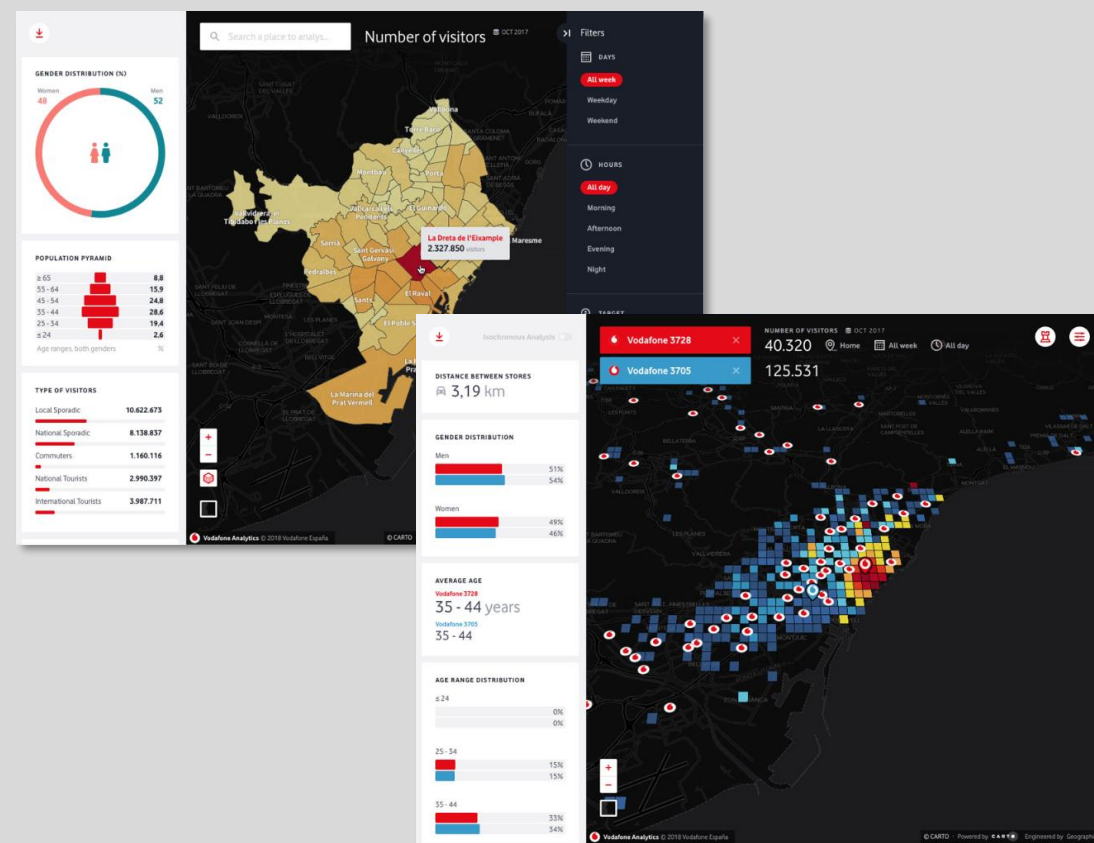
Carto stod for struktureringen af data og visualisering af bevægelsesmønstre i et dynamisk dashboardplatform (illustreret til højre). Blandt andet sammenkoblede Carto lokaliseringsdata med en række andre datapunkter for at kunne segmentere resultaterne på baggrund af køn, alder, nationalitet og status som fastboende eller turist.

Platformen blev udviklet til både private virksomheder og offentlige organisationer, der vil forstå, hvordan folk bevæger sig i Barcelona, og hvorfor de bevæger sig, som de gør, samt til at forudsige mobilitet i fremtiden til for eksempel events, fodboldkampe, koncerter mv. En af brugerne af denne platform var Barcelona kommune, som brugte platformen til følgende:

- Analyse af byens trafikmønstre: analyse af, hvor der hyppigst opstår trafikproblemer i byen. Platformen blev brugt i drøftelser af eventuelle nye infrastrukturløsninger for byen.
- Analyse af turisme i byen: opgørelse af antal unikke besøgende og deres nationalitet. Platformen blev brugt til at forstå turisternes bevægelsesmønstre for at forbedre oplevelsen for turister, når de bevæger sig rundt i Barcelona.
- Forbedring af affaldsopsamling: Med det formål at gøre Barcelona til en miljøvenlig og ren by blev de mest besøgte områder i byen identificeret med henblik på at intensivere affaldsopsamlingen der.



CARTO



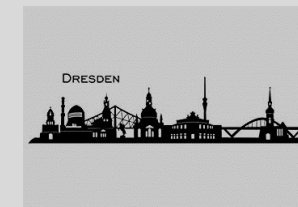
Motionlogic og Dresden

Nedenfor præsenteres en case, hvor location intelligence-virksomheden Motionlogic udarbejdede analyser vedrørende det årlige julemarked i den tyske by Dresden.

Dresdens julemarked er et af Europas største og mest berømte med dets cirka tre millioner årlige besøgende. I 2018 brugte arrangørerne af Dresdens julemarked location intelligence-virksomheden Motionlogic – et spinoff-firma af Deutsche Telekom – til at forstå de besøgendes bevægelsesmønstre bedre. Ydermere ønskede de at få indblik i antallet af besøgende, samt hvor de besøgende kom fra i verden. Projektet blev gennemført som et pilotprojekt og foregik i perioden fra 27. november til 24. december i 2018, hvor julemarkedet har åbent.

Motionlogic baserede projektet på passive lokaliseringsdata indsamlet af Deutsche Telekom. Som teleskab havde Deutsche Telekom ansvaret for at stå for anonymiseringen af data, inden Motionlogic fik adgang til dem. Det skyldes, at der i Tyskland er der særligt strenge regler for anonymisering af data, som Deutsche Telekom, i kraft af deres status som teleselskab, har ansvaret for bliver overholdt. Deutsche Telekom skulle derfor følge en fast anonymiseringsproces udarbejdet af de tyske myndigheder, der blandt andet foreskriver, at data skal rehashes én gang i døgnet, og at data skal indeles i perioder a én time. Derudover skal firmaer, der ønsker at arbejde med lokaliseringsdata, søge tilladelse til dette hos myndighederne.

For at teste nøjagtigheden af de indsamlede lokaliseringsdata blev data sammenlignet med manuelle optællinger af personer på julemarkedet. De manuelle optællinger blev foretaget 25 gange i løbet af projektperioden. Konklusionen fra optællingerne var, at lokaliseringsdata alene var en usikker kilde til opgørelse af antallet af personer indenfor et begrænset geografisk område med ingen faste ind- og udgangspunkter. Det var derfor nødvendigt at kombinere lokaliseringsdata med de manuelle optællinger og anvende Motionlogics algoritmer til at skabe et mere præcist billede af de besøgendes færden på julemarkedet.



Bilag 1

Identificerede cases

Bruttoliste over cases

Introduktion

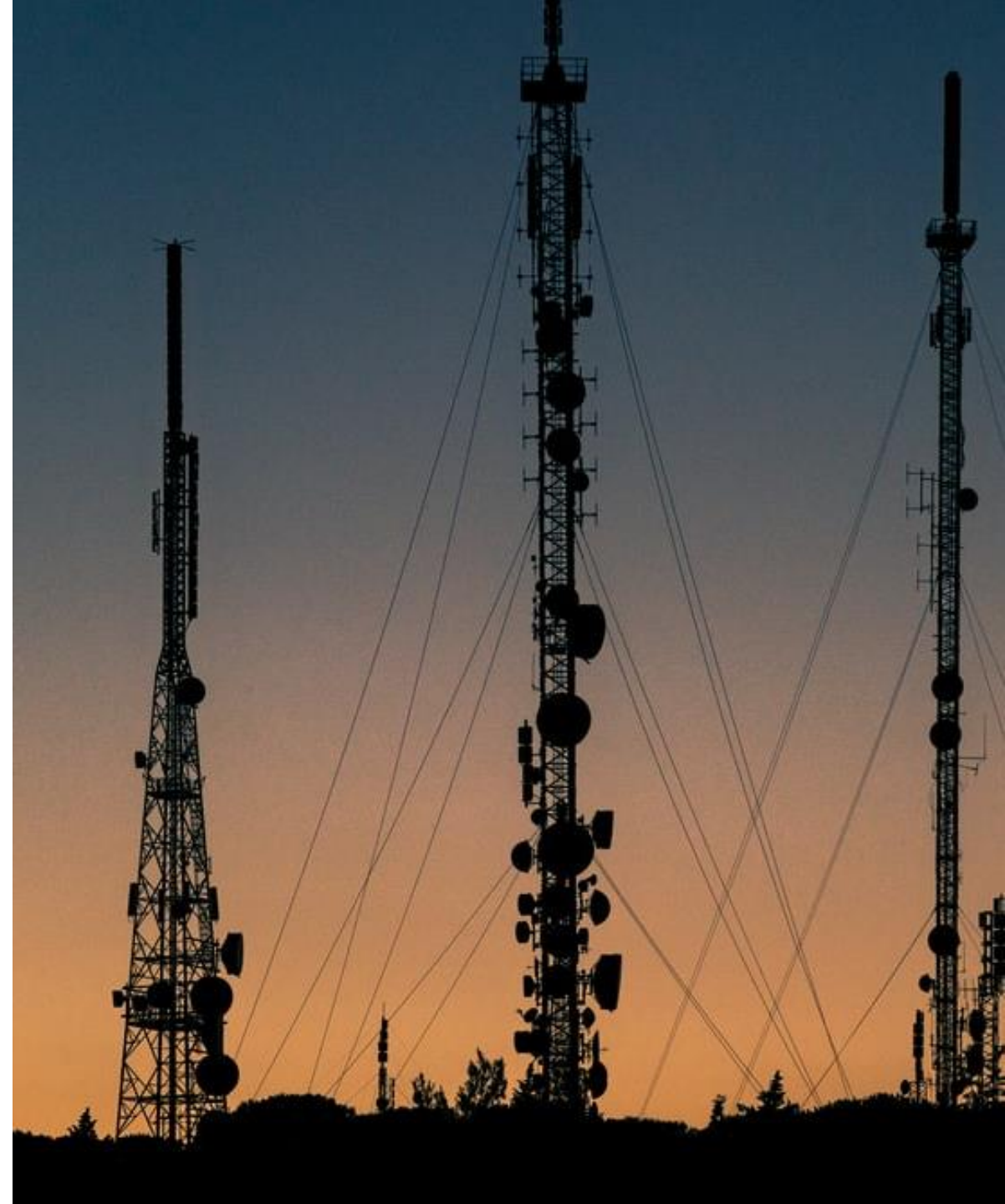
Oplisting af cases

Som led i denne undersøgelses indledende afdækning af anvendelsen af lokaliseringsdata er der identificeret 34 cases, der viser, hvordan lokaliseringsdata anvendes i kommercielt øjemed.

På de følgende sider er de identificerede cases oplistet. Casene er opdelt efter, hvilken branche de tilhører. Først præsenteres cases indenfor transport, detailhandel, turisme, kultur og den finansielle sektor. Til sidst er der en gennemgang af de teleselskaber og firmaer, der udbyder lokaliseringsplatforme og -løsninger af forskellig karakter.

Samtidig er de identificerede cases inddelt i tre typer på baggrund af, hvilken aktør der er involveret i casen:

- **Creators** er virksomheder, som udvikler teknologier, der muliggør indsamling og anvendelse af lokaliseringsdata. Creators vil typisk være teleselskaber.
- **Enablers** er virksomheder, som behandler, strukturerer og samkører lokaliseringsdata og på den måde omdanner data fra individuelle datapunkter til analyser af sammenhængene mobilitetsmønstre. Enablers vil typisk være såkaldte LI-leverandører, men kan også være teleselskaber. LI-leverandører – eller leverandører af location intelligence – er firmaer med særlig ekspertise i bearbejdning af lokaliseringsdata. De samarbejder ofte med teleselskaber om at omdanne de rå lokaliseringsdata til platforme og koncepter for anvendelse af mobildata.
- **Operators** er virksomheder eller offentlige organisationer, der anvender lokaliseringsdata til yderligere raffinering af eksisterende produkter eller distribution/produktion af disse. Operators kan med andre ord karakteriseres som aftagere af lokaliseringsprodukter.



Cases – transport

Identificerede cases indenfor transport

Nr.	Land	FIRMA	CASE	PERIODE	TYPE
1	Storbritannien	Transport for London (TfL)	Project EDMOND (Estimating Demand using Mobile Network Data): forudsigelse af trafikmønstre ved hjælp af mobildata knyttet sammen med andre komplementære datasæt. TfL har også gennemført et pilotprojekt med teleselskabet O2 om bevægelsesmønstre blandt metropassagerer.	2016	Operator
2	Storbritannien	Immense	Immense: joint partnership med O2 om at anvende lokaliseringsdata i forbindelse med mobilitetsundersøgelser.	2016-nu	Enabler
3	Schweiz	Teralytics	Brug af lokaliseringsdata fra mobiltelefoner til at hjælpe offentlige instanser samt private virksomheder med at analysere bevægelsesmønstre og udvikle infrastrukturer og den offentlige transport.	2012-nu	Enabler
4	Tyskland	Deutsche Bahn	Analysen af transportmønstre gennem Telefónica NEXTs Transport Analytics.	Lukket i 2019*	Operator
5	Tyskland	München lufthavn	Pilotprojekt med Telefónica NEXT om monitorering af bevægelsesmønstre i lufthavnen.	Lukket i 2019*	Operator
6	Tyskland	Deutsche Telekom Motionlogic GmbH	Levering af analyser af trafikmønstre. Har blandt andet samarbejdet med selskaber inden for offentlig transport.	2013-nu	Creator/ enabler
7	Estland	Bystyret i Tartu	Analyse og optimering af offentlig transport i Tartu med afsæt i en kombination af mobil- og rejsekortsdata.	2019	Operator
8	Internationalt	TomTom	Partnerskab om anvendelse af Vodafone-data i TomToms navigationstjenester i en lang række europæiske lande.	Igangværende	Operator

* Projektet blev gennemført i samarbejde med Telefónica NEXT, der var Telefónica Tysklands datterselskab indenfor udnyttelse af lokaliseringsdata. Telefónica NEXT lukkede i 2019 grundet Telefónicas ønske om at prioritere flere ressourcer til udvidelse af 5G-netværket.

Cases – detailhandel

Identificerede cases indenfor detailhandel

Nr.	Land	FIRMA	CASE	Periode	TYPE
9	Spanien	Vodafone Analytics	I samarbejde med Carto og Geographica blev der udviklet en platform, der måler turisternes bevægelsesmønstre i Barcelona, samt hvornår butikker modtager flest besøgende.	2019	Operator
10	Spanien	Ralph Lauren	Teleselskabet Orange og analysefirmaet Ipsos foretog en undersøgelse af, hvor det bedst kunne betale sig for Ralph Lauren at anbringe deres reklamer i Madrid og Barcelona baseret på antallet af forbipasserende ved forskellige reklameskærme.	2019	Operator
11	Tyskland	Telefónica NEXT	I Tyskland hjalp Telefónica NEXT virksomheder med løsninger indenfor smart media og detailhandel. De udviklede blandt andet løsninger, hvor beskeder og reklamer blev sendt målrettet til personer baseret på deres lokation. Disse blev blandt andet anvendt af Coca-Cola, Audi og Commerzbank. Derudover udviklede de også løsninger til måling af mobiltrafik omkring butikker og leverede socioøkonomiske oplysninger om besøgende i butikken. Telefónica NEXT lukkede i 2019 grundet Telefónicas ønske om at prioritere flere ressourcer til udvidelse af 5G-netværket.	Lukket i 2019	Enabler
12	Estland	Positium, Newbase	Pilotprojekt udviklet til shoppingcentre, detailkæder, fastfoodrestauranter mv. for at optimere marketingaktiviteter og planlægge kampagner bedre på baggrund af lokaliseringsdata fra mobiltelefoner.	2012	Enabler
13	Israel	TrendIT, CE-Traffic	På baggrund af lokaliseringsdata fra mobiltelefoner leverede TrendIT realtidsdata om menneskers bevægelsesmønstre, adfærd og andre karakteristika fra specifikke lokationer såsom butikker og supermarkeder. Analyser af data blev solgt til offentlige instanser, detailhandel, marketingbureauer mv.	2010-2017	Enabler

Cases – turisme, kultur og finansiel sektor

Identificerede cases indenfor turisme, kultur og finansiel sektor

Nr.	Land	FIRMA	CASE	Periode	TYPE
14	Tyskland	Hamburg Marketing	En af de første mobilitetsanalyser af turisternes bevægelsesmønstre i Tyskland foretaget i samarbejde med Deutsche Telekom.	2017-2018	Operator
15	Tyskland	German National Tourist Board	Analyse af bevægelsesmønstre blandt turister fra seks store turistgrupper i Tyskland, blandt andet Brasilien, Rusland og Kina.	-	Operator
16	Tyskland	City of Dresden	Analyse af bevægelsesmønstre blandt besøgende på julemarkeder i Dresden.	2018	Operator
17	Estland	Nationalbanken i Estland	I samarbejde med Positium bruger nationalbanken i Estland lokaliseringssdata fra mobiltelefoner for at få bedre overblik over turisme i Estland. data anvendes i beregningen af Estlands betalingsbalance.	2009-nu	Operator
18	Estland/Finland	Statistics Estonia, Positium, Statistics Finland og Finnish Transport Infrastructure Agency	Opgørelse af antallet af rejsende i Estland og Finland ved hjælp af lokaliseringssdata fra mobiltelefoner.	2014	Operator
19	Tjekkiet	CzechTourism	For at forbedre turismestrategier og marketingkampagner for Tjekkiet brugte CzechTourism lokaliseringssdata fra mobiltelefoner ved 45 af de mest besøgte attraktioner til bedre at forstå turisternes bevægelsesmønstre.	2012-2013	Operator
20	Storbritannien	Olympics 2012 i London	Monitorering af bevægelsesmønstre for folkemængder, belastninger ved ind- og udgange samt monitorering af transport af ansatte.	2012	Operator

Cases – udbydere af lokalisingsdataløsninger

Identificerede cases blandt udbydere af lokalisingsdataløsninger

Nr.	Land	FIRMA	CASE	Periode	TYPE
21	Storbritannien	O2 Motion	Teleselskabet O2's innovationsteam, der arbejder med produktudvikling indenfor salg af lokalisingsdata.	2012-nu	Creator/ enabler
22	Spanien	Nommon, Kineo Mobility Analytics	Bruger lokalisingsdata fra mobiltelefoner til at analysere bevægelsesmønstre i forbindelse med transport og turisme samt til geomarketing. Har både offentlige og private kunder, herunder McDonald's.	2015-nu	Enabler
23	Spanien	Orange	Orange har indgået partnerskaber med en lang række LI-leverandører, herunder Kido Dynamics, om udnyttelse af lokalisingsdata. Orange har sammen med partnervirksomheder blandt andet gennemført projekter indenfor transport, turisme, detailhandel og marketing.	2015-nu	Creator
24	Internationalt/Spanien	Carto og Geographica	Carto er en af verdens største udbydere af location intelligence-platforme, der blandt andet benytter mobildata til analyser indenfor transport og mobilitet. Carto har opkøbt det spanske firma Geographica, der tilbyder løsninger indenfor big data og kunstig intelligens, herunder anvendelse af mobildata.	2012-nu	Enabler
25	Schweiz	Kido Dynamics	Kido Dynamics bruger lokalisingsdata fra mobiltelefoner i samspil med andre data til at forudsige mobilitetsmønstre. Deres analyser er målrettet turisme, transport, geomarketing, smart cities, detailhandel og ejendomsmarkedet.	2018-nu	Enabler
26	Estland	Positium	Positium leverer en teknologisk platform og metode til behandling af mobildata for at hjælpe beslutningstagere med at forstå menneskelig mobilitet. data bliver brugt til analyser af turisme marketing, byplanlægning, sikkerhedsanalyser, analyse af bevægelsesmønstre mv.	2002-nu	Enabler

Bilag 2

Metoder til anonymisering af lokaliseringsdata

Introduktion til brug af konkrete anonymiseringsteknikker

Ved anonymisering af data skal de registeransvarlige tage højde for særligt tre risici

Risici ved anonymisering

Selvom man anvender flere anonymiseringsmetoder, kan der stadig opstå risici for, at en udefrakommende kan identificere en person eller en værdi, såfremt personen ved, hvad der skal kigges efter. Det drejer sig særligt om tre forskellige risici, som er illustreret og beskrevet nedenfor. Disse tre risici er de hyppigste årsager til, at personer bliver identificeret i datamængden.

Man skal derfor sikre, at de anvendte anonymiseringsmetoder er nøje gennemtænkte og anvendt på en måde, som minimerer risikoen for, at der kan ske en udskilning, sammenkobling eller udledning.



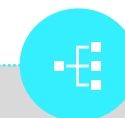
Udskilning

Den første risiko omhandler muligheden for at udskille nogle af eller alle de poster i et datasæt, som kan identificere en bestemt person. Det vil sige, at man gennem isolering af rækker i et datasæt kan finde frem til en bestemt person.



Sammenkobling

Den anden risiko omhandler muligheden for at koble mindst to poster vedrørende den samme registrerede, eller en gruppe registrerede, sammen. Det vil sige, at man igennem sammenkobling af flere poster kan finde frem til en bestemt person eller en bestemt gruppe personer.



Udledning

Muligheden for at man med stor sandsynlighed kan udlede værdien af én bestemt attribut i et datasæt ud fra værdierne af et andet sæt attributter. Det vil sige, at man igennem sammenholdning af bestemte attributværdier kan finde frem til en anden bestemt attributværdi i datasættet.

Formål og kontekst for anvendelse er afgørende

Ingen anonymiseringsmetoder er uden ulemper, og for at imødekomme de tre identificerede risici skal man derfor sikre robusthed mod genidentificering ved at udvælge metoder, som passer til formål og kontekst for anvendelsen af datasættet.

Der påhviler de registreringsansvarlige et ansvar, da det er dem, som skal sikre, at anonymiseringsmetoderne anvendes korrekt og ansvarligt.

Introduktion til brug af konkrete anonymiseringsteknikker

Der er fem konkrete teknikker til anonymisering, som enten mindsker nøjagtigheden af data eller skalerer værdier

De præsenterede anonymiseringsmetoder omfatter:

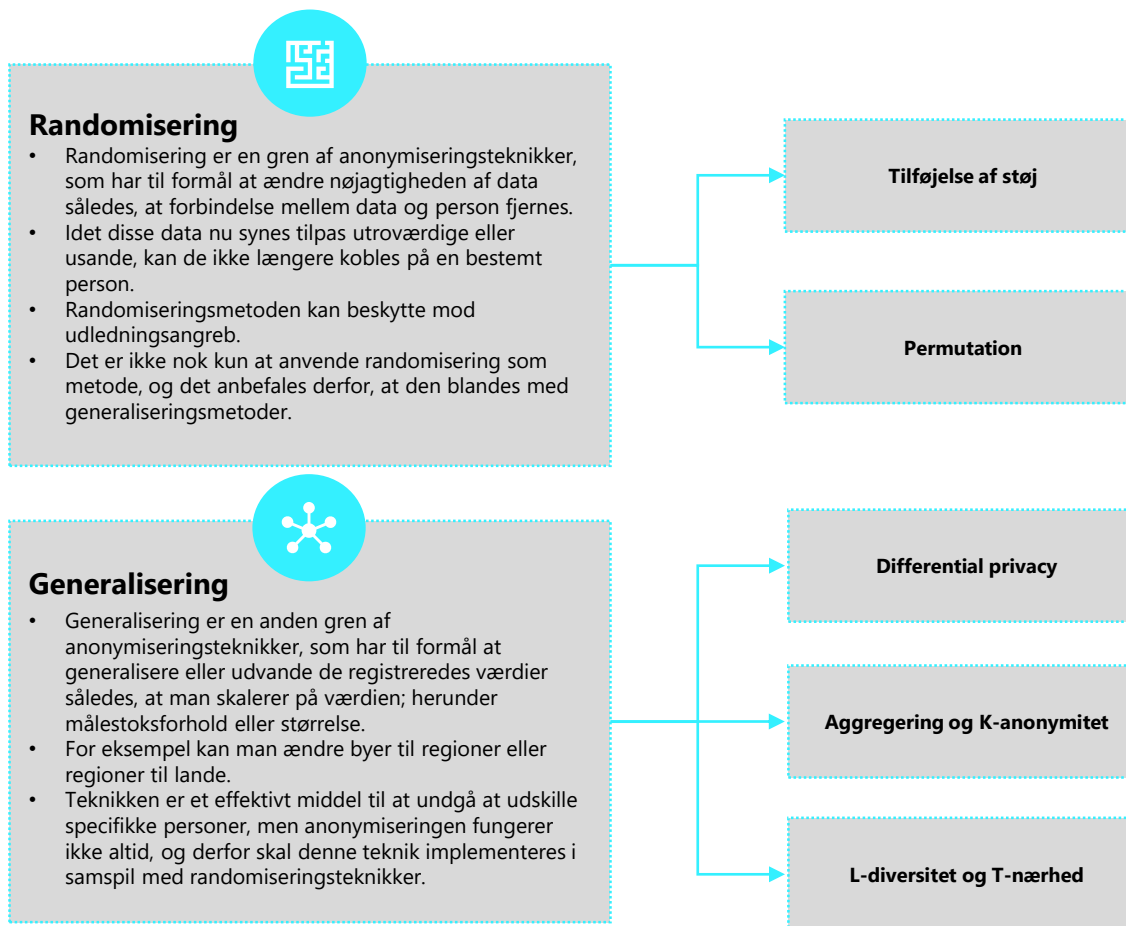
1. Randomisering, hvor nøjagtigheden af data mindskes.

Teknikkerne herunder er:

- *Tilføjelse af støj*, hvor data gøres unøjagtige til en vis grad. Denne teknik anvendes især, når der er personhenførbare informationer i data.
- Permutation, hvor værdier i et datasæt rykkes rundt. Denne teknik anvendes især, når man skal bevare værdier i et datasæt.

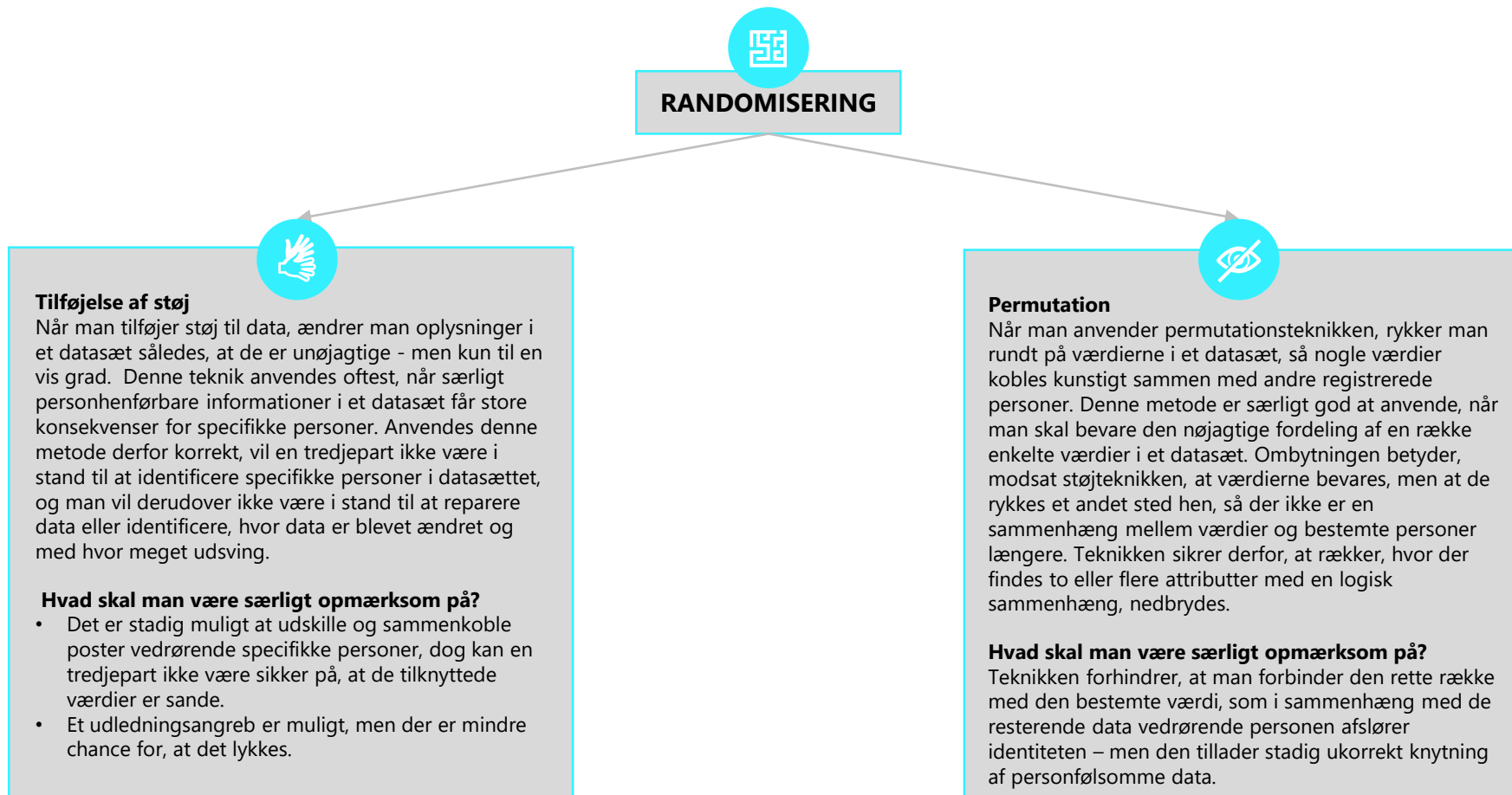
2. Generalisering, hvor værdierne i data generaliseres. Teknikkerne herunder er:

- *Differential privacy*, hvor der pålægges støj, og formatet tilpasses afhængigt af, hvem tredjeparten er.
- Aggregering og K-anonymitet, hvor data anonymiseres ved at gruppere data.
- L-diversitet og T-nærhed er en forlængelse af andre teknikker, herunder K-anonymitet, hvor der anvendes forskellige teknikker til at sikre mod udledningsangreb.



Anonymiseringsteknik: Randomisering

Randomisering har til formål at ændre nøjagtigheden mellem datapunkt og person. Det kan gøres ved brug af to teknikker kendt som "tilføjelse af støj" og "permutation"



Anonymiseringsteknik: Generalisering

Generalisering har til formål at udvande de registreredes værdier således, at man skalerer på værdien; herunder målestoksforhold eller størrelse. Det kan gøres ved brug af tre forskellige teknikker kendt som "Differential privacy", "Aggregering og K-anonymitet", og "L-diversitet og T-nærhed"





Om Deloitte

Deloitte leverer ydelser indenfor revision, consulting, financial advisory, risikostyring, skat og dertil knyttede ydelser til både offentlige og private kunder i en lang række brancher. Deloitte betjener fire ud af fem virksomheder på listen over verdens største selskaber, Fortune Global 500®, gennem et globalt forbundet netværk af medlemsfirmaer i over 150 lande, der leverer kompetencer og viden i verdensklasse og service af høj kvalitet til at håndtere kundernes mest komplekse forretningsmæssige udfordringer. Vil du vide mere om, hvordan Deloittes omkring 312.000 medarbejdere gør en forskel, der betyder noget, så besøg os på Facebook, LinkedIn eller Twitter.

Deloitte Touche Tohmatsu Limited

Deloitte er en betegnelse for en eller flere af Deloitte Touche Tohmatsu Limited ("DTTL"), dets netværk af medlemsfirmaer og deres tilknyttede virksomheder. DTTL (der også omtales som "Deloitte Global") og alle dets medlemsfirmaer udgør separate og uafhængige juridiske enheder. DTTL leverer ikke ydelser til kunderne. Vi henviser til www.deloitte.com/about for nærmere oplysninger.