Legal tech för dataskydd

Kan teknik lösa de problem som tekniken skapat?

Staffan Malmgren

1. Inledning

Dataskyddet har genom tiderna försökt att, med rättsliga förpliktelser och förbud, skydda den enskildes integritet mot de intrång som registrering och behandling av personuppgifter kan medföra (prop. 1973:33 s. 1). Registrering och annan behandling av personuppgifter har skett i alla tider men informationstekniken har möjliggjort en allt mer omfattande insamling av uppgifter och allt bredare användningsområden för de insamlade uppgifterna. Tekniken kan därför – oaktat att det är människor, företag och myndigheter som väljer att använda tekniken och bestämmer användningens ändamål – sägas medföra ett hot för den personliga integriteten, genom att den sänker kostnader och ökar värde på sådan behandling som exempelvis kan kartlägga enskilda personer. Man kan därför prata om *teknikens problem* för den personliga integriteten.

Rättsreglerna kring dataskyddet – exempelvis i dess utformning i dataskyddsförordningen – kan därför ses som *en juridisk lösning på teknikens problem*. Dessa rättsregler innehåller förutom konkreta förbud mot behandling i vissa situationer (exempelvis om det saknas en hållbar rättslig grund) även många förpliktelser för den som ansvarar för behandlingen. Exakt vilka förpliktelserna är har varierat genom åren och i takt med informationteknikens utveckling och samhällets normer. Den konstruktion som anvisades i 1973 års datalag, där varje registeransvarig behövde en licens för detta utgiven av Datainspektionen var kanske ändamålsenlig när upprättandet av ett register var en omfattande process. Det fungerar dock inte i dagens samhälle där en genomsnittlig kontorsarbetare snart sagt skapar ett nytt personregister så fort denne öppnar Excel.

Hur mycket man än skulle önska att alla problem kunde lösas med en välskriven lagparagraf så har den rättsliga regleringen av dataskyddet även gått hand i hand med en teknikutveckling som – åtminstone

i viss mån – strävar efter att på automatiserad väg stärka skyddet för den personliga integriteten. Detta kan ske genom att automatisera eller skapa verktyg för att tillämpa rättsreglerna på ett enhetligt och effektivt sätt, ibland genom att utforma systemen för själva personuppgiftsbehandlingen på ett sådant sätt att riskerna för integriteten minskas. I sammanhanget pratar man om Privacy Enhancing Technologies (PET), Privacy by Design (and Default) eller inbyggt dataskydd för att indikera att det handlar om teknik- och systemutformning som strävar efter att skydda den enskildes personliga integritet i samband med att behandling utförs.

Kan man då prata om en *teknisk lösning på teknikens problem*?

2. Legal tech och personuppgiftshantering

Med Legal tech avses i denna text sådana tekniska system som syftar till att stödja och underlätta juridiskt arbete på olika sätt. Marknaden för legal tech kan kategoriseras såväl utifrån vilken typ av arbete som systemen utför eller underlättar (dokumentgranskning, rättsinformationssökning, avtalsanalys mm) såväl som vilket rättsområde som systemen är inriktade på. I det följande inriktar vi oss på sådana system som på ett eller annat vis är inriktade på dataskyddsfrågor. Det juridiska arbetet inom dataskyddsområdet har varit föremål för omfattande Legal tech-utveckling och det finns idag en stor marknad för sådana system och tjänster. Vi gör i det följande en grov indelning i sådana verktyg som underlättar regelefterlevnad respektive sådana som utgör eller bidrar till tekniska skyddsåtgärder vid personuppgiftsbehandling.

3. Regelefterlevnad

Det finns ett stort antal rättsregler i dataskyddsförordningen som ålägger den personuppgiftsansvarige att i olika situationer utföra vissa åtgärder. Exempelvis föra ett register över personuppgiftsbehandlingar, utföra en konsekvensbedömning, skapa ett registerutdrag, ingå ett personuppgiftsbiträdesavtal eller anmäla en personuppgiftsincident till en tillsynsmyndighet.

Informationssystem som syftar till att hjälpa den personuppgiftsansvarige att säkerställa att sådana regler efterlevs (*compliance*) är en stor och växande marknad. Systemen tillhandahålls vanligen som molntjänster, och utvecklas kontinuerligt för att vara i linje med dataskyddslagstiftningen. Ett exempel är de överföringsbedömningar (transfer impact

assessment) som inte uttryckligen regleras i dataskyddsförordningen, men som Schrems II-domen och den efterföljande EDPB-rekommendationen (01/2020) anvisat som krav innan en personuppgiftsöverföring till ett tredjeland genomförs vid användning av s.k. standardavtalsklausuler. De system som finns på marknaden har relativt snabbt infört stöd för de sex steg som EDPB-rekommendationen anvisar.

Det stora värdet i ett system för regelefterlevnad är förstås inte begränsat till dataskydd. Marknaden för regelefterlevnadssystem är inte heller en exklusiv fråga för dataskyddsregleringen. Olika systemleverantörer är mer eller mindre ambitiösa, och riktar sig till delvis olika målgrupper, när de bygger in stöd för regelefterlevnad av övriga rättsområden, exempelvis NIS-lagstiftningen, visselblåsarregleringen eller penningtvättslagstiftning.

3.1 Kartläggning av och register över behandlingar

En vanlig utgångspunkt för regelefterlevnadssystem för dataskydd är den förteckning över personuppgiftsbehandling som den personuppgiftsansvarige eller ett personuppgiftsbiträde som huvudregel måste föra enligt artikel 30. Ett sådant register ska innehålla ett antal uppgifter, bl.a. ändamål med behandlingen, kategorier av registrerade och kategorier av uppgifter som behandlas, kategorier av mottagare mm. Utöver de skyldigheter som följer direkt av artikel 30 behöver en personuppgiftsansvarig även hålla ordning på exempelvis rättslig grund för behandlingen (art. 13.1 c), var de behandlade uppgifterna kommer från (art. 15.1 g) m.m. – sådana uppgifter kan också med fördel lagras i ett artikel 30-register. Utöver sådana krav som direkt följer av lagstiftningen kan ett sådant register även användas för att hålla ordning på vilka underleverantörsrelationer som finns för behandlingen och tillhörande personuppgiftsbiträdesavtal, bedömningar av tredjelandsöverföringar m.m.

Även om många kan komma en lång bit på vägen mot att hålla ordning på sina behandlingar genom en enkel förteckning upprättad i exempelvis Excel, så finns ett värde i ett mer komplett register som är sammanflätat med funktionalitet för exempelvis riskbedömningar, samt att kunna skapa rapporter eller ”dashboards” över en organisations totala personuppgiftsbehandling och i vilken utsträckning man efterlever lagstiftningen i alla delar. Detta är kärnan i de flesta system för regelefterlevnad på marknaden.

3.2 Stöd för dataskyddsrättsliga bedömningar

För varje behandling behöver en personuppgiftsansvarig göra en åtminstone översiktlig bedömning av om behandlingen sannolikt kan leda till en hög risk för de registrerade. Om så är fallet behöver den personuppgiftsansvarige göra en mer formaliserad konsekvensbedömning (Data Protection Impact Assessment, DPIA). Integritetsskyddsmyndigheten har även upprättat en ”svart lista” över sådana behandlingar som alltid ska anses omfattas av kravet på konsekvensbedömning.[1] (Tillsynsmyndigheter har även möjlighet att upprätta en ”vit lista” på behandlingar som aldrig kräver konsekvensbedömning, men Integritetsskyddsmyndigheten har för närvarande inte planer på att göra en sådan förteckning.[2]) Dataskyddsförordningen anger inte exakt hur en sådan bedömning ska göras men anger i artikel 35.7 vissa grundkrav på vad den färdiga bedömningen ska innehålla. Som kompletterande vägledning har Europeiska dataskyddstyrelsens (EDPB) föregångare, Artikel 29-gruppen, gett ut riktlinjer om konsekvensbedömningar, inklusive hur den inledande bedömningen om huruvida en viss behandling sannolikt leder till hög risk, ska göras.[3]

Utöver detta har även vissa tillsynsmyndigheter, dock inte Integritetsskyddsmyndigheten, utvecklat metoder för hur en konsekvensbedömning ska göras. Dessa metoder förutsätter bl.a. ett strukturerat arbetssätt där risker kartläggs, värderas och mappas mot skyddsåtgärder avsedda att minska dessa risker. Arbetsgången är på hög nivå inte olik den inventering av behandlingar inom en organisation som måste göras enligt artikel 30 i det att det handlar om att på förhand försöka inventera och bedöma olika aspekter av behandlingen.

Franska tillsynsmyndigheten Commission Nationale de l’Informatique et des Libertés (CNIL) har kompletterat sin metodik för att utföra konsekvensbedömningar[4] med ett mjukvaruverktyg (PIA) som underlättar kartläggning- och mappningsarbetet samt den slutliga bedömningen av hur stor risken för de registrerade blir med beaktande av planerade skyddsåtgärder.[5] Verktyget erbjuds, till skillnad mot vad som är vanligast på den kommersiella marknaden, inte som en SaaS-tjänst, utan istället

[1] IMY, ”Förteckning enligt artikel 35.4 i Dataskyddsförordningen”, Dnr DI-2018-13200, https://www.imy.se/globalassets/dokument/ovrigt/forteckning---konsekvensbedomningar.pdf.
[2] https://www.imy.se/verksamhet/dataskydd/det-har-galler-enligt-gdpr/konsekvensbedomningar-och-forhandssamrad/dataskyddsforordningen-om-konsekvensbedomningar-och-forhandssamrad/.
[3] Artikel 29-arbetsgruppen för skydd av personuppgifter, ”Riktlinjer om konsekvensbedömning avseende dataskydd och fastställande av huruvidabehandlingen ”sannolikt leder till en hög risk” i den mening som avses i förordning 2016/679”, WP 248 rev. 01.
[4] https://www.cnil.fr/en/PIA-privacy-impact-assessment-en.
[5] https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment.

är källkoden tillgänglig under open source-licensen GPL 3.0, vilket gör det möjligt för en organisation att själva driftsätta och hantera de uppgifter om högriskbehandlingar som krävs för en konsekvensbedömning.

Utöver konsekvensbedömningar har liknande bedömningsprocesser vuxit fram inom dataskyddsområdet, exempelvis överföringsbedömningar (Transfer Impact Assessment, TIA) och bedömningar av legitimt intresse (Legitimate Interest Assessment, LIA). På marknaden börjar även verktygsstöd för sådana bedömningar att erbjudas, ofta inom ramen för sådana övergripande kartläggningsverktyg som beskrivs i avsnitt 3.1

Gemensamt för dessa approacher är att verktygen följer en metodik, som i sin tur är utformad för att leda till ett resultat som följer lagstiftningen. Metodiken kräver en viss utbildningsinsats hos de som ska använda den på ett effektivt sätt, och verktygens förmåga att med intuitiva gränssnitt m.m. utbilda i denna metodik är begränsade. Å andra sidan, när den person som ska utföra bedömningarna har tillägnat sig metodiken kan denne ofta ställa sig frågan vad verktygen egentligen tillför jämfört med till exempel en enklare Excellista eller Wordmall.

3.3 Ärendehantering för utövande av registrerades rättigheter

Redan vid sitt införande 1973 innehöll datalagen i sin 10 § vissa regler för registrerades rättigheter och hur dessa skulle utövas vad gällde registerutdrag (idag oftare benämnt rätten till tillgång). Dataskyddsförordningen har i jämförelse en mer omfattande katalog av rättigheter i artiklarna 15–22 som den registrerade kan på begäran kan göra gällande mot den personuppgiftsansvarige, och en mer detaljerad reglering i artikel 12 kring hur den personuppgiftsansvarige ska se till att dessa rättigheter kan utövas enkelt. Detta omfattar aspekter som tidsfrister, hur en begäran ska ställas, hur den registrerade ska identifieras i samband med utövande av sina rättigheter och på vilket format som begärda uppgifter ska lämnas.

Allt detta kan sammantaget läsas som en *kravspecifikation på ett ärendehanteringssystem* för utövande av registrerades rättigheter. Ett sådant system – varav marknaden främst har inriktat sig på begäran om tillgång (Data Subject Access Request, DSAR) – hjälper den personuppgiftsansvarige med att ta emot begäran och besluta om hur begäran ska hanteras. Vad gäller de faktiska åtgärderna för att exempelvis sammanställa de uppgifter som behandlas, eller för att utföra en radering av uppgifter enligt rätten att bli glömd, varierar det i vilken utsträckning ett sådant ärendehanteringssystem faktiskt kan hjälpa till med sådana åtgärder. Såvida inte ärendesystemet är integrerat med det eller

de verksamhetssystem där uppgifterna behandlas krävs ofta manuella handgrepp av en anställd eller uppdragstagare hos den personuppgiftsansvariga.

För verksamheter som har omfattande kontroll över de verktyg och system som behandlar personuppgifter (exempelvis för att verksamheten byggt dessa system själva och/eller hanteringen är en del av kärnverksamheten) är det vanligare att system för både tillgång enligt artikel 15 och dataportabilitet enligt artikel 20 är integrerade med de verksamhetssystem som hanterar uppgifterna på så sätt att processen kan göras helt automatiskt.[6] Som Integritetsskyddsmyndighetens beslut mot Spotify visar är detta dock inte en garanti för att en tillsynsmyndighet ska anse att de registrerades rättigheter är fullständigt tillgodosedda.[7]

3.4 Incidenthantering

Alla verksamheter som behandlar personuppgifter behöver ha en beredskap för att hantera personuppgiftsincidenter enligt dataskyddsförordningens krav. Särskilt de tidsfrister som anges i artikel 34 ställer hårda krav på en organisations förmåga att snabbt utreda, bedöma och åtgärda incidenter. Detta låter sig, likt systemen som beskrivs i avsnitt 3.3, utformas som ett ärendehanteringssystem. Eftersom utredning av en personuppgiftsincident ofta kan kräva samverkan mellan olika delar av verksamheten blir det viktigt att ett sådant system håller reda på vilket steg incidenthanteringen befinner sig i, vem som har ansvaret för att driva processen vidare och att inga tidsfrister överskrids.

Av särskild vikt för ett sådant system är dels hjälp med att löpande dokumentera incidentutredningen allt eftersom mer kunskap framkommer, att erbjuda stöd för att på ett enhetligt sätt bedöma incidentens allvarlighetsgrad för att bedöma om den måste anmälas till tillsynsmyndigheten samt om de registrerade som påverkas av den behöver informeras.

Integritetsskyddsmyndigheten har tagit fram ett verktyg för anmälan av personuppgiftsincidenter.[8] Detta verktyg adresserar dock bara en del av det fullständiga ansvaret för att hantera incidenter som faller på en personuppgiftsansvarig, varför det inte kan ersätta de system som finns på marknaden.

[6] Exempelvis Google Takeout (https://takeout.google.com/), LinkedIn Export your data (https://www.linkedin.com/mypreferences/d/download-my-data) eller Spotify Download your data (https://www.spotify.com/us/account/privacy/).
[7] Integritetsskyddsmyndighetens beslut 2023-06-12, Dnr DI-2019-6696.
[8] https://www.imy.se/verksamhet/utfora-arenden/anmal-personuppgiftsincident/risk-for-registrerade/.

3.5 Sammanfattning regelefterlevnad

Verktyg för regelefterlevnad har, som termen antyder, de juridiska reglerna som utgångspunkt. En övergripande ”risk” med dessa verktyg är att de fokuserar dataskyddsarbetet på att efterleva de regler som är till för att skydda de registrerades personliga integritet snarare än att vara direkt inriktat på att skydda denna. Det är förstås inte det sämsta, och utifrån ett affärsperspektiv ofta det enda rimliga. Ett alternativt förhållningssätt är dock att använda verktyg för att direkt minska eller eliminera risker för de enskilda. Ett verktyg som är utformat för att säkerställa regelefterlevnad av vissa specifika regler (för vår del den allmänna dataskyddsförordningen, för andra jurisdiktioner kanske Europarådets dataskyddskonvention eller amerikansk delstatslagstiftning) går sällan längre än vad dessa regler kräver, även om ytterligare funktionalitet hade kunnat underlätta ett höjande av skyddsnivån.

I kontrast förhåller sig tekniska skyddsåtgärder för att skydda registrerades integritet mer fristående från dataskyddsregleringen och erbjuder ett mer ”hands on”-perspektiv för att åstadkomma ett effektivt skydd. Detta leder i sig inte nödvändigtvis till bättre (eller sämre) integritetsskydd jämfört med verktyg för regelefterlevnad, och en heltäckande approach för dataskydd behöver förstås ta med båda förhållningssätten.

4. Tekniska skyddsåtgärder

4.1 Inbyggt dataskydd och dataskydd som standard

En utgångspunkt för dataskyddsförordningen är att den behandling av personuppgifter som utförs ska utformas på ett sådant sätt att själva behandlingen ska innehålla åtgärder för att säkerställa att kraven enligt förordningen uppfylls och att de registrerades rättigheter skyddas (Data protection by design and by default, DPbDD). Denna skyldighet beskrivs i artikel 25, som tillsammans med artikel 32 innehåller den kortfattade vägledning som dataskyddsförordningen erbjuder till vilka tekniska skyddsåtgärder som kan tänkas vara lämpliga. Två åtgärder som särskilt lyfts fram är kryptering respektive pseudonymisering av personuppgifter.

Det här är dock bara två tänkbara tekniska åtgärder till skydd för de registrerade. Marknaden för kompletterande skyddsåtgärder är omfattande och under snabb utveckling. Det är inte ovanligt med pressreleaser som lovar tekniska skyddsåtgärder som löser all dataskyddsrättslig huvudvärk. Sådana åtgärder behöver dock analyseras noggrant för att förstå i vilka sammanhang de överhuvudtaget är tillämpliga, vilka

säkerhetsgarantier de erbjuder och i vilken utsträckning detta påverkar skyddsnivån för de personuppgifter som ska behandlas.

4.2 Kryptering

Kryptering är den process enligt vilken information (*cleartext*) kan transformeras till data som innehåller samma information men som är oläslig (*ciphertext*) såvida man inte har tillgång till kompletterande information (*secret*, vanligtvis benämnd ”nyckel”) som låter en reversera transformationen.[9] Den som lagrar information, inklusive personuppgifter, i krypterad form behöver inte vara orolig för att denna röjs så länge inte nyckeln röjs och krypteringslösningen i övrigt är korrekt implementerad.

Problemet är dock att det i de flesta scenarion är svårt eller omöjligt att göra något intressant med den lagrade informationen – mer än att just lagra den – utan att nyckeln hanteras strax i närheten av samma part. Detta är särskilt fallet när man använder molntjänster. Beroende på vilken ansvarsfördelning mellan användare och leverantör av molntjänst (om det är fråga om infrastrukturtjänster/IaaS, plattformstjänster/PaaS, slutanvändartjänster/SaaS eller någon annan leveransform) kan det finnas större eller mindre möjlighet att som användare av molntjänster utöva en meningsfull kontroll över nyckeln. Det är dock svårt och i många sammanhang omöjligt att säkerställa att den nyckel som används, eller att den information som krypterats med densamma, aldrig kan komma att röjas, om nyckeln lagras hos en annan part.

4.3 Homomorfisk kryptering och confidential computing

I avsnitt 4.2 anfördes att det är svårt till omöjligt att säkerställa att krypterad information inte kan komma att röjas i klartextform om man inte utövar exklusiv kontroll över nyckeln. Detta är också EDBPs utgångspunkt i sina rekommendationer om kompletterande åtgärder för överföringsmekanismer.[10] I punkt 84.6 anges att ett villkor för att kryptering ska vara en effektiv skyddsåtgärd vid överföring till tredjeland är att nyckeln bevaras under den personuppgiftsansvariges exklusiva

[9] Vi beskriver i detta avsnitt endast s.k. symmetrisk kryptering och krypterad *lagring* (encryption at rest). Assymetrisk eller public key-kryptering har andra egenskaper särskilt vad gäller nyckelhantering, som gör den till en mer självklar del av krypterade *överföringar* (encryption in transit).

[10] EDPB, “Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data”.

kontroll.[11] I punkt 94 anges vidare som utgångspunkt att molntjänster som för sin funktion kräver att molntjänstleverantören har tillgång till uppgifterna i klartext inte kan förses med kompletterade tillräckliga skyddsåtgärder. EDPB utesluter dock inte att den tekniska utvecklingen kan leda till åtgärder som låter molntjänstleverantören utföra den önskade affärsnyttan utan att ha tillgång till uppgifterna i klartext.

Två tänkbara åtgärder för att möjliggöra detta är s.k. homomorfisk kryptering och olika hårdvaruskydd som med ett samlingsnamn beskrivs som ”confidential computing”.

Homomorfisk kryptering kan enkelt beskrivas som en kryptering där man kan utföra beräkningar eller vidta andra åtgärder på krypterad data, och få ett krypterat resultat som kan göras läsbart för den som har nyckeln. En molntjänstleverantör kan alltså ges i uppdrag att köra ett datorprogram som enbart behandlar data i krypterad form, utan att ges tillgång till nyckeln. För att detta ska fungera krävs att molntjänstkunden (eller någon som uppdragits av denne) skriver och kompilerar själva programmet med hjälp av mjukvarukomponenter för att åstadkomma beräkningar genom homomorfisk kryptering. En ytterligare svaghet med tekniken är att program för honomorfisk kryptering är i storleksordningen en miljon gånger långsammare än traditionella program, vilket omöjliggör de flesta praktiska tjänster.

Confidential computing åstadkommer liknande resultat men genom att använda de hårdvarufunktioner som finns i moderna processorer för att skapa avskilda utrymmen (trusted execution environment, TEE) inom datorns exekveringsmiljö där endast de program som användaren angett har tillgång till lagrad information. Genom att lagra krypteringsnyckeln för de skyddade uppgifterna i den TEE som molntjänstanvändaren har exklusiv tillgång till kan denne säkerställa att nyckeln inte röjs samtidigt som användaren har full kontroll över all behandling som sker inuti den TEE som används. Detta gäller dock endast under förutsättning att hårdvarufunktionerna fungerar som avsett i alla aspekter och att inga säkerhetsbrister som skulle möjliggöra extrahering av information i en TEE till den omgivande exekveringsmiljön. Sådana och liknande hårdvarusäkerhetshål har dessvärre upptäckts under senare år. En ytterligare svårighet (som hårdvaru- och molntjänstleverantörer arbetar på) är hur användaren på distans ska kunna säkerställa att den TEE och den hårdvaruplattform man som användare har tillgång till faktiskt är uppsatt och fungerar som avsett.

[11] Eller ”by an entity trusted by the exporter in the EEA or under a jurisdiction offering an essentially equivalent level of protection to that guaranteed within the EEA”, vilket utesluter molntjänstleverantörer i andra jurisdiktioner.

Gemensamt för båda dessa skyddsåtgärder är att de kräver att ansvarsfördelningen i molntjänster viktas kraftigt åt kundens håll – det är kunden som måste ansvara för vilken kod som körs, trots att en drivande anledning till att använda molntjänster är att molntjänstleverantören ofta är mer lämpad att ta detta ansvar. Detta utesluter i dagsläget att åtgärderna kan användas för färdiga slutanvändartjänster (SaaS) där affärsmodellen utgår från att tjänsteleverantören tar just detta ansvar.

4.4 Anonymisering och pseudonymisering

Anonymisering och pseudonymisering är två åtgärder för att transformera information som innehåller personuppgifter som har det gemensamt att uppgifter som kan användas för att identifiera fysiska personer avlägsnas. För att informationen ska vara anonymiserad i dataskyddsförordningens mening krävs att det inte finns något sätt att återidentifiera den som informationen avser, oavsett av vem som gör det eller med vilken kompletterande information (jfr skäl 26) det görs. För pseudonymisering tillåts däremot existensen av sådana kompletterande uppgifter som tillåter återidentifiering (exempelvis en ”kodnyckel” som mappar id-nummer till identifierande uppgifter) så länge som dessa förvaras separat och skyddas (jfr artikel 4.5). Om uppgifterna faktiskt är anonymiserade och det inte finns någon väg att tillskriva uppgifterna till en fysisk person så omfattas behandling av dessa inte av dataskyddsförordningen.[12]

För en begränsad mängd användningsområden – framförallt inom dataanalys och statistisk bearbetning – kan pseudonymisering utgöra en användningsbar skyddsåtgärd. Pseudonymisering är dock en betydligt mer omfattande åtgärd än att bara byta ut direkt identifierande uppgifter som namn, personnummer och adresser. Anledningen är att ju fler uppgifter man har om en individuell fysisk person, desto lättare är det att genom uteslutning säkerställa vem personen är – särskilt med hjälp av andra publika eller privata uppgiftssamlingar.

Verktyg för anonymisering och pseudonymisering behöver ta höjd för risken för sådana baklängesidentifieringar genom att exempelvis bedöma varje enskild uppgift, och varje kombination av uppgifter, förmåga att särskilja en individ från en annan. Beroende på vilka uppgiftssamlingar som anonymiseras eller pseudonymiseras kan krävas att uppgifter generaliseras (exempelvis att lagra ålderskategorier istället för födelseår) eller obfuskeras (exempelvis att löneuppgifter ändras uppåt

[12] Själva processen att transformera uppgifterna till anonymiserad form är dock en personuppgiftsbehandling som omfattas av förordningen.

eller nedåt på ett sätt som försvårar identifikation men utan att ändra uppgifternas statistiska egenskaper).

4.5 Syntetisk data

I många sammanhang, exempelvis i samband med dataanalys enligt avsnitt 4.4, men också i samband med mjukvaruutveckling eller maskinlärning/AI, finns behov av datamängder som inte nödvändigtvis innehåller riktiga personuppgifter, men som på en övergripande nivå ”uppför sig” som sådan data. Det kan handla om fejkade kundregister med en fördelning i ålder, kön, boendeort mm som motsvarar ett riktigt kundregister. Att skapa sådan syntetisk data kan ses som nästa steg av anonymisering där man skapar datamängder som inte på något vis härstammar från individuella personuppgifter men som ändå kan användas som om det var riktiga personuppgifter.

Det finns ett fåtal produkter på marknaden för att skapa sådana syntetiska datamängder med bas från den personuppgiftsansvariges riktiga datamängder. Det finns även ett stort intresse från EU-nivå där EDPS exempelvis tidigare har utsett syntetisk data som en kommande trend under 2022/2023.[13]

4.6 Identifiera personuppgiftsbehandling i ostrukturerat material

Att veta att man behandlar personuppgifter, och vilka dessa personuppgifter är, låter sig ganska lätt göras i strukturerat material, exempelvis ett kundregister. Men verksamheter behandlar stora mängder personuppgifter i ostrukturerat och semistrukturerat format, exempelvis epost (där sändare och mottagare anges i strukturerad form, medan information i brevkroppen är mestadels ostrukturerat) eller övervakningsvideomaterial (där frågan om huruvida personuppgifter förekommer beror på vad som förekommit framför linsen).

En generellt användbar kategori av tekniska åtgärder är därför verktyg som hittar och agerar på personuppgifter i ostrukturerad format. Detta kan användas för att säkerställa att en organisation inte skapar personuppgiftsbehandlingar och personregister utanför den personuppgiftsansvariges kontroll.

Med hjälp av upptränade AI-modeller kan sådana verktyg med relativt god träffsäkerhet identifiera personuppgifter i löpande text, bild och video så länge som personuppgifterna liknar vad AI-modellen tränats

[13] https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en.

på (om modellen endast har tränats på att identifiera personnamn och adresser kommer den inte automatiskt att kunna identifiera registreringsnummer eller fastighetsbeteckningar). De kan dock sällan garantera en hundraprocentig träffsäkerhet ens på sådana uppgifter som den tränats på. Användningen av sådana modeller kan i sig medföra risker för den personliga integriteten och utgöra självständiga personuppgiftsbehandlingar. Som ett av flera verktyg kan de dock utgöra en meningsfull komponent i ett generellt regelefterlevnadsarbete, särskilt vad gäller inventering av personuppgiftsbehandling som förekommer.

5. Sammanfattning

De två övergripande kategorier av tekniska lösningar – regelefterlevnad respektive tekniska skyddsåtgärder – som beskrivits i denna text innehåller lösningar på många små problem för den personliga integriteten som en omfattande informationslagring och – behandling för med sig. De kräver dock mycket av sina användare för att både användas på ett effektivt sätt som för att förstå när de ger någon nytta att använda överhuvudtaget. Drömmen om en heltäckande DPaaS-lösning (Data Protection as a Service) lever i vissa delar av marknaden men kommer knappast kunna vara en ersättare till kunniga och samvetsgranna dataskyddsspecialister. Men även om tekniken inte erbjuder en paketlösning för de problem som den skapat kan den tillhandahålla en verktygslåda med vars hjälp enskildas integritet kan skyddas långt mer effektivt än genom uteslutande juridiskt arbete.