Datatilsynet

FINANSDEPARTEMENTET
Postboks 8008 DEP
0030 OSLO

Deres referanse	Vår referanse	Dato
18/1077	18/00878-2/EOL	26.06.2018

Datatilsynets høringsuttalelse - Forslag til regler om at Skatteetaten og Tolletaten kan bruke personopplysninger til å utvikle og teste it-system

Vi viser til Finansdepartementets høringsbrev av 24. mai 2018 vedrørende forslag til regler om at Skatteetaten og Tolletaten kan bruke personopplysninger til å utvikle og teste it-system.

Oppsummering
Datatilsynet støtter ikke Finansdepartementets forslag om egen hjemmel for å tillate Skatteetaten og Tolletaten å bruke personopplysninger til å utvikle og teste sine informasjonssystemer. Vi mener det er fullt ut mulig å lage syntetiske testdata for både utvikling og integrasjon av nye systemer, samt test av eksisterende systemer. Bruk av produksjonsdata vil derfor ikke være nødvendig for formålet.

Datatilsynet oppfordrer Finansdepartementet til å gi Skatteetaten, som forvalter av Folkeregisteret, de ressurser som skal til for å etablere et eller flere syntetiske testdatasett som flere offentlige etater, kommuner og virksomheter fritt kan bruke. Dette er et tiltak som vil tjene både økonomi, kvalitet og personvern.

Syntetiske data er løsningen for testing
Finansdepartementets utgangspunkt for forslag til regler om at Skatteetaten og Tolletaten kan bruke personopplysninger til å utvikle og teste informasjonssystemer er at fiktive testopplysninger vil ofte ha for lite variasjon, være for statistiske eller ikke gjenspeile produksjonssituasjonen.

Denne påstanden vil vi tilbakevise med følgende argumenter:

For det første er hensynet til *variasjon* nettopp en god grunn til å *ikke* bruke produksjonsdata. Når man skal teste et system ønsker man å teste på alle mulige scenarier – også de som man ikke finner i produksjon. Slike scenarier vil man imidlertid kunne generere ved hjelp av syntetiske data. Ved å skape et testgrunnlag bestemmer man variasjonen selv.

For det andre er *dynamikk* i datasett ikke nødvendigvis en fordel med tanke på formålet test. Det kan for eksempel bli vanskelig å automatisere test når grunnlaget i produksjonsdataene endrer seg og man må laste inn produksjonsdata på nytt.

Postadresse:	Kontoradresse:	Telefon:	Telefaks:	Org.nr:	Hjemmeside:
Postboks 8177 Dep 0034 OSLO	Tollbugt 3	22 39 69 00	22 42 23 50	974 761 467	www.datatilsynet.no

Skatt og Toll argumenterer med at de må kunne teste på tvers av systemer. I så fall vil endringer i personopplysninger som følge av saksbehandling hos Toll kunne gjøre at en test hos Skatt feiler, og som igjen forårsaker tid brukt til feilsøking.

Det å simulere dynamikk er for øvrig enkelt å få til ved å legge inn i programvaren at visse hendelser skal genereres jevnlig (f.eks fødsler, at populasjonen blir eldre, etc).

Bedre kvalitet og hurtigere tilgang ved bruk av syntetiske data
Da står vi igjen med det som synes som hovedargumentet for å bruke reelle personopplysninger til test og utvikling: «ved å gjøre det på denne måten slipper Skatteetaten og Tolletaten å bruke penger og ressurser på å lage syntetiske testdata».

Er dette et gyldig argument? Datatilsynet mener nei, og vi vet fra vårt arbeid med veileder for programvareutvikling med innebygd personvern at kompetansemiljø relatert til testing også er av samme oppfatning. I et kort perspektiv kan det nok hende at etatene sparer noen kroner. I det lange perspektiv vil det bli kostbart å leve med risikoen det innebærer å bruke personopplysninger til testformål, tidsbruken dette tar og den mangelfulle kvaliteten.

I en artikkel som omhandler testdata og personvern i NAV kommer det frem at 2/3 av innsatsen som kunne vært benyttet til faktisk testing av NAV sine systemer blir benyttet til feilsøking, flytting og annen preparering av produksjonsdata[1]. Datatilsynet er også kjent med at NAV allerede har utarbeidet en syntetisk basispopulasjon som benyttes i testing.

Et annet argument for bruk av produksjonsdata har vært at det kun er her du finner spesialtilfellene og grensetilfellene. Det som ikke kommer frem er at det brukes store mengder tid for å søke opp disse grensetilfellene. Dersom man isteden produserer personopplysninger for det formål å teste grenser og kompliserte tilfeller vil dette gå raskere.

Dette tilsier at det å skape syntetiske data har et stor potensiale både hva gjelder kvalitet og økonomi.

Innebygd personvern i utviklingsfasen, inklusiv testfase
Med personvernforordningen får virksomheter en plikt til å sørge for å bygge personvern inn i sine informasjonssystemer.

Ved utvikling av ny teknologi skal man ta hensyn til de krav som gis ved personvernprinsippene, de registrertes rettigheter og friheter. Kravene bygges inn i alle utviklingsfaser, dette inkluderer testing av løsningen. I utviklingsaktiviteten test skal testerne sjekke at krav til personopplysningsvern og sikkerhetskrav satt i design – og sikkerhetsfasene faktisk er implementert og riktig implementert. Sikkerhetstesting innebærer en omfattende testing av programvaren for å avdekke sårbarheter og for å være sikker på at koden ivaretar sikkerhet og personvern på tilstrekkelig måte, i hele utviklingsløpet.

[1] Promiss-qualify.no/index.php/artikler/82-gdpr-og-hvorfor-vi-ikke-kan-bruke-skarpe-testdata

Ved utvikling eller bestilling av applikasjoner, systemer og løsninger som skal behandle personopplysninger plikter behandlingsansvarlig å sikre at disse har innebygd personvern og personvern som standardinnstilling.
Slik Datatilsynet forstår forslaget fra Finansdepartementet står det i delvis motstrid til plikten til å sørge for innebygd personvern.

Nederst på side 8 i høringsnotatet står det om utviklingsarbeidet og hvem som gjør hva. Her introduseres begrepet «smidig utvikling». Poenget i moderne konsepter som Smidig, Agile, DevOps osv. er at man ikke lengre skiller klart på hvor test faktisk foregår i utviklingsløpet, man tester hele tiden.

Nederst på side 10 i høringsnotatet omtales test gjennomført av ordinære saksbehandlere. Vi oppfatter dette som «User Acceptance Test» (UAT) og ikke test som skal gjøres i fortløpende i hele utviklingsløpet (ved for eksempel bruk av Smidig, Agile, DevOps).

Dersom man bruker syntetiske data er denne måten å gjøre utvikling og test på ikke problematisk. Dersom man bruker reelle data derimot vil en slik fremgangsmåte stride mot flere av grunnprinsippene i personvernforordningen;

To helt sentrale grunnprinsipper for behandling av personopplysninger i personvernforordningen er prinsippet om formålsbegrensing[2] og prinsippet om dataminimering[3]. Formålet med behandlingen av personopplysninger skal klart defineres, og behandlingsansvarlig har en plikt til å kun bruke de personopplysningene som er relevante og nødvendige for det på forhånd definerte formålet. Som beskrevet over vil utvikling og test gå så over i hverandre at dette gjør det vanskelig å ramme inn hver enkelt formål med behandling av personopplysninger, og følgelig også tilsvarende vanskelig å spesifisere hvilke data som er nødvendige for henholdsvis utvikling og test.

Faglitteratur og –artikler skrevet av personer som har tyngde i testmiljø tilsier dessuten at testdata kan bli en flaskehals dersom man ikke begynner å benytte syntetiske data i DevOps, Smidig, Algile osv. Det er altså ønskelig å benytte syntetiske data fordi man med dette øker effektiviteten.

Siden man i Smidig, DevOps, Agile osv. tester gjennom hele utviklingsløpet er tanken at test ikke bør begrenses til UAT alene. Dermed blir nettopp bruk av Smidig, DevOps, Agile osv. et argument for at det *også* er lønnsomt å etablere og benytte syntetiske data fordi man tester i hele utviklingsløpet.

Fuzz-testing er også en del av testingen. Poenget med Fuzz-testing er å finne sårbarheter ved å benytte «tilfeldige og ikke-forventede data». Heller ikke denne type testing tilsier at man skal benytte reelle data.

---

[2] Personvernforordningen art. 5 (1) bokstav b
[3] Personvernforordningen art. 5 (1) bokstav c

Ved bruk av kunstig intelligens (KI) eller «deep learning» i en utviklingsfase gjenbrukes ofte tredjeparts programvare. Det er viktig at behandlingsansvarlig forvisser seg om at tredjeparts programvare, støttekomponenter, utviklingsverktøy mm er godkjente verktøy og rammer.

Kunstig intelligens er for øvrig ikke nevnt i Finansdepartementets høringsnotat. Datatilsynet kjenner imidlertid til at stadig flere virksomheter definerer opplæring av KI som en del av utviklingen av programvare. Datatilsynet mener det hadde er interessant for allmennheten å vite hvilket ståsted Skatteetaten og Tolletaten har i dette spørsmålet - definerer disse etatene opplæring av KI som utvikling?

Ved maskinlæring dukker det opp egne problemstillinger som må vurderes. Datatilsynet har skrevet en rapport om KI og personvern hvor teknologien vurderes opp mot personvernprinsippene.

Vi anbefaler å lese denne dersom statlige etater vurderer KI som programvareutvikling. Rapporten ligger på våre sider www.datatilsynet.no og heter «Kunstig intelligens og personvern».

Vurdering av personvernkonsekvenser
Skatteetaten og Tolletaten behandler personopplysninger i stort omfang, og det kommer stadig nye lovforslag som har til hensikt å utvide etatenes adgang til å sammenstille og utlevere personopplysninger[4].

Lovforslagene viser at Skatteetaten og Tolletaten fortløpende ser på hvordan teknologi og ny teknologi kan bistå dem i effektivisering i å kontrollere den enkelte borgers etterlevelse av regelverket. Det er derfor spesielt viktig at disse etatene sørger for innebygd personvern, at konsekvenser for personvernet vurderes og at det iverksettes tiltak som reelt sett ivaretar borgernes rettigheter og friheter.

Det er lovgivers oppgave å redegjøre for personvernkonsekvensene slik at høringsinstansene kan få seg presentert de ulike sidene ved forslaget og gjøre seg opp en kvalifisert mening. Datatilsynet mener det er en klar mangel ved høringsnotatet at det ikke er foretatt noen reell vurdering av personvernkonsekvenser.

Finansdepartementet konkluderer kort med at nytten av å kunne bruke reelle data for test og utvikling overstiger personvernulempene for de registrerte. Det eneste tiltaket vi kan se som skal begrense inngrep i personvernet er at det skal beskrives i retningslinjer for utvikling og test. Dette er retningslinjer som ikke er en del av forslaget og det er dermed ikke mulig å vurdere om de inneholder reelle garantier for å sikre de registrertes rettigheter og friheter.

Uten å gå inn på hvilke implikasjoner lovforslaget kan tenkes å få for de registrertes rettigheter, eller å beskrive reelle garantier for ivaretagelse av rettigheter og friheter etter personvernforordningen, er det ikke gjort en reell vurdering av personvernkonsekvenser.

[4] Se høring om forslag om endringer i reglene om informasjonsbehandling i Skatteetaten med frist 25. juni 2018 som gjelder utvidelse av hjemler for sammenstilling og utlevering av personopplysninger.

Avsluttende merknader
Datatilsynet støtter ikke Finansdepartementets forslag om hjemmel til å bruke reelle produksjonsdata (personopplysninger) til test og utvikling av informasjonssystemer i Skatteetaten og Tolletaten.

Datatilsynet mener at Finansdepartementet isteden må bevilge de ressurser som skal til for at statlige etater kan ta i bruk syntetiske datasett for formålet test og utvikling.

Til sammen har de store offentlige etatene i Norge veldig mange systemer (NAV har alene over 200systemer), men de forholder seg i stor grad til samme datasett – den norske befolkningen. Vi vil oppfordre Finansdepartementet til å gi Skatteetaten, som forvalter av Folkeregisteret, de ressurser som skal til for å etablere et eller flere syntetiske testdatasett som flere etater fritt kan bruke.

Utvikling og bruk av syntetiske data vil være den beste løsningen både av hensyn til økonomi, kvalitet og personvern.

Vi stiller oss til rådighet dersom dere ønsker utfyllende informasjon eller har spørsmål til vårt høringssvar. Dere kan i så fall ta kontakt med Eirin Oda Lauvset på telefon 22 39 69 11.

Med vennlig hilsen

Bjørn Erik Thon
direktør

Eirin Oda Lauvset
seniorrådgiver