Promemoria

Stordataanalyser och datasjöar – begrepp och rättsliga förutsättningar

ES2023-2

Februari 2023

Innehåll

1. Inledning

Stordataanalyser och datasjöar framhålls ofta som en viktig förutsättning för att stärka det datadrivna arbetssättet. Vinsterna med utökad analys inom offentlig sektor kan vara många. Exempelvis skulle hälsofrämjande verksamheter kunna bli bättre på att förutse ärftliga sjukdomar och bidragsutbetalande myndigheter skulle lättare kunna upptäcka bedrägerier. Samtidigt finns det rättsliga utmaningar, bland annat avseende möjligheten att sammanfoga och dela data samt kraven på informationssäkerhet och integritetsskydd.

I denna promemoria tittar vi närmare på begreppen och de rättsliga förutsättningarna.

1.1 Syfte

Syftet med denna promemoria är att belysa innebörden av begreppen *stordataanalyser* och *datasjöar*. Syftet är vidare att beskriva de rättsliga förutsättningarna för stordataanalyser och datasjöar, samt vilka möjligheter respektive utmaningar som kan föreligga för myndigheter som vill arbeta med stordataanalyser eller datasjöar.

Promemorian riktar sig i första hand till jurister, arkitekter och utvecklare, men även till myndighetsledning och en vidare krets såsom leverantörer till offentlig sektor.

1.2 Avgränsningar

Denna promemoria är avgränsad till rättsliga förutsättningar för stordataanalyser och datasjöar. Promemorian går inte i detalj in på på hur de rättsliga förutsättningarna kan hanteras i praktiken. Datalager beskrivs som begrepp för att tydliggöra skillnaden mellan ett datalager och en datasjö, men ingen fördjupad rättslig bedömning görs av förutsättningar för datalager.

1.3 Medverkande

Arbetet har genomförts av en arbetsgrupp bestående av Johanna Sahlman, Catrine Nylén, Ann Svensson, Tina Chavoshi och Linda Lindström. Kvalitetssäkring har skett i eSams rättsliga expertgrupp, expertgruppen i säkerhet samt koordineringsgruppen för arkitektur. Beredning har skett via eSams samordningsgrupp.

2. Begreppen stordataanalys och datasjöar

2.1 Inledning

Stordataanalyser och datasjöar kommer ofta upp i sammanhang där verksamhetstveckling diskuterats, särskilt när det kommer till utveckling med hjälp av AI. Dock kan konstateras att begreppen används lite olika varvid förvirring lätt uppstår om vad det är som avses i praktiken. I följande avsnitt görs beskrivningar i syfte att ge ett stöd till myndigheter vid användning av begreppen. Beskrivningarna ska inte ses som uttömmande definitioner.

2.2 Stordata (Big Data)

Det finns ingen enhetlig definition av stordata eller det engelska begreppet Big Data. Ofta nämns det i sammanhanget omfattande datamängder där målet är att analysera stora mängder data så att datan blir meningsfull och skapar värde.[1]

Enligt Europaparlamentet avses med stordata datamängder som är så stora och komplexa att de kräver ny teknik, såsom AI, för att kunna bearbetas.[2]

Ofta förklaras stordata utifrån tre V:n– Volume, Variety, Velocity (volym, variation, hastighet). Volym refererar till magnituden av data som kommer från olika källor och det är fråga om datavolymer på flera tera- och petabytes eller till och med exa- och zettabytes. Storleken ökar hela tiden. Variation avser att data kommer i alla typer av format, strukturerad, ostrukturerad och semistrukturerad. Hastighet tar sikte på hur snabbt data skapas och hur fort den eventuellt analyseras. Hastigheten i vilken data skapas ökar för varje år, vilket i sin tur har ökat kraven för realtidsanalys.

Utöver dessa tre V nämns också andra dimensioner. Exempelvis används upp till sjutton V:n och ett C för att beskriva stordata (volume, velocity, value, variety, veracity, validity, visualization, virality, viscosity, variability, volatility, venue, vocabulary, vagueness, verbosity, voluntariness, and versatility and complexity).[3]

[1] Bello-Orgaz, Jung & Camacho, Social big data: Recent achievements and new challenges (2016)
[2] https://www.europarl.europa.eu/news/sv/headlines/society/20210211STO97614/vad-ar-big-data-definition-fordelar-och-utmaningar
[3] International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 09 | Sep-2017

2.3 Stordataanalys (Big Data Analytics)

Stordataanalyser handlar om att analysera stordata, både strukturerad och ostrukturerad (se avsnitt 2.2), för att genom analysen kunna hitta mönster, okända korrelationer eller annan, ännu inte känd, information. Ostrukturerad data är exempelvis data i dokumentform, men skulle också kunna vara uppgifter i system, såsom till exempel handläggarnoteringar i fritext. Med strukturerad data avses vanligen data som finns i databasform. Det finns ingen enskild teknik som utgör stordataanalys, utan det är ofta flera typer av teknologier som samverkar, exempelvis maskininlärning, datautvinning och prediktiva analyser.[4]

2.4 Datalager (Data Warehouse)

Ett datalager beskrivs vanligen som en databas med information från ett flertal datakällor, sammanställd på ett sätt som underlättar sökningar och analys. Datalager används ofta för informationsutvinning och beslutsstöd. Ett datalager innehåller strukturerad och filtrerad data som avses att användas för, eller redan har bearbetats för, något specifikt ändamål. Inom myndigheter kan datalager innehålla strukturerad data från olika källsystem som används i kärnverksamheten, till exempel från myndighetens handläggningssystem.

2.5 Datasjö (Data Lake)

En datasjö innehåller stora mängder uppgifter både i strukturerad och ostrukturerad form. Uppgifter kan komma både från den egna verksamheten eller från externa aktörer. Till skillnad från traditionella datalager (se avsnitt 2.4) så tvättas och sorteras inte informationen utan data lagras i sin ursprungsform, i olika former och format på en samlad lagringsyta. Syftet med data i en datasjö är inte alltid definierat i förväg. Det kan istället betraktas som en lagring av rådata fram till dess att den behövs. Data i datasjöar är mer svårtillgängligt än i datalager och kräver mer vana och andra verktyg för att bearbetas och analyseras.

[4] Prediktiv analys innebär att man gör statistiska analyser av stora mängder data för att hitta mönster som sedan kan användas för att göra förutsägelser. Detta görs ofta med hjälp av maskininlärning. (https://main.exedsse.se/vad-betyder/prediktiv-analys)

3. Nuläge och tendenser inom offentlig verksamhet

Som framförts inledningsvis kan det finnas många möjligheter med stordataanalyser inom offentlig sektor. Till exempel skulle Polisen med sådan teknik kunna bli bättre på att förutse var det kan begås kriminella handlingar. Hälsosektorn skulle bättre kunna förutse sjukdomar och ge träffsäkrare behandling av patienter. Skatteverket och de bidragsutbetalande myndigheterna skulle kunna bli bättre på att upptäcka bedrägerier. Trafiksektorn skulle också kunna styra trafik efter väderförhållanden och reagera snabbare på exempelvis miljökatastrofer.

Vid analyser av stora mängder uppgifter kan uppgifterna lagras i en datasjö hos den egna myndigheten. Det förekommer också så kallade federerade lösningar, se avsnitt 3.2.

3.1 Nuläge inom eSam

Inom ramen för denna promemoria har frågan ställts till några av eSams medlemmar för att få användningsfall beskrivna. Det har visat sig att det bland medlemmarna inte finns så många exempel på datasjöar enligt den beskrivning som anges i avsnitt 2.5. Däremot beskriver flera att de har pågående utvecklingsarbeten med olika former av datalager, se exempel i 3.1.1. Det finns inte heller några tydliga exempel på stordataanalyser, vilket kan förklaras av att det är få medlemsmyndigheter som har tillräckligt stora datamängder eller möjlighet att hantera data med den hastighet som stordataanalyser innebär. Det kan även bero på rättsliga begränsningar, se avsnitt 4.

Det framstår också som att behovet bland eSams medlemmar mer handlar om att kunna analysera den data myndigheten själv har samt att även kunna hämta in data från andra myndigheter (exempelvis i brottsbekämpande syfte). Till exempel anges att myndigheter skulle vilja kunna titta på en persons alla ärenden inom myndigheten, oavsett ärendekategori, eller inom en särskild kategori såsom utbetalning. Det uttrycks också ett behov av att kunna analysera data direkt mot den som innehar datan och inte behöva lagra data själv, i syfte att undvika onödig dubbellagring av stora datamängder.

3.1.1 Exempel på arbete med datalager

Ett exempel på utvecklingsarbete med olika former av datalager är Skatteverkets arbete. Skatteverket använder inte stordata (enligt beskrivningen ovan) men utför omfattande analyser av den data som finns i myndigheten. Myndigheten har inte någon datasjö.

I dag används i huvudsak Informationslagret, som är Skatteverkets datalager (Data Warehouse). Här samlas olika verksamheters information på ett strukturerat sätt för att kunna användas för analys. Det innebär lagring av analysdata och en fysisk koppling mellan olika data. Informationslagret består också av ett gränssnitt som ger åtkomst till data för myndighetens analys och rapporteringsverktyg.

Data i Informationslagret struktureras och hanteras på ett sådant sätt att respektive verksamhet bara får tillgång till sin egen data. Ingen verksamhet har alltså tillgång till all data. Internt kallas det för logisk separation och är nödvändigt för att bland annat kunna efterleva tillämplig sekretesslagstiftning och dataskyddsreglering.

För att utveckla Skatteverkets analysförmåga behöver nya it-lösningar tas fram eller upphandlas. Målsättningen är att kunna hämta data direkt från källan och bli mindre beroende av att ha data i separata lösningar som Informationslagret. Ett Logiskt datalager ska därför införas. En del av den arkitektur som benämns som Logiskt datalager utgörs av Accesslagret. Accesslagret blir kopplingen mellan olika data, virtuellt eller logiskt. Det innebär att myndigheten kommer att kunna nå data från flera källor och blir inte lika beroende av datalager som Informationslagret.

På samma sätt som för den data som finns i Informationslagret behöver ett Logiskt datalager kunna separera data så att Skatteverket med flera olika verksamheter ska kunna efterleva tillämplig sekretesslagstiftning och dataskyddsreglering.

3.1.2 Exempel på användning av datasjö

Arbetsförmedlingen har ett äldre klassiskt datalager, ett nyare EDW (enterprise datawarehouse) och en datasjö. Uppgifter i datasjön används bland annat för profilering av arbetssökande när de skriver in sig på Arbetsförmedlingen. Verktyget gör en uppskattning utifrån personens egna uppgifter, som de lämnar vid sitt inskrivningstillfälle, tillsammans med kompletterande uppgifter från andra datakällor exempelvis Migrationsverket och SCB. Personen får en kategorisering på hur nära de bedöms vara arbetsmarknaden. Kategoriseringen baseras på statistik som Arbetsförmedlingen tagit fram om vilka parametrar som gör det snabbt eller inte att komma i arbete. Bedömningsstödet hjälper Arbetsförmedlingen att anpassa lämpliga

åtgärder. Det samlade underlaget presenteras för en beslutande handläggare, men stegen dessförinnan sker automatiserat med utgångspunkt i uppgifter som bland annat finns lagrade i Arbetsförmedlingens datasjö.

Det finns inga begränsningar för vilka uppgifter som kan lagras i datasjön eller hur länge de kan lagras där. Däremot finns det tydliga behörighetsbegränsningar som innebär att endast behöriga personer får åtkomst till uppgifterna i en viss zon av datasjön. Rättsavdelningen gör laglighetsbedömningar innan olika uppgifter kan plockas ut ur datasjön. Inför införandet av datasjön gjorde också Arbetsförmedlingen en större laglighetbedömning avseende de generella rättsliga förutsättningarna för att lagra uppgifter i datasjön. Genom att specificera uppgifterna i datasjön i en datakatalog, kan Arbetsförmedlingen genom metadatan hålla reda på vilka uppgifter som finns i datasjön och vilket som är ändamålet med behandingen.

3.2 Federerade lösningar

Ett sätt att genomföra analyser är genom så kallad federerad analys där principen är att själva analysen flyttar till platsen med data istället för att data samlas på platsen för analys. Analysen genomförs på noder där data finns och endast resultatet tillsammans med metadata överförs till en central analysdator.[5]

På liknande sätt är vid så kallad federerad inlärning tanken att träna en maskininlärningsmodell på lokal data och inte behöva lagra data centralt. Den enda information som överförs till den centrala modellen är resultatet av den lokala träningen och de inlärda kunskaperna.[6]

Nedan beskrivs några intiativ med federerade lösningar.

3.2.1 Federerade analyser

Läkemedelsverket har tagit fram två rapporter[7] om federerade analyser. Syftet med rapporterna är bland annat att beskriva och lyfta fram de rättsliga överväganden som uppkommer i förhållande till behandlingar av personuppgifter vid registerforskning med federerad analys. Läkemedelsverket har bland annat försökt att beskriva hur till exempel personuppgiftsansvaret kan anses vara fördelat vid tre olika modeller med federerad analys. Integritetsskyddsmyndigheten har yttrat sig över Läkemedelsverkets rapport och fört fram att de inte i alla delar instämmer i Läkemedelsverkets

[5] Federerade analyser, Rättsliga överväganden, Rapport från Läkemedelsverket, Datum:2021-12-23 Dnr: 4.3.1-2020-017988
[6] Edvin Listo Zec, Olof Mogren, John Martinsson, Leon Ren´e Sütfeld, Daniel Gillblad, Specialized federated learning using a mixture of experts
[7] Federerade analyser, Rättsliga överväganden, Rapport från Läkemedelsverket, Datum:2021-12-23 Dnr: 4.3.1-2020-017988 och Federerade analyser– IT-arkitektur, Rapport från Läkemedelsverket, Datum: 2021-12-03 Dnr: 4.3.1-2020-017988

bedömning.[8] Även om federerad analys tillämpas får uppgifterna endast användas för det ändamål som avses och det faktiska personuppgiftsansvaret måste upprätthållas. Det innebär att den som bestämmer över ändamål och medel är den som är personuppgiftsansvarig även om analysen sker hos den ursprungliga datainnehavaren utan insyn från den som har bestämt för vilket ändamål uppgifterna ska analyseras.

3.2.2 Regulatorisk testverksamhet och federerad inlärning

Integritetsskyddsmyndigheten genomför ett pilotprojekt i form av regulatorisk testverksamhet tillsammans med Sahlgrenska Universitetssjukhuset och Region Halland. Pilotprojektet avser decentraliserad AI med tillämpning av federerad inlärning och drivs med stöd av AI Sweden. Målsättningen är att kunna dela lärdomar och utveckla AI tillsammans utan att behöva dela personuppgifterna. Genom den regulatoriska testverksamheten kan aktörerna testa sina idéer i dialog med Integritetsskyddsmyndigheten och få stöd i rättsliga bedömningar. De idéer som testas bedöms mot befintlig gällande rätt. I början av år 2023 kommer pilotprojektet att utmynna i en publik rapport från Integritetsskyddsmyndigheten.

3.3 På gång inom EU

I detta avsnitt redogörs för några tendenser inom EU. Redogörelsen består av några exempel och gör inte anspråk på att vara uttömmande.

3.3.1 Förslag till AI-förordning

Inom EU förhandlas förslaget till Europaparlamentets och rådets förordning om harmoniserade regler för artificiell intelligens (rättsakt om artificiell intelligens) och om ändring av vissa unionslagstiftningsakter (AI-förordningen).[9] Syftet med AI-förordningen är att harmonisera reglerna för AI inom EU.

I förslag till kommande AI-förordning finns bland annat bestämmelser för regulatoriska sandlådor[10] där utveckling kan ske och personuppgifter behandlas för andra syften än insamlingsändamålet, under vissa kontrollerade former. Det föreslås att ytterligare behandling av personuppgifter ska få ske för utveckling, test och träning av innovativa AI-system.

[8] IMYs Yttrande över Läkemedelsverkets rapport Federerade analyser, DI-2021-4199
[9] Förslag till Europaparlamentets och Rådets förordning om harmoniserade regler för artificiell intelligens (Rättsakt om artificiell intelligens) och om ändring av vissa unionslagstiftningsakter
[10] Regulatoriska sandlådor kan beskrivas som försöksverksamhet, begränsad i tid och rum, med syfte att utveckla teknik och regelverk under trygga former. Se t.ex. Rådets slutsatser om regulatoriska sandlådor och experimentklausuler som verktyg för ett innovationsvänligt, framtidssäkrat och motståndskraftigt regelverk som hanterar omvälvande utmaningar i en digital tidsålder, Bryssel den 16 november 2020, ST 13026/20.

3.3.2 Förslag till förordning om ett europeiskt hälsodataområde

Kommissionen har lagt fram ett förslag till förordning om det europeiska hälsodataområdet.[11] Förslaget till förordning är nu föremål för förhandlingar i EU.

Syftet med den föreslagna förordningen är dels att ge enskilda inom EU en ökad kontroll över sina hälsodata, dels att göra det lättare att dela och få tillgång till olika typer av hälsodata. Detta gäller både det som kallas primär användning (inom vården), exempelvis möjlighet att hämta ut e-recept över landsgränser, och för det som kallas sekundär användning och som omfattar bland annat forskning, precisionsmedicin, innovation och beslutsfattande.

Den föreslagna strukturen för att möjliggöra att dela och få tillgång till hälsodata bygger på en federerad lösning där hälsodata är tänkt att ligga kvar hos den som ursprungligen innehar data (datainnehavare) och delas först när en dataanvändare efterfrågar och har ett tillstånd för användningen av data för ett särskilt ändamål. Förordningen förutsätter att det inom varje medlemsstat inrättas flera olika organ, bland annat en myndighet för den primära användningen och ett organ med ansvar för tillgång till hälsodata för sekundär användning.

Organ med ansvar för tillgång till hälsodata ska bland annat ansvara för tillståndsgivning för hälsodata och erbjuda en så kallad säker behandlingsmiljö för sekundär användning av hälsodata. Säkra behandlingsmiljöer kan exempelvis användas av forskare som ansöker om att få ta del av hälsodata. Varje datainnehavare ska ha en skyldighet att upprätta en datasetskatalog som de ska tillhandahålla organet med ansvar för tillgång till hälsodata.

Av ovan redovisning framgår att det inte är fråga om en ”datasjö” där all data samlas ostrukturerad. Snarare är syftet att kunna ge tillgång till vissa specifika datamängder för särskilda syften, som är fördefinierade och föregås av en tillståndsprocess.

Förslaget till förordning i sin nuvarande utformning förutsätter att endast pseudonymiserade eller anonymiserade uppgifter ska kunna tillgängliggöras i de säkra behandlingsmiljöerna. Detta har ifrågasatts av ett antal svenska remissinstanser då det i Sverige finns möjlighet att forska på personuppgifter utan någon sådan begränsning.

[11] Förslag till Europaparlamentets och rådets förordning om ett europeiskt hälsodataområde, COM(2022) 197 final

4. Rättsliga förutsättningar för stordataanalyser och datasjöar

4.1 Allmänna rättsliga förutsättningar

Många gånger är de rättsliga frågorna för stordataanalyser och datasjöar desamma som vid annan verksamhetsutveckling. I eSams *Checklista för jurister Introduktion i rättsliga förutsättningar i utvecklingsinsatser*,[12] ges ett stöd i vilka rättsliga frågor som kan behöva ställas i en utvecklingsinsats.

Vid den rättsliga bedömningen kan det vara en fördel att utgå från olika faser i verksamhetsutvecklingen. Det blir då tydligt att flera av rättsområdena aktualiseras i varje fas. Det finns olika sätt att beskriva dessa faser eller områden, men en grov indelning kan vara:

- Behovsbedömning
- Insamling
- Lagring
- Bearbetning
- Tillgängliggörande
- Drift/Avveckling

Nedan redovisas i korthet några av de rättsområden som vanligen aktualiseras utifrån den struktur som används i Checklista för jurister och eSams checklista *Juridik vid användning av AI*.[13] För att underlätta samläsning av dokumenten har samma rubriksättning använts (vilket i detta dokument kan innebära att rubriken inte till fullo återspeglar innehållet). För en mer fullständig genomgång av de rättsområden och den lagstiftning som aktualiseras hänvisas till Checklista för jurister samt för rättsliga frågor i samband med AI-användning till eSams checklista *Juridik vid användning av AI*. För frågor rörande pseudonymisering hänvisas till eSams vägledning *Pseudonymisering av personuppgifter*.[14]

4.1.1 Kompetensområde

Regeringsformens bestämmelser om allas likhet inför lagen, saklighet och opartiskhet måste beaktas[15], liksom legalitetsprincipen[16] och proportionalitetsprincipen.[17] Var och

[12] Checklista för jurister Introduktion i rättsliga förutsättningar i utvecklingsinsatser, version 2.0, eSam juni 2019
[13] ES2022-08 Checklista Juridik vid användning av AI
[14] ES2022-01 Vägledning Pseudonymisering av personuppgifter
[15] 1 kap. 9 § och 2 kap. 6 § regeringsformen (1974:152)
[16] 1 kap. 1 § tredje stycket regeringsformen och 5 § förvaltningslagen (2017:900)
[17] 2 kap. 21 § regeringsformen och 5 § förvaltningslagen

en är dessutom skyddad från det allmänna mot betydande intrång i den personliga integriteten, om det gäller övervakning eller kartläggning av personliga förhållanden och det sker utan samtycke.[18] Vid bedömningen av hur ingripande intrånget i den personliga integriteten är, i samband med insamling, lagring och bearbetning eller utlämnande av uppgifter om enskildas personliga förhållanden, är det enligt förarbetena naturligt att stor vikt läggs vid uppgifternas karaktär. Ju känsligare uppgifterna är, desto mer ingripande måste det allmännas hantering av uppgifterna normalt anses vara.[19]

Myndigheten bör ställa sig frågan om vilket behov användning av stordataanalys och/eller en datasjö ska fylla, det vill säga vilken verksamhetsnytta ska stordataanalysen eller datasjön leda till? Utifrån detta behöver värderas vilka behandlingar myndigheter avser att genomföra och om det finns stöd för detta i myndighetens uppdrag. Kanske måste datamängd och innehåll anpassas för den tilltänkta användningen för att det ska finnas rättsliga förutsättningar.

Att myndigheter inte går utanför sitt uppdrag eller vidtar oproportionerliga åtgärder med tillgänglig data är viktigt för rättssäkerheten och tilliten till den offentliga förvaltningen.

4.1.2 God offentlighetsstruktur

En god ordning behövs för den enskildes möjlighet att ta del av allmänna handlingar och myndighetens förmåga att gallra och bevara allmänna handlingar. Bestämmelserna i bland annat tryckfrihetsförordningen (1949:105), offentlighets- och sekretesslagen (2009:400) (OSL) och arkivlagen (1990:782) måste beaktas.

Vid genomförande av en stordataanalys eller vid lagring i en datasjö behöver beaktas om informationen som ska analyseras eller lagras utgör allmänna handlingar och om uppgifterna är föremål för sekretess. Myndigheten behöver värdera vilka källor som används, egna eller andra aktörers, och utifrån det bedöma om sekretess mellan verksamhetsgrenar eller överföring av sekretess aktualiseras. Det kan i vissa fall vara så att sekretess inte längre föreligger eller att andra sekretessbestämmelser aktualiseras för uppgifterna i detta sammanhang. Det behöver också identifieras om det uppstår nya handlingar (inklusive potentiella handlingar) och om dessa är allmänna. Detta kan vara utmanande i en bearbetningsfas, särskilt om AI används i analysen. Det kan behövas finnas möjlighet till ”taggning” eller vara möjligt att göra sekretessmarkeringar. Uppgiftsmängderna kan också behöva hållas separerade (logiskt eller fysiskt).

[18] 2 kap. 6 § andra stycket regeringsformen (1974:152)
[19] Prop. 2009/10:80 s. 183

Vidare behöver krav på bevarande och gallring säkerställas. Myndigheten behöver fundera över hur handlingar som blivit allmänna omhändertas för arkivering för det fall de ska bevaras. Det kan behövas specifika gallringsbestämmelser i myndighetens tillämpningsbeslut som ger möjlighet till gallring av sådana handlingar som inte ska bevaras. Det kan exempelvis vara handlingar som uppstått i bearbetningsfasen och är av uppenbart ringa betydelse eller av generisk karaktär (RA-FS 2021:6).

En god ordning behövs också för att uppmärksamma eventuella immaterialrättsliga begränsningar i möjligheten att använda reultatet av analysen. Upphovsrättsliga bestämmelser finns bland annat i lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk. Se även avsnitt 4.1.6 om immaterialrättsliga begränsningar i avtal.

Det kan vara önskvärt att resultatet kan delas vidare och då behöver det redan i ett initialt skede skapas förutsättningar för detta. Bestämmelser om vidareutnyttjande och öppna data kan aktualiseras, se lagen (2022:818) om den offentliga sektorns tillgängliggörande av data.

4.1.3 Personuppgifter

Information eller data som lagras i en datajsö eller behandlas i en stordataanalys kan innehålla personuppgifter. Om stordataanalysen eller lagringen i en datasjö innebär att personuppgifter behandlas, aktualiseras bestämmelserna i dataskyddsförordningen[20]. Även brottsdatalagen (2018:1177), dataskyddslagen (2018:218) med tillhörande förordning samt myndighetsspecifika registerförfattningar kan aktualiseras. Dataskyddsprinciperna för behandling av personuppgifter behöver beaktas i alla situationer där personuppgifter behandlas.[21]

Myndigheten måste säkerställa att det finns ett berättigat ändamål och en rättslig grund för behandlingen. Den rättsliga grund som främst aktualiseras för myndigheter i detta sammanhang är att behandlingen är nödvändig för att utföra en uppgift av allmänt intresse eller som ett led i en myndighetsutövning (art 6.1 e dataskyddsförordningen). Det innebär att vid inrättandet av en datasjö eller genomförande av en stordataanalys måste personuppgiftsbehandlingen bedömas utifrån ett nödvändighetskriterium. Det är myndighetens grunduppdrag i materiell rätt, dess instruktion samt ändamålsbestämmelserna i de aktuella författningarna som måste bedömas.

Ändamålsbeskrivningar kan vara mer eller mindre detaljerat formulerade och de behöver inte vara uttömmande. Finalitetsprincipen sätter den yttersta ramen för vilken

[20] Europaparlamentets och rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning)
[21] Artikel 5 dataskyddsförordningen

personuppgiftsbehandling som är tillåten. Finalitetsprincipen innebär att det kan vara tillåtet att behandla personuppgifter för andra ändamål än de som uttryckligen anges i olika författningar, så länge behandlingen inte är oförenlig med insamlingsändamålen. Myndigheten behöver således titta på för vilket ändamål uppgifterna ursprungligen samlades in och bedöma om den analys eller uppföljning som ska följa av stordataanalysen eller lagringen i en datasjö är oförenlig med det ändamålet. En sådan prövning måste dokumenteras och kopplingen till det ursprungliga ändamålet bevaras.

Myndigheten behöver också säkerställa de registrerades rättigheter, såsom information om personuppgiftsbehandlingen och möjlighet till radering och rättelse av felaktiga uppgifter. Det finns krav på dokumentation för att möjliggöra transparens. Det ska även vidtas lämpliga tekniska och organisatoriska åtgärder samt finnas en process för hantering av eventuella personuppgiftsincidenter.[22] Rent praktiskt innebär de tekniska och organisatoriska åtgärderna bland annat att myndigheten måste ta ställning till hur behörighetsstyrning, sökbegränsningar och loggningsfunktionalitet bör utformas.

I lagringsfasen kan det bli fråga om en egen lagringsyta, men det finns även federerade lösningar där data inte flyttas till en särskild lagringsyta för analys. Det kan också bli fråga om att utkontraktera lagringen varvid bland annat överföringsbestämmelser kan aktualiseras, se eSams vägledning *Outsourcing 2.0 En vägledning om sekretess och dataskydd*.[23] Personuppgiftshanteringen behöver anpassas utifrån förutsättningarna i varje enskild lösning. Om AI används för att göra en analys kan också frågor om transparens och förklarbarhet aktualiseras, se eSams checklista *Juridik vid användninga av AI*.

Särskilt om myndigheternas registerförfattningar

Av artikel 6.2 dataskyddsförordningen framgår att medlemsstaterna får behålla eller införa mer specifika bestämmelser för att anpassa tillämpningen av förordningen genom att närmare fastställa specifika krav för uppgiftsbehandlingen och andra åtgärder för att säkerställa en laglig och rättvis behandling.

I många myndigheters registerförfattningar finns reglering kring ändamålsbestämmelser och andra regleringar som begränsar hur personuppgifter får behandlas. I de fall en myndighet har en registerförfattning som är snävt utformad kan det innebära att myndighetens möjligheter att behandla personuppgifter begränsas i förhållande till det uppdrag som myndigheten har erhållit i exempelvis myndighetens instruktion. Det kan därmed skilja sig mellan olika myndigheters möjlighet att behandla personuppgifter i

[22] Artikel 32-34 dataskyddsförordningen
[23] Outsourcing 2.0 En vägledning om sekretess och dataskydd, 2019

stordataanalyser eller i en datasjö beroende på vilka förutsättningar som föreligger i myndighetens registerförfattning.

Skillnad mellan privata och offentliga aktörer
De rättsliga förutsättningarna för att behandla personuppgifter skiljer sig också åt mellan privat och offentlig verksamhet.

En privat aktör kan i större utsträckning använda de rättsliga grunderna *avtal* (artikel 6 1. b dataskyddsförordningen) och *samtycke* (artikel 6 1. a dataskyddsförordningen) för behandling av personuppgifter. Sociala medieplattformar och andra privata kommunkationstjänster grundar vanligtvis sin behandling av personuppgifter på att den enskilde ingår ett avtal med företaget och däri ofta även lämnar sitt samtycke till behandlingen av personuppgifter.

Myndigheter har inte samma möjlighet som en privat aktör att behandla personuppgifter på dessa grunder. Som redovisats ovan är istället den vanligaste grunden för myndigheters personuppgiftsbehandlingar att uppgifterna är nödvändiga för en uppgift av allmänt intresse.

Privata aktörer har därmed i jämförelse med en myndighet större möjligheter att behandla personuppgifter för stordataanalyser och i datasjöar.

4.1.4 Service, tillgänglighet och ärendehandläggning

Bestämmelserna om service och tillgänglighet i bland annat regeringsformen, förvaltningslagen och myndighetsförordningen gäller såväl vid ärendehandläggning som vid annan förvaltningsverksamhet. Det är viktigt för allmänhetens förtroende att de beslut som myndigheterna fattar är riktiga och grundar sig på korrekta beslutsunderlag. Utökade analyser kan möjliggöra en större enhetlighet och rättssäkerhet i myndighetens beslut.

En viktig faktor är dock att processen och grunderna för myndighetens beslut och bedömningar är transparenta. Det kan finnas särskild anledning att se över frågor om transparens när stordataanalys eller lagring i en datasjö resulterar i automatiserade beslut. Det behöver också beaktas att risker för diskriminering kan uppstå i samband med urval, databearbetning och beslutsfattande, se checklista *Juridik vid användning av AI*.

4.1.5 Informationssäkerhet

Informationssäkerhet handlar om att skydda information utifrån dess värde. Värdet bestäms oftast utifrån konsekvens vid utebliven konfidentialitet, riktighet eller

tillgänglighet, så kallad informationsklassning. Utifrån klassningen behöver myndigheten göra en riskanalys. Informationssäkerhet uppnås sedan genom att säkerhetsåtgärder genomförs, exempelvis genom policyer, regler, processer, rutiner, organisatoriska strukturer samt funktioner i program och hårdvara. Detta kan vara särskilt relevant i bearbetningsfasen men även i övriga faser. Det är mycket viktigt att det finns en god dokumentation särskilt vad gäller behörigheter, åtkomst och loggning och att myndigheten upprätthåller en god intern styrning och kontroll. Myndigheten behöver också ha förmåga att uppmärksamma och hantera incidenter.

Utifrån att en datasjö förutsätts bestå av ostrukturerat data är det utmanande för en myndighet att kunna säkerställa informationssäkerhetskraven vid behandling av uppgifter i datasjöar.

Det behöver också bedömas om säkerhetsskyddslagens bestämmelser kan bli tillämpliga.

4.1.6 Upphandling och konkurrensfrågor

För uppgifter som inhämtats eller köpts in till myndigheten så kan det finnas begränsningar i avtal för hur uppgifterna får användas av myndigheten. Exempelvis kan det finnas immaterialrättsliga begränsningar för hur datat kan nyttjas. Det är därmed av vikt att redan i insamlingsfasen bedöma om det finns begränsningar som gör att datat inte är lämpligt att ingå i en datasjö eller stordatanalys. Se även avsnitt 4.1.2.

4.1.7 Drift och förvaltning

De rättsliga förutsättningarna blir inte bara aktuella vid initieringen av en datasjö eller stordataanalys. För att den nya hanteringen ska leva upp till alla rättsliga krav även vid drift och förvaltning krävs ett livscykelperspektiv på uppgiftshanteringen. Det är därför viktigt att det finns en tydlighet och spårbarhet i de överväganden som gjorts i de olika faserna. Myndighetens överväganden behöver dokumenteras och det behöver säkerställas att det finns rutiner framtagna för delar av hanteringen. Det är också viktigt att det är definierat vem som har ansvar för vad (exempelvis personuppgiftsbehandling och gallring) i de olika faserna. Det behöver exempelvis klargöras vem som har rätt att bestämma vilka uppgifter som samlas in och vilka som ska tillfrågas om den insamlade datan ska användas för nya eller ändrade ändamål. Lagring i datasjöar eller genomförande av stordataanalyser kan innebära att ansvarsskyldigheter uppstår inom nya områden.

4.2 Summering rättsliga förutsättningar

Det är lätt att se fördelar med analys av stora datamängder och rent tekniskt är möjligheterna med datasjöar och stordataanalyser många. De rättsliga kraven, bland annat med koppling till it- och informationssäkerhet samt krav på integritetsskydd, är desamma som då information eller personuppgifter behandlas i ett datalager eller används i en verksamhetsanalys. Att genomföra stordataanalyser eller att lagra information och personuppgifter i en datasjö innebär alltså inte förändrade möjligheter eller lättnader juridiskt.

Som framgår av redogörelsen är det flera rättsliga förutsättningar som måste beaktas. Legalitetsperspektivet blir en viktig faktor utifrån att det sannolikt blir fråga om något som myndigheten inte har ett tydligt stöd för i instruktion eller registerlagstiftning. Vid stordatanalys kan exempelvis aggregerad data visa på nya mönster eller ge nya kunskaper som skapar verksamhetsnytta för myndigheten. Eftersom behovet och nyttan av olika analyser kan vara föränderligt är det viktigt att kontinuerligt bedöma om de föreslagna åtgärderna ryms inom myndighetens uppdrag eller om ett särskilt förtydligande behöver inhämtas från uppdragsgivaren.

Personuppgiftsbehandlingen är ett område som särskilt bedöms ge upphov till utmaningar om en myndighet vill genomföra stordataanalyser eller inrätta en datasjö. Många gånger kommer datamängden sannolikt innehålla personuppgifter. Det finns en risk att även om det varit fråga om anonym information kan denna aggreggerad komma att ha sådan koppling till en person att blir fråga om personuppgifter. Att samla data som innehåller personuppgifter kräver att det är förutbestämt för vilka berättigade ändamål och med vilken rättslig grund som personuppgiftsbehandlingen utförs. Ju mer ändamålen ändras desto svårare blir det att säkerställa dataskyddet. Dataskyddsförordningens principer gäller vid all behandling av personuppgifter och måste alltid beaktas av den personuppgiftsansvarige. Det innebär exempelvis att principerna om nödvändighet och proportionalitet ska beaktas, samtidigt som det vid en stordataanalys många gånger inte går att ange vilka uppgifter som är relevanta att behandla förrän analysen är genomförd.

Därtill kan myndighetsspecifik registerförfattning innehålla ytterligare begränsningar vad gäller behandling av personuppgifter, exempelvis vad gäller sökbegränsningar eller hur känsliga personuppgifter får behandlas. Många registerförfattningar har också uppdelningar mellan myndighetens olika databaser, vilket innebär att verksamhetsdata från en del av verksamheten inte får sammanblandas med data från en annan. Det begränsar möjligheten både till stordaraanalys och lagring i datasjö.

En offentlig aktör behöver ha interna förhållningsregler för hur behandlingen av personuppgifter får och ska gå till. Tekniken för att strukturera, hantera och behandla personuppgifter måste vara förenlig med juridiken. Om datat har olika förutsättningar till exempel olika rättsliga grunder eller olika gallringsfrister, krävs ofta mer avancerad teknik som i sin tur kan innebära högre krav på informationssäkerhet och integritetsskydd.

Det finns förslag bland annat på EU-nivå som ser ut att innebära möjligheter för behandling av personuppgifter i större utsträckning, genom så kallade regulatoriska sandlådor. Det kan dock konstateras att dessa förslag kommer med strikta regler om hantering och tillsyn. I några av dessa förslag nämns att pseudonymisering kan användas som en åtgärd för att personuppgifterna ska kunna behandlas. Detta kan med fördel utforskas vidare i eSams arbete med pseudonymisering.

Sammanfattningsvis kräver ett genomförande av en stordataanalys mycket god kontroll av vilka uppgifter som behandlas och för vilka syften, samt förmåga att hantera förvaltningsrättsliga krav. Detsamma gäller vid inrättandet av en datasjö, vilket kan vara en utmaning utifrån att en datasjö är tänkt att också kunna innehålla ostrukturerat material. Gemensamt för flera av de rättsliga förutsättningar som behandlas i denna promemoria är att de förutsätter struktur. Krav på bland annat gallringsfrister, uppgiftsminimering, ändamålsbegränsning, adekvat skydd efter informationsklass och tillgänglighet ställer krav på ordning och kontroll. Det innebär att det behöver finnas en struktur även för det ostrukturerade materialet.

Utifrån de rättsliga förutsättningarna är därför bedömningen att myndigheters möjlighet att inrätta en datasjö som uppfyller regelverken är mycket begränsad. Möjligheten att genomföra stordataanalyser begränsas också av de rättsliga förutsättningarna. Eventuellt skulle en tydligare reglering kring uppdrag och ändring av ändamålsbestämmelser i registerförfattningar kunna skapa möjlighet att genomföra sådana analyser.

eSam är ett medlemsdrivet program för samverkan mellan myndigheter för att underlätta och påskynda digitaliseringen inom det offentliga. eSam bildades 2015 som en frivillig fortsättning på E-delegationen. En viktig uppgift för eSam är att ta fram stöd och vägledningar som ger förutsättningar för att öka den digitala samverkan inom offentlig förvaltning.

Alla stöddokument finns på esamverka.se

I eSam ingår Arbetsförmedlingen, Bolagsverket, Boverket, Centrala Studiestödsnämnden, Domstolsverket, eHälso-myndigheten, Ekonomistyrningsverket, Folkhälsomyndigheten, Försäkringskassan, Havs- och vattenmyndigheten, Inspektionen för vård och omsorg, Jordbruksverket, Kriminalvården, Kronofogdemyndigheten, Lantmäteriet, Länsstyrelserna, Migrationsverket, Naturvårdsverket, Patent- och Registreringsverket, Pensionsmyndigheten, Riksarkivet, Rättsmedicinalverket, Sida, Skatteverket, Skolverket, Statens institutionsstyrelse, Statens servicecenter, Statens tjänstepensionsverk, Statistiska centralbyrån, Tillväxtverket, Trafikverket, Transportstyrelsen, Tullverket och Universitets- och högskolerådet.