



Deloitte  
Statsautoriseret  
Revisionspartnerselskab  
CVR-nr. 33 96 35 56  
Weidekampsgade 6  
Postboks 1600  
0900 København C

Telefon: 36102030  
Fax: 36102040  
www.deloitte.dk

13. januar 2020

# Undersøgelse af Rigspolitiets håndtering af historiske teledata – ajourføring

Notatet indeholder en ajourføring af resultaterne i kapitel 5 'Sammenligning af rådata og konverteret data' i Deloitte's afrapportering af *Undersøgelse af Rigspolitiets håndtering af historiske teledata* (herefter undersøgelsen) af 1. oktober 2019.

Baggrunden for ajourføringen fremgår af notatets afsnit 1, hvorefter betydningen for undersøgelsens konklusioner sammenfattes i ledelsesresuméet i afsnit 2. Afsnit 3 indeholder opdaterede hovedtabeller fra undersøgelsens kapitel 5.

## **1 Baggrund for ajourføring af undersøgelse af Rigspolitiets håndtering af historiske teledata**

Det er i forbindelse med udsendelsen af datapakker<sup>1</sup>, der skal fungere som et hjælpeværktøj til efterforskere og anklagere i forbindelse med gennemgangen og kvalitetssikringen af historiske teledata i verserende og afsluttede straffesager, blevet afdækket, at undersøgelsen har omfattet 803 genkonverterede filer, der ikke skulle have været inkluderet i datagrundlaget.

De 803 filer er blevet genkonverteret af Rigspolitiet efter den 8. marts 2019 i forlængelse af, at Rigspolitiet identificerede konverteringsfejl i historiske teledata. Således har den samme konverterede data optrådt to gange i datagrundlaget, der blev lagt til grund for undersøgelsen.

I alt er de 803 filer blevet identificeret i 370 individuelle mapper som indeholder genkonverterede filer, idet det samlede datagrundlag i forhold til konverterede data består af 30.072 individuelle mapper, der er organiseret efter kreds og journalnummer, og hvor ikke alle mapper er navngivet ensartet på tværs af journalnumre. Deloitte og Rigspolitiet har ikke været opmærksomme på, at indholdet af disse 370 mapper har været indlæst i Deloitte's analysedatabase, som har ligget til grund for undersøgelsen.

Deloitte har i forbindelse med ajourføringen gennemført yderligere undersøgelser af om der eventuelt er andre genkonverterede data, der fejlagtigt er indgået i undersøgelsens datagrundlag. I den forbindelse har Deloitte og Telecenteret i samarbejde identificeret en liste over de konverterede data, der er relevante for undersøgelsen. Deloitte har i dialog med Telecenteret verificeret, at alle relevante data fra denne liste indgår, og at der ikke længere indgår genkonverterede data i undersøgelsen.

<sup>1</sup> Datapakker indeholder en excel-fil (valideringsfil), der indeholder oplysninger om eventuelle konverteringsfejl (feltfejl, rækkefølge mv.) samt reference til de anvendte rådatafiler og konverterede filer, som er lagt til grund for valideringsfilen.

Tilstedeværelsen af de genkonverterede filer i undersøgelsens datagrundlag har medført, at der optræder tidsdubletter (teleaktiviteter med samme tidsstempel) i datagrundlaget, der ikke er reelle tidsdubletter, men derimod et udtryk for, at data har optrådt flere gange henholdsvis som konverterede data og igen som genkonverterede data.

Tidsdubletter kan ikke analyseres for fejl i undersøgelsen, idet tidspunktet for en teleaktivitet har været anvendt som en 'nøgle' til at adskille flere teleaktiviteter fra hinanden. Nøglen er nødvendig for at kunne sammenligne teleaktiviteterne i de konverterede data op imod teleaktiviteterne i rådata. Et overlap i tidspunktet for en eller flere teleaktiviteter betyder derfor, at en analyse af de pågældende teleaktiviteter for konverteringsfejl ikke har været mulig. Datagrundlaget er således som følge af, at de genkonverterede filer indgik i undersøgelsen, blevet reduceret unødvendigt.

Forekomsten af de genkonverterede filer i undersøgelsens datagrundlag betyder tillige, at omfanget af fejltypen rækketab i undersøgelsen er underestimeret, da de genkonverterede filer udligner det rækketab, som har eksisteret i de oprindeligt konverterede filer. Rækketab betyder, at der i forbindelse med Rigspolitiets konvertering af historiske teledata er sket et datatab, således at den konverterede fil har indeholdt færre teleaktiviteter end rådatafilen.

På baggrund af ovenstående valgte Rigsadvokaten den 29. oktober at tilbagekalde de til politiet og anklagemyndigheden udsendte datapakker. Rigspolitiet anmodede samtidig Deloitte om, at gennemføre en ajourføring af undersøgelsens resultater, hvor de genkonverterede filer ikke indgik i datagrundlaget. Resultaterne af ajourføringen fremgår af nærværende ajourføringsnotat. Der vil snarest muligt i forlængelse af ajourføringen blive udsendt nye datapakker baseret på det opdaterede datagrundlag.

Ud over de 803 filer, der er genkonverteret i perioden fra den 8. marts 2019 og frem til den 27. august 2019 for at korrigere for rækketab, er der identificeret en række yderligere genkonverteringer. I perioden mellem den 25. november 2010 og den 26. august 2019 er yderligere 1.137 filer blevet genkonverteret en eller flere gange, typisk fordi Telecenteret har modtaget rådata fra teleudbydere med fejl eller mangler, og efterfølgende har fået tilsendt ny rådata med et delvist overlappende indhold, der korrigerede for fejlen.

Genkonverteringerne af de 1.137 filer foranlediger, at der optræder mindst to filer, der i nogen grad er baseret på de samme rådata, hvilket medfører dubletter i både rådata og konverterede data. Samtlige af de 1.137 filer er tilsvarende indgået i undersøgelsens oprindelige datagrundlag, men der er som et led i ajourføringen taget højde for de dublerede filer ved alene at anvende den senest konverterede fil ud fra en generel vurdering af, at dette vil være den bedste fil at lægge til grund for sammenligningen af konverteret data op imod rådata. Dog er en enkeltstående større sag<sup>2</sup> håndteret som en undtagelse til denne regel.

Undersøgelsens resultater er blevet ajourført på baggrund af det verificerede datagrundlag. Resultaterne heraf gennemgås i det følgende.

## 1.1 Sammenfatning

Sammenfattende har tilstedeværelsen af de 803 genkonverterede filer, samt de 1.137 dublerede filer i undersøgelsens datagrundlag, *ikke* betydning for undersøgelsens konklusioner, da fejlilledet fortsat er det samme.

Undersøgelsens konklusioner om de identificerede typer af konverteringsfejl, herunder rækketab samt fejltyper i forhold til feltindhold, er således fortsat gældende i forlængelse af ajourføringen. Som følge af det ajourførte datagrundlag, er de identificerede fejlratere dog steget for en række fejltyper. Konklusionerne

---

<sup>2</sup> Den pågældende sag har et meget betydeligt omfang af rådata og konverteret data, hvilket påvirker den samlede statistik.

i forhold til række-tab og konverteringsfejl på feltniveau som følge af ajourføringen uddybes i de efterfølgende afsnit.

Tilstedeværelsen af de 803 genkonverterede filer i datagrundlaget har medført, at undersøgelsen har underestimeret omfanget af række-tab. Ajourføringen, hvor de genkonverterede filer er udeladt i datagrundlaget, identificerer, at samlet set 1.538 rekvisitioner er berørt af konverteringsfejlen række-tab sammenlignet med de 1.131 rekvisitioner, der var angivet i undersøgelsen. Der er således i forlængelse af ajourføringen sket en stigning i antallet af rekvisitioner berørt af konverteringsfejlen række-tab fra ca. 5,8 procent i undersøgelsen til ca. 7,9 procent. Der henvises til tabel 2 i ajourføringsnotatets afsnit 3.

Omfanget af række-tab i de 1.538 (tidligere 1.131) rekvisitioner fordeler sig nogenlunde som det række-tab, der fremgik af undersøgelsen. Der henvises til tabel 3 afsnit 3 for en detaljeret visning af omfanget af række-tab på rekvisitionsniveau.

Udeladelsen af de genkonverterede filer i undersøgelsens datagrundlag har dertil foranlediget, at det i ajourføringen har været muligt at sammenligne indholdet på feltniveau for et større antal rækker end de 48,7 millioner rækker, der indgik i undersøgelsen. Ajourføringen af datagrundlaget medfører således, at det har været muligt at gennemføre analyser af feltindholdet af 56,5 millioner rækker.

Det opdaterede datagrundlag viser, at der for en række fejltyper på feltindhold overvejende er de samme fejlratere, som er angivet i undersøgelsen, eller hvor afvigelserne ligger inden for det, der kan betegnes som statistisk usikkerhed. Fejltyperne på indholdet tjenestetype og opkaldsvarighed viser således minimale afvigelser, der falder inden for den statistiske usikkerhed, idet fx 'tjenestetype' ved ajourføring er på 2,1 procent sammenlignet med tidligere 2,0 procent i undersøgelsen, mens fejlraten på indholdet 'varighed' ved ajourføring udgør 32,4 procent mod tidligere 33,3 procent.

Udeladelsen af de genkonverterede filer har imidlertid også foranlediget, at konverteringsfejl i forhold til visse fejltyper på feltindhold er steget. Det drejer sig om konvertering af start- og særligt slutmaste-positioner, som angiver maste-positionen ved teleaktivitetens start og afslutning. Maste-positioner leveres i forskellige koordinat-formater af teleudbydere og konverteres til ét nyt format af Telecenteret. Derfor er det ikke muligt at lave en 1:1 sammenligning af om rådata svarer til konverterede data. Det er i stedet blevet valideret om maste-positioner samt celleretninger er korrekt konverteret, herunder at de to konverterede maste-positioner og celleretningen når frem til samme resultat.

I forhold til startmaste-positioner er der sket en mindre stigning i andelen af fejlkonverteringer fra 1,2 procent til 1,4 procent. Stigningen skyldes formentlig, at genkonverteringerne har nedbragt fejlraten. I forhold til feltet "celleretning startmast", som vedrører hvilken retning i grader på startmasten, som den registrerede teleaktivitet har fundet sted på, ses en stigning fra 0,5 procent til 0,8 procent, hvilket ligeledes skyldes, at de genkonverterede filer har nedbragt fejlraten.

I forhold til slutmaste-positioner er der sket en stigning i andelen af fejlkonverteringer fra 0,7 procent til 4,3 procent. Stigningen kan i det væsentligste henføres til, at de genkonverterede data havde rettet fejlen, og dermed dækket over at slutmaste-position manglede i de oprindeligt konverterede data.

Dette illustrerer samtidig, at de genkonverteringer, som har været gennemført af Telecenteret efter den 8. marts på baggrund af identificerede konverteringsfejl, har nedbragt fejlraten.

Der har som led i ajourføringen været foretaget en mere præcis sammenligning<sup>3</sup> af oplysningerne IMEI samt A, B og C-numre, hvilket har foranlediget, at der er identificeret flere uoverensstemmelser mellem rådata og

<sup>3</sup> Den mere præcise sammenligning skyldes, at der tidligere var IMEI numre og lange telefonnumre – typisk udenlandske numre som ved en fejl var blevet underlagt såkaldt "scientific notation" (hvor eksempel 457812369612457 bliver til 4.57812E15+). Således var valideringen ikke foretaget på alle cifre, men alene et udsnit af cifrene. Ved opdateringen er alle cifre blevet valideret.

konverteret data. Således er fejlraten for IMEI-numre steget fra 19,4 procent i undersøgelsen til 28,0 procent i forlængelse af ajourføringen, samt for A-numre (fra 0,0 procent til 0,6 procent), B-numre (fra 0,0 procent til 0,1 procent) og C-numre (fra 0,0 procent til 0,1 procent). Stigningen i fejlratene skyldes, at de fulde talrækker, som primært vedrører udenlandske telefonnumre, er blevet sammenlignet, hvor det tidligere i undersøgelsen alene var muligt at undersøge en delmængde af talrækken.

Der har ligeledes som led i ajourføringen været foretaget en justering i sammenligningen af Fokus ABC. Denne justering har foranlediget en faldende fejlrate fra 6 procent til 1,2 procent. Fokus ABC er et "beregnet feltindhold"<sup>4</sup>, som identificerer om det er henholdsvis A-, B- eller C-nummeret, der er genstand for indgrebet. Specielt for én teleudbyder er konverteringsreglerne komplicerede og involverer analyse af en lang tekststreng. Den justerede validering af fokus ABC korrigerer således for små ændringer i tekststrengen, som ikke har nogen indholdsmæssig betydning, men som tidligere indgik i opgørelsen af fejlkonverteringer, hvilket er baggrunden for faldet i fejlraten.

Opsummerende er det således fortsat de samme fejltypen og det samme generelle fejl billede, der optræder i forlængelse af ajourføringen. De stigninger i fejlratene, der gør sig gældende, skyldes, at mængden af analyseret data samlet set har været større i ajourføringen end i undersøgelsen på trods af at datagrundlaget er blevet reduceret. Det illustrerer samtidig, at de tiltag, som Telecenteret har gennemført i forbindelse med genkonverteringerne, har reduceret fejlratene.

Hertil kommer, at en styrket validering af flere IMEI- og A, B og C-numre har afdækket nogle flere konverteringsfejl, mens fejlraten for Fokus ABC er faldet.

Det vil af de ajourførte datapakker fremgå hvilke rådatafiler og konverterede filer, som er lagt til grund for valideringen af Rigspolitiets konvertering af historiske teledata.

## 2 Ajourførte statistikker

I dette afsnit fremgår en opdatering af udvalgte tabeller fra kapitel 5 'Sammenligning af rådata og konverteret data' i Deloitte's afrapportering af *Undersøgelse af Rigspolitiets håndtering af historiske teledata* (herefter undersøgelsen) af 1. oktober 2019. Tal, der er angivet i parentes, er undersøgelsens oprindelige tal.

**Tablet 1: Antal rekvisitioner med rådata og konverteret data (tal fra oprindelig undersøgelse i parentes)**

Rekvisitioner	Antal	Procent hvor både rådata og konverteret data kan forefindes
Rådata	41.845 (41.842)	-
Konverteret data	19.664 (24.245)	47,0 % (57,9 %)
- Heraf hvor der findes både rådata og konverteret data	19.523 (24.028)	46,7 % (57,4 %)
- Heraf hvor der findes både rådata og konverteret data, og der er indhold <sup>5</sup>	19.501 (19.386)	46,6 % (46,3 %)

Andelen af rekvisitioner, hvor både rådata og konverteret data kan forefindes, er faldet med 10,9 procent fra 57,9 procent til 47,0 procent. Undersøgelsen afslørede mange filer, der ikke indeholdte relevante data,

<sup>4</sup> Et beregnet feltindhold betyder, at værdien ikke sendes fra teleudbyderne, men beregnes af Telecenteret.

<sup>5</sup> Tidligere fald fra 24.048 til 19.386 skyldes, at der har været forespørgsler på teleaktiviteter, hvor der ikke har været teleaktiviteter. Disse tomme rekvisitioner er ikke medtaget i denne opdatering.

typisk fordi teleudbyderne returnerede tomme filer i de tilfælde, hvor teleudbyderne ikke lå inde med relevante data i relation til den aktuelle rekvisition. Disse tomme filer fremgik i undersøgelsens opgørelse overantallet af rekvisitioner, hvor der findes både rådata og konverteret data, men indgik ikke i øvrigt i undersøgelsen. Med afsæt i denne indsigt har Deloitte målrettet indlæsningen af data i forbindelse med ajourføringen ved at undlade at indlæse disse tomme Zip filer (filer uden indhold) i ajourføringsgrundlaget. Dertil har Deloitte sammen med Telecenteret kvalificeret hvilke filer, og dermed hvilke rekvisitioner, det var relevant at lægge til grund for undersøgelsen.

Blandt de relevante 19.664 rekvisitioner har 19.501 indeholdt brugbare data. Stigningen på 115 filer i forhold til undersøgelsen skyldes, at programmet, der modtager, ud-pakker og konverterer de modtagne data, er blevet forbedret, hvorfor det har været muligt at behandle yderligere 115 filer.

**Tabel 2: Rekvisitioner med rækketab på rekvisionsniveau (tal fra oprindelig undersøgelse i parentes)**

	Antal unikke rekvisitioner med overensstemmelse	Antal unikke rekvisitioner uden overensstemmelse
Total antal	17.613 (18.047)	1.888 (1.339)
Heraf flere rækker i rådata end konverteret data (sessionsrapport)	N/A	1.538 (1.131)
Heraf færre rækker i rådata end konverteret data (sessionsrapport)	N/A	350 (208)

Som det fremgår af tabellen, er antallet af rekvisitioner berørt af konverteringsfejlen rækketab steget som følge af det ajourførte datagrundlag. Dette er jf. sammenfatningen forventeligt, fordi de genkonverterede filer i undersøgelsens resultater har udlignet rækketabet i en række rekvisitioner, og således har konverteringsfejlen rækketab været underestimeret i undersøgelsen.

**Tabel 3: Omfang af rækketab på rekvisionsniveau (tal fra oprindelig undersøgelse i parentes)**

	Antal unikke rekvisitioner med flere rækker i rådata end konverteret data	Procent af rekvisitioner med flere rækker i rådata end konverteret data
Total antal	1.538 (1.131)	100 %
Rækketab på 1 række	248 (142)	16,1 % (12,6 %)
Rækketab på 2-10 rækker	404 (261)	26,3 % (23,1 %)
Rækketab på 11-100 rækker	364 (275)	23,7 % (24,3 %)
Rækketab på >100 rækker	522 (453)	33,9 % (40,0 %)

Rækketabet fordeler sig nogenlunde tilsvarende det rækketab, der fremgik af undersøgelsen, dog således, at der er forholdsmæssigt flere rekvisitioner med få rækketab. Det kan konstateres, at det absolutte rækketab opgjort i antal øges for alle intervaller som følge af justeringen.

**Tabel 4: Analyse af feltindhold (tal fra oprindelig undersøgelse i parentes)**

Feltindhold	Rækker uden overensstemmelse som procent af totalt antal rækker i feltsammenligning (*)
Tjenestetype	2,1 (2,0)
Varighed	32,4 (33,3)
A-nummer	0,6 (0,0)
B-nummer	0,1 (0,0)
C-nummer	0,1 (0,0)

Startmast	2,1 <sup>6</sup> (1,5)
- Heraf koordinat	1,4 (1,2)
- Heraf celleretning	0,8 (0,3)
Slutmast	4,4 (0,7)
- Heraf koordinat	4,3 (0,7)
- Heraf celleretning	0,1 (0,0)
IMEI	28,0 (19,4)
Fokus ABC	1,2 (6,9)

(\*) Udtrykker dermed rækker med konverteringsfejl i feltsammenligningen som procent af total antal rækker

Fejlratene i forhold til fejltypene på feltindhold viser, at ajourføringen for nogle felter har foranlediget en stigning, mens der for andre er sket et fald, jf. sammenfatningen ovenfor.

Opsummerende er der således fortsat de samme fejltypen og det samme generelle fejl billede, der optræder i ajourføringen. Stigningerne i fejlratene skyldes generelt, at mængden af analyseret data samlet set har været større i ajourføringen end i undersøgelsen, givet at rækker med dubletter er reduceret, og dermed øger mulighederne for at foretage sammenligning af rådata op imod konverteret data.

Hertil kommer, at en styrket validering og validering af flere IMEI- og A, B og C-numre har afdækket flere konverteringsfejl af denne type. Resultaterne illustrerer samtidig, at de tiltag, som Telecenteret har gennemført i forbindelse med genkonverteringerne, har reduceret fejlratene.

---

<sup>6</sup> Afrundet fra 2,13