DATAETHICS

DataEthics Whitepaper 2

Kortlægning af dataflows og konsekvensanalyse

Af Birgitte Kofod Olsen, partner, PhD, Carve Consulting og medstifter af Dataethics

Dette whitepaper gennemgår de faser, en virksomhed eller myndighed skal igennem får at kortlægge dataflows. Det vil sige *dataindhentningen* - herunder også dataanalyse og såkaldte følsomme data; *dataanvendelsen* - herunder også datasikkerheden samt anvendelse af data fra leverandører og andre samarbejdsparter; samt *ophøret med databehandlingen*. Endelig forklares, hvad en konsekvens- og riskikoanalyse bør indholde hvilke data der er særligt risikable, og hvad de negative konsekvenser af behandling af persondata kan være.

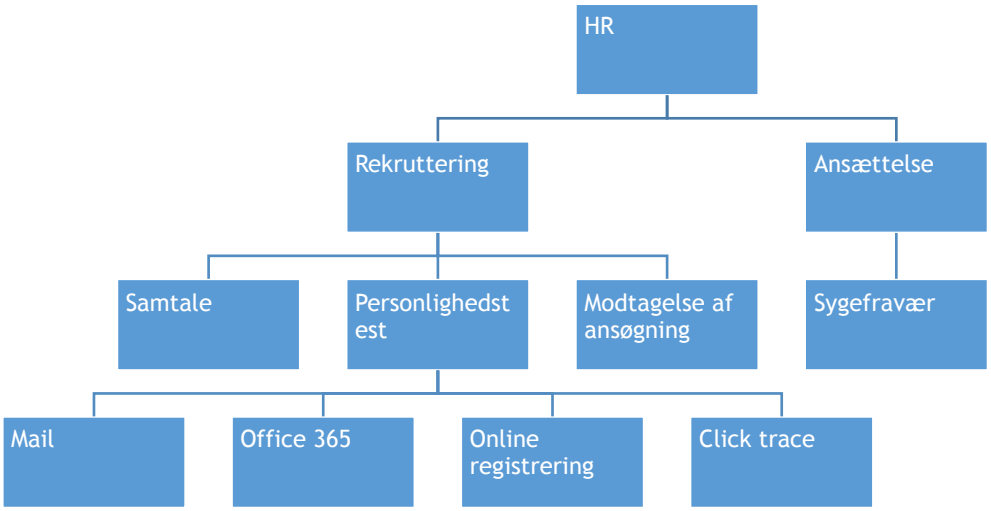
En kortlægning af en virksomheds eller myndigheds dataflows skaber indsigt i databeskyttelsens effektivitet og et oplyst grundlag for beslutninger om organisatoriske, tekniske og sikkerhedsmæssige foranstaltninger.

Ved at kortlægge organisationens dataflows bliver vi også i stand til at vurdere, hvordan databehandlingen påvirker kundernes, borgernes eller medarbejdernes persondatabeskyttelse. Hvis deres rettigheder og friheder påvirkes negativt, og der er høj sandsynlighed for, at det sker, skal vi tilpasse både databehandlingen og sikkerhedsforanstaltningerne til risikoniveauet – eller helt undlade at anvende persondata, der har sådanne konsekvenser. Med dataforordningens artikel 35 indføres en retlig forpligtelse til at udføre konsekvensanalyse, når der er høj risiko for datasubjektets rettigheder og frihedsrettigheder. Undladelse vil kunne udløse et bødekrav på 10 mio eller op til 2% af årlig omsætning.

Med en dataflow-kortlægning identificeres alle dele i de behandlede persondata's livscyklus, fra de strømmer ind i organisationen til de spredes og anvendes i forskellige enheder, videregives til tredjeparter og til sidst slettes.

Det er afgørende, at de data, der indgår i kortlægningen, vedrører identificerbare eller identificerede fysiske personer. Hvis data er anonymiserede, krypterede eller på anden måde beskyttet, så de ikke kan henføres til en bestemt person, er der ikke tale om persondata i forordningens forstand. Men hvis identifikation kan ske direkte eller indirekte gennem en indikator som fx navn, cpr-nummer, lokaliseringsdata eller en IP-adresse – eller sammenstilling heraf – udløser det persondatabeskyttelse. Det samme gælder, når indikatorerne afslører en persons fysiske, fysiologiske og genetiske kendetegn eller vedkommendes psykiske, økonomiske, kulturelle eller sociale identitet.

Før dataflowet kan tegnes og beskrives på en overskuelig måde, er det nødvendigt at udvælge de forretningsområder, der håndterer persondata. Det kan være en salgsafdeling, HR-administration, forskning og udvikling eller IT-driften. For hvert af disse områder vil det typisk være nødvendigt at opdele dem i funktionsområder. HR kan fx opdeles i funktionsområderne rekruttering, ansættelsesforholdet og afskedigelse. Hvert funktionsområde vil have nogle processer, der styrer relevante aktiviteter, og it-systemer, der understøtter begge dele. Fx benytter mange organisationer online-ansøgning, hvor cv og dokumenter uploades online til en database og efterfølgende evt. trækkes ud til behandling i et tekstbehandlingssystem.



Hvor dybt i niveauerne man skal gå afhænger af såvel den samlede it-arkitektur som af de systemkomponenter, der anvendes. Nogle algoritmer kan benytte følsomme personidentifikatorer som variable og skal derfor beskrives som en komponent i dataflowet, eller der kan være etableret interfaces mod cloudløsninger, som også kræver kortlægning. Derudover kan selvstændige datasæt, fx bestående af sundhedsoplysninger, være af en sådan karakter, at de skal beskrives som selvstændige komponenter i dataflowet.

Alle dele af værdikæden og alle niveauer skal derfor screenes for persondatahåndtering. På den baggrund vil det være muligt at identificere og prioritere de områder, hvis dataflows skal tegnes og beskrives.

Dataflowets input

Et dataflow kan opdeles i tre overordnede faser. Dataindhentningen, dataanvendelsen og ophøret med databehandlingen.

Den første fase angår *data input*, d.v.s. **indsamlingen eller indhentningen af persondata**. Data kan være tilvejebragt af datasubjektet selv eller indhentet hos en myndighed eller en anden virksomhed eller tredjepart. Denne del af kortlægningen bør derfor beskrive følgende:

Datakilderne

- Hvem indhentes persondata fra?
- Har datasubjektet givet samtykke til dataindhentningen - og hvordan?
- Eller er der et andet lovligt grundlag for dataindhentningen?
- Er datakilderne geografisk placereret indenfor eller udenfor EU
- Beriges eller analyseres data hos en tredjepart inden modtagelse?

Særligt på grund af mulighederne for at kombinere data med fx statistiske oplysninger eller oplysninger fra sociale medier, er det vigtigt at dataflowsbeskrivelsen indeholder abonnementsordninger eller andre services hos tredjeparter, der håndterer eller sammenstiller data på

vegne af den dataansvarlige virksomhed, der rekvirerer persondata. Hvis driften af et online-ansøgningssytem fx er outsourcet til en tredjepart, som ved brug af en algoritme sorterer ansøgningerne efter uddannelsesinstitution eller bopæl inden de sendes til virksomheden, udgør aktiviteten et datainput i dataflow-beskrivelsen.

Dataelementer

- Hvilke persondata indhentes?

I praksis rubriceres data i såkaldte felter. De kan fx indeholde telefonnummer, adresse, kommune, alder, betalingskortnummer eller indkøbssted. Det er vigtigt at beskrive alle datafelter i det funktionsområde, proces eller system, som kortlægges. Hvis antallet af felter er meget stort, vil de ofte kunne samles i emneområder. Fx kan et cpr-nummer udmøntes i langt flere felter end alder og køn, bl.a. ved at validere det hos cpr-registeret i forhold til adresse. Adressen kan nemlig omsættes til nye felter, der oplyser om bopæl, lokalmiljø og boligtype. Her kan der være behov for at opdele felterne i områder som fx personoplysninger, ejendomsoplysninger, lokationsoplysninger.

Datatyper

- Er de indhentede data almindelige persondata?
- Er der tale om kumulering af flere forskellige typer af almindelige persondata?
- Kan de i så fald afdække oplysninger af privat karakter?
- Er der tale om særlige kategorier af oplysninger?
- Omfatter data oplysninger om strafbare forhold og lovovertrædelser

I beskrivelsen bør det fremgå, om datasubjektet let kan identificeres ud fra de persondata, der behandles om dem. Identificerbarheden kan som nævnt ovenfor være umiddelbar eller direkte, fx ved registrering af navn, adresse og CPR-nummer, eller indirekte, fx via oplysninger om geografiskab placering og tidspunkt, som evt. sammenstilles med oplysninger om betalingskort, medlemskab, nummerplade eller andre oplysninger, der knytter de behandlede data til en bestemt person.

Med de forbedrede muligheder for at sammenstille data fra forskellige kilder og udføre *data mining* på *big data*, er der en øget risiko for at almindelige persondata som navn og adresse, IP-adresse, betalingskort, købstidspunkt og sted, når de vurderes samlet, kan afspejle oplysninger om de personer, hvis data indgår i databehandlingen, herunder om deres opholdssteder, bevægemønstre samt interesser og vaner. I sådanne tilfælde vil der kunne opstå risiko for, at der skabes indsigt i private forhold, som ikke med rimelighed kan forventes af datasubjektet, og som må anses for irrelevante eller uproportionale i forhold til formålet med dataindhentningen.

De datatyper, vi kender som følsomme persondata, kaldes i dataforordningen for særlige kategorier af data. De er oplistet i artikel 9 og omfatter personoplysninger om racemæssig eller etnisk oprindelse, politisk, religiøs eller filosofisk overbevisning og fagforeningsmæssigt tilhørsforhold samt behandling af genetiske data, biometriske data for entydigt at identificere en person eller oplysninger om helbredsforhold eller seksuelle forhold og seksuel orientering. Det er som udgangspunkt forbudt at behandle denne type personoplysninger.

Typebestemmelsen af de indhentede persondata er væsentlig for den senere konsekvensanalyse og risikovurdering, og bør derfor beskrives grundigt og præcist.

Databehandlingen

Den anden fase i dataflowet omfatter den konkrete **anvendelse af persondata** for at imødekomme kundens ønsker om service og oplysning eller borgerens krav på en offentlig ydelse. Når data anvendes skal principperne for databehandling i artikel 5 overholdes og det er derfor nødvendigt for hvert funktionsområde, proces eller datasæt at beskrive følgende:

- Hvad er formålet med indsamling af persondata?
- Indsamles kun tilstrækkelige og relevante data?
- Er data nødvendige for at opfylde formålet?

For at skabe grundlag for vurdering af, om kravene til dataansvarlighed er opfyldt, d.v.s. at databehandlingen er lovlig, rimelig og sker på en gennemsigtig måde i forhold til datasubjektet, samt kravet til den dataansvarlige om at gennemføre passende organisatoriske og tekniske foranstaltninger (artikel 24), kan dataflow-kortlægningen med fordel suppleres med disse oplysninger:

- Hvem er ansvarlig for indhentning?
- Hvordan sikres det, at dataene kun anvendes til det formål, de er indhentet til?
- Hvordan håndteres et eventuelt behov for at anvende data til et andet formål (formålsforksydning)?
- Hvordan indhentes samtykke – og hvordan kan det dokumenteres?
- Hvornår indhentes fornyet samtykke?
- Hvornår indhentes forældres samtykke til behandling af børns persondata?
- Hvordan sikres det, at data er korrekte og bliver opdateret?
- Hvem kontrollerer, at data anvendes i overensstemmelse med persondataloven

Dataflowets output

Som led i databehandlingen kan den dataansvarlige have indgået aftaler med leverandører eller samarbejdspartnere om løsning af behandlingsopgaver. Det fx være nødvendigt at validere modtagne persondata hos en tredjepart, fx cpr-registeret eller en kreditvurderingsbureau. Eller den dataansvarlige kan have valgt at outsource dele af databehandlingen til en tredjepart, fx lønregnskab eller en it helpdesk, eller have indgået aftale om placering af tredjeparts cookies på en hjemmeside.

Der vil derfor typisk også være brug for at afdække dataflowets *output*. Kortlægningen heraf bør omfatte:

- Identifikation af leverandører og deres opgaver
- Leverandørens geografiske placering, herunder om det er indenfor eller udenfor EU
- Beskrivelse af de dataelementer, der videregives

- Kategorisering af dataelementerne i datatyper
- Beskrivelse af det lovlige grundlag for videregivelse
- Hvordan datasubjektet oplyses om videregivelse af persondata og formålet hermed

Datasikkerheden

For at sikre kontrol med, hvem der behandler persondata, og reducere risikoen for uautoriseret adgang til data, misbrug og læk af data, bør det i tilknytning til databehandlingsfasen også afdækkes, om der er vedtaget interne procedurer og processer for at opfylde kravene til passende behandlingssikkerhed (artikel 32), herunder om:

- Hvem har adgang til persondata?
- Hvem tildeler adgangsrettigheder til data?
- Hvem kontrollerer adgangen til data og anvendelsen af dem?
- Hvordan opbevares persondata? – lokalt, i cloud-løsning eller anden host
- Videregives data til en tredjepart? Fx til validering eller analyse
- Hvornår anvendes pseudonymisering og kryptering til at sikre persondata? under transmission og/eller opbevaring
- Hvor ofte foretages der back up på de systemer, hvor persondata behandles?
- Hvor ofte foretages pachting af anvendt software til databehandlingen?
- Hvordan håndteres sikkerhedsbrud? – er der fx vedtaget beredskabsplaner
- Hvem er ansvarlig for underretning om sikkerhedsbrud? Og hvordan eskaleres de til ledelsen?
- Hvor ofte afprøves, vurderes og evalueres sikkerhedsforanstaltningernes effektivitet?

Databehandlings ophør

Den afsluttende fase i datas livscyklus er dataflowets og behandlingens ophør. Databehandlingens ophør kan enten skyldes, at relationen til kunden ophører, fx ved at kunden opsiger tjenesten, trækker sit samtykke tilbage eller anmoder om at få overflyttet sine persondata til en anden udbyder eller benytter sin ret til at få data slettet. Eller der kan ske ændringer i det lovlige grundlag for indsamling og behandling af persondata, fx ved at en samfundsmæssig eller anden legitim interesse ikke længere er så tungtvejende, at hensynet til datasubjektets beskyttelse kan tilsidesættes.

I dataflowkortlægningen bør derfor indgå følgende oplysninger:

- Hvordan bringes databehandlingen til ophør?
- Hvem er ansvarlig for at slette data?
- Hvem kontrollerer, at data ikke opbevares længere end nødvendigt?

Forankring af dataflowkortlægninger

Alle virksomheder, der anvender persondata, bør kende deres dataflows. En kortlægning af dataflows skaber både grundlaget for at udføre konsekvensanalyser og vurdere organisationens risiko for at tilsidesætte beskyttelsen af datasubjektet, men bidrager også til opfyldelse af det nye krav om den dataansvarlige og databehandlerens dokumentation af behandlingsaktiviteter, som er indført med dataforordningens artikel 30.

For organisationer, hvor persondatabehandling indgår i deres kerneaktivitet, vil det derfor være hensigtsmæssigt at forankre opgaven med at kortlægge dataflows i organisationen og fastlægge processer for udarbejdelse af dataflowkortlægninger samt procedurer for, hvornår de skal udføres og hvem, der fører kontrol med at det sker.

Konsevkensanalyse

Med kravet om udførelse af konsekvensanalyser, introducerer dataforordningen et risikoparagdigme i persondataretten. De fleste virksomheder vil være fortrolig med metoden, fordi risikovurderinger allerede indgår i monitoreringen af forretningsdriften og sikkerhedsniveauet. Eksisterende redskaber til afdækning af aktuelle og potentielle risici, vurdering af sandsynligheden for, at de opstår, samt planer for mitigering, eliminering, forebyggelse samt kontroller og beredskab vil derfor kunne anvendes til håndtering af risikoen på persondataområdet.

Det nye element er imidlertid perspektivet: der skal nu også skal være fokus på den risiko, der er for at påvirke borgerne, kunderne og medarbejdernes rettigheder negativt. Det er med andre ord ikke længere alene risikoen for forretningen, der er i spil, men også risikoen for de konkrete personer, hvis data bliver behandlet som led i drift og opgaveløsning.

Høj risiko

Dataforordningen nævner en række specifikke risici og rubricerer nogle af disse som udgørende *høj risiko* for datasubjektets beskyttelse. Artikel 35 foreskriver kun, at der skal udarbejdes konsekvensanalyser vedrørende databeskyttelse i tilknytning til behandling, der *"sandsynligvis vil indebære en høj risiko for fysiske persons rettigheder og frihedsrettigheder".* Brug af nye teknologier nævnes som en konkret behandlingstype, der kan udløse høj risiko. Databehandling, der på grund af sin karakter, omfang, sammenhæng og formål, kan også udløse høj risiko medfører også et krav om at gennemføre konsekvensanalyse.

Sådan høj risiko er efter artikel 35, stk. 3 navnlig knyttet til behandling af persondata, der omfatter

- følsomme persondata og oplysninger om strafbare forhold i stort omfang
- systematisk overvågning af offentligt tilgængelige steder
- profilering, d.v.s. en form for automatisk databehandling, der har som formål at vurdere personlige forhold systematisk og omfattende, så de kan danne grundlag for afgørelser, der har retsvirkning for datasubjektet.

Andre typer risici

I tidligere udgaver af dataforordningen var der nævnt en række andre risikotyper, fx bestemmelserne om den dataansvarliges ansvar og om behandlingssikkerhed. I den vedtagne tekst henvises alene til risikoen for datasubjektets rettigheder og frihedsrettigheder. De tidligere oplistede eksempler på sådanne risici fremgår imidlertid fortsat af præambulære bestemmelser (75 og 85), og er derfor med til at udfylde kravene til den dataansvarlige i den nuværende artikel 24 og til behandlingssikkerheden i artikel 32.

Når man vurderer, om en organisations databehandling eller sikkerhedsniveau medfører en høj risiko for datasubjektet, kan præamblens eksempler anvendes som tjekliste. De negative konsekvenser af behandling af persondata kan omfatte:

- fysisk, materiel eller immateriel skade
- diskrimination
- identitetstyveri eller -svig
- finansielle tab
- skade på omdømme
- tab af fortrolighed for personoplysninger, der er omfattet af tavshedspligt
- uautoriseret ophævelse af pseudonymisering
- betydelige økonomiske eller sociale konsekvenser
- forhindring i at udøve kontrol med deres persondata
- evaluering af personlige forhold, navnlig analyse eller forudsigelse af forhold vedrørende indsats på arbejdspladsen, økonomisk situation, helbred, personlige præferencer eller interesser, pålidelighed eller adfærd eller geografisk position eller bevægelser, med henblik på at oprette eller anvende personlige profiler
- behandling af persondata om sårbare fysiske personer, navnlig børn

Den konkrete vurdering af behovet

Den konkrete vurdering af, om man som organisation er forpligtet til at udføre konsekvensanalyser, foretages bedst på baggrund af en dataflowkortlægning. Den skaber overblik og indsigt i såvel anvendte teknologier, som i formålet og omfanget af databehandlingen, behandlingens karakter og den sammenhæng den indgår i, men også i dataelementer og typer, og deres konkrete anvendelse.

Datatilsynet vil på et senere tidspunkt udarbejde lister over de behandlingsaktiviteter, der kræver konsekvensanalyser og evt. også over dem, der ikke gør. Indtil de er offentliggjort er et en god idé i hvert fald af foretage en indledende screening og en dataflowkortlægning som beskrevet ovenfor.

Konsekvensanalysens indhold

Det fremgår af artikel 35, at konsekvensanalysen mindst skal omfatte:

- en systematisk beskrivelse af de planlagte behandlingsaktiviteter
- formålene med behandlingen
- behandlingens lovlige grundlag
- en vurdering af, om der er et rimeligt forhold mellem formål og behandlingen (proportionalitet)
- en vurdering af risikoen for at påvirke datasubjektets rettigheder og frihedsrettigheder negativt
- etablerede eller påtænkte risikoreducerende foranstaltninger.
- overholdelse af godkendt adfærdskodekser efter forordningen kan også indgå i vurderingen.

De tre første punkter (og med tiden det sidste, når adfærdskodekser er en realitet) vil være afdækket med dataflowkortlægningen, ligesom den vil kunne danne grundlag for at foretage de beskrevne vurderinger af proportionalitet og risikoen for datasubjektets beskyttelse. Disse risikovurderinger vil så være bestemmende for, hvilke risikoreducerende foranstaltninger, der skal iværksættes.

Risikovurderingen

Vurderingen af risikoen for datasubjektets rettigheder og frihedsrettigheder kan man tage afsæt i en klassisk risikovurdering, hvor man starter med at forstå sit risikobillede. Når denne metode tilpasses persondataområdet, kan følgende spørgsmål oplyse beslutningsgrundlaget:

- Hvilke datatyper behandler vi? – kumulerer vi almindelige persondata eller anvender følsomme
- Hvor mange personer berører vores databehandling?
- Hvem udveksler vi persondata med? – er der modtagere udenfor EU?
- Hvad bliver data brugt til? – af os selv, vores databehandlere og deres underleverandører
- Hvordan er vores sikkerhed? Både i forhold til tilgængelighed, fortrolighed, integritet og robusthed
- Hvad er organisationens risikoprofil? – hvordan påvirker størrelse, struktur, kompleksitet, industri/branche og kultur risikoen for at tilsidesætte databeskyttelsen
- Hvori består truslen? – har vi ikke minimeret datamængden, undgået formålsforskydning, er sikkerheden ikke effektiv, eller mangler der databehandlingsaftaler

Mange af disse spørgsmål vil kunne besvares på baggrund af dataflowkortlægningen, fordi den netop vil afdække svagheder i den persondatabeskyttelse, der er etableret i tilknytning til konkrete behandlingsaktiviteter. Ud fra beskrivelserne af dataflowets tre faser, vil det således være muligt at vurdere det samlede risikobillede og de konkrete tiltag, som vil være nødvendige for at reducere eller eliminere risikoen via foranstaltninger, garantier og mekanismer, der sikrer beskyttelsen af persondata.

Sandsynligheden

Til risikovurderingen hører imidlertid også en vurdering af sandsynligheden for, at databeskyttelsen reelt tilsidesættes. Dataforordningen foreskriver, at den dataansvarlige skal vurdere den høje risikos specifikke sandsynlighed og alvor. Den sandsynlighedsvurdering skal inddrage databehandlingens karakter, omfang, sammenhæng og formål samt risikokilderne.

Også denne del af konsekvensanalysen kan med fordel integreres organisationens øvrige risikohåndtering- og vurdering, fx på informationssikkerhedsområdet. Et tilpasningsmoment her vil være den anvendte sandsynlighedsskala. Måles sikkerhedsrisikoen fx på en skala 1-5 eller 1-3, eller benyttes høj, medium og lav til at beskrive sandsynligheden for at risikoen får konkrete følger, må man tage stilling til, om den anvendte skala også er egnet til at måle sandsynligheden for negativ konsekvens og alvoren af en given data- behandling med høj risiko for det eller de berørte datasubjekter.

Forretningsmæssige fordele

Udover at beskytte kunder, borgere, medarbejdere og andre berørte personers rettigheder og legitime interesser bedre i forbindelse med behandling af persondata om dem, er der nogle forretnings- eller driftsmæssige fordele at kende sit dataflow.

Ved at arbejde systematisk med dataflowkortlægning og konsekvensanalyser opnår man som virksomhed og myndighed således større indsigt i databehandlingsaktiviteter og behandlingssikkerhed. Det er en viden, der kan bruges til at effektivisere arbejds- og kontrolprocesser, bl.a. gennem interne politikker, procedurer og retningslinjer, øge kvaliteten af databehandlingsaktiviteterne gennem medarbejderuddannelse og leverandørstyring, men også til at inspirere til innovativ produkt- eller procesudvikling. Sidstnævnte kan fx være en udløber af, at konsekvensanalysen afdækker behovet for at anvende redskaber til sikring af *data protection by design* og data *protection by default*.

Tilrettelæggelsen og udførelsen af dataflowkortlægningen og konsekvensanalysen vil derudover samlet set bidrage til, at et andet af de nye krav til persondatabehandling, nemlig kravet om dokumentation i dataforordningens artikel 30, bliver opfyldt og løbende opdateret.

Vil man være på plads med en effektiv persondatabeskyttelse i 2018 er dataflowkortlægningen på alle måde et godt første skridt.