

2018-05-24

Delrapport 3 Slutrapport

KARTLÄGGNING AV PERSONUPPGIFTER I KOMPLEXA
INFORMATIONSSYSTEM MED STÖD AV AI - INFÖR
DATASKYDDSFÖRORDNINGEN 2018

PETER MANKENSKIÖLD





Sammanfattning

Rapporten omfattar beskrivning och resultat från Föreningen Sambruks projekt, samt beskrivning av den leveransklara tjänsten, som har varit en del av projektets målsättning.

Utvecklingen av lösningen har genomförts utanför projektets budget och organisation, men baseras på kravställningen från projektet.

Projektet har genomförts med i första hand ekonomiskt stöd från medlemmar i Föreningen Sambruk; Eskilstuna kommun, Gävle kommun, Göteborgs stad, Helsingborgs stad, Nacka kommun, Skellefteå kommun, Sundsvalls kommun, Täby kommun, Västerås stad samt Västra Götalandsregionen - VGR.

De inledande aktiviteterna i etapp 1 och 2a har omfattat utvärdering av tekniskt genomförbara krav samt juridiska förutsättningar, i första hand kopplat till nya [Allmänna Dataskyddsförordningen \(GDPR- General Data Protection Regulation\)](#), vars syfte är att skydda individen och dennes uppgifter, som idag behandlas och lagras i ett otal former och i en stor mängd informationssystem. Att kartlägga detta är ett mycket stort och komplext arbete, som för alla organisationer innebär hundratals, t o m tusentals, arbetstimmar (både före och efter den 25 maj 2018 då förordningen träder ikraft).

Den sista etappen, 2b, har hanterat utvärdering av förslag på paketerad lösning och hantering av licensmodeller, teknisk paketering och prismodeller, i första hand för kommuner.

Projektets sista etapp är, vid tidpunkten för redovisning av denna rapport, inte fullt finansierad.

Milstolpeplan

Etapp 1

2017-09-25 Projektstart (Rev till 2017-10-05)
2017-10-27 Leverans POT (Rev till 2017-11-28)
2017-11-02 Beslut om etapp 2a (Rev till 2017-11-28)
2017-11-30 Delrapport 1 – etapp 1

Etapp 2a

2017-12-22 Beslut om etapp 2b (Rev till 2018-02-16)
2018-02-16 Delrapport 2 – etapp 2a

Etapp 2b

2018-03-02 Delrapport 3 – avstämning etapp 2b
2017-12-22 Slutrapport (Rev till 2018-05-24)



Slutsatser och resultat från förstudien:

- ✓ Befintliga verktyg på marknaden kan analysera omfattningen av lagrade personuppgifter, dock saknas en lösning som hanterar utmaningen att granska och strukturera data enligt Dataskyddsförordningen
- ✓ Det är praktiskt omöjligt att manuellt hantera analys av den extrem stora mängden data – en teoretisk uppskattning indikerar 12 månår per Terabyte (TB) data (vilket motsvara mindre än en tiondel av vad som finns i en medelstor kommun)
- ✓ Ostrukturerad data är ett område som hittills inte adresserats och saknar tydlig systemägare
- ✓ Licensmodeller för befintliga system på marknaden är ofta baserade på volym eller antal sökningar, vilket försvårar både anskaffning och hantering av systemen

Sammanfattning av kravställning från projektet

- ✓ Systemet ska kunna hantera olika former av ostrukturerad data
- ✓ Stöd för att kunna koppla genomlysningssystemet till respektive identifierad datakälla
- ✓ Arbetsflöde för inventering
 - Stöd för angivande av förekomster av personuppgifter
 - Stöd för angivande av laglig grund
 - Koppling till externa källor för stöd i beslut
 - Stöd för dokumentation av juridiska beslut (Registerförteckning)
 - Stöd för angivande av icke tidigare identifierade personuppgifter
- ✓ Säkerställande att ingen data lämnar den lokala installationen i en organisations IT-miljö
 - Lokalt kognitivt lärande och annotationsteknik
 - Central AI-motor som omvandlar kognitiva förmågor till nya sökalgoritmer
- ✓ Ett stöd för att **lösa** problemet – inte bara visa problemets omfattning

Ett separat utvecklingsprojekt, för att realisera en komplett, levererbar tjänst har initierats som en spin-off till utrednings- och utvärderingsprojektet, enligt Föreningen Sambruks ursprungliga projektspecifikation.

Utvecklingsarbetet har alltså skett utanför förstudieprojektet och har finansierats och genomförts av företagen Aigine AB, IBM Svenska AB och Elinar OY Ltd (Finland).

Avstämning mellan utvecklingsgruppen och projektets kravställning har skett kontinuerligt.

*En mera utförlig beskrivning av lösningen finns i **Bilaga** till denna rapport. Utvecklingsteamet bakom tjänsten har också tagit fram en separat produktbulletin. Delar av texten i förstudierapportens kapitel 2 och 3 har hämtats från denna produktbulletin; [Aigine produktbulletin](#)*

Innehåll

Sammanfattning.....	2
Projektet.....	2
Milstolpeplan.....	2
Slutsatser och resultat från förstudien konstaterar att:	3
Sammanfattning av kravställning från projektet.....	3
Ett separat utvecklingsprojekt.....	3
1 GDPR-Projektet.....	8
1.1 Bakgrund till pilotprojektet	8
2 GDPR / Dataskyddsförordningen.....	8
2.1 Om PUL.....	8
2.2 Utökad definition av personuppgift	8
2.3 Utökad definition av datakälla	9
2.4 Missbruksregeln försvinner	9
2.5 Hantering av personuppgifter	9
2.6 Registerförteckning	9
2.7 Lagliga grunder	10
2.7.1 Samtycke	10
2.7.2 Avtal.....	10
2.7.3 Rättslig förpliktelse.....	10
2.7.4 Skydd för grundläggande intresse.....	10
2.7.5 Uppgift av allmänt intresse och myndighetsutövning	10
2.7.6 Efter en intresseavvägning	10
2.8 Eventuella konsekvenser	10
2.9 Viten	10
2.10 Skadestånd	11
2.11 Bad will	11
3 Utmaningen.....	11
3.1 Vad behöver göras med ostrukturerade data.....	11
3.1.1 Läs igenom allt.....	11
3.1.2 Avgöra om det finns personuppgifter	12
3.1.3 Hitta laglig grund	12
3.1.4 Dokumentera.....	12
4 Projektets syfte.....	12

5	Projektets mål	12
6	Projektets förväntade nytta/resultat	12
6.1	Projektplan (ursprunglig).....	12
AP001	Projektledning/-koordinering	12
AP002	Etablering av labbmiljö/test.....	13
AP003	Persondatautsökning & verifiering	13
AP004	Etablering av labbmiljö/prod	13
AP005	Persondatautsökning & analys	13
AP006	Konkretisering av konceptet som erbjudande.....	13
AP007	Avrapportering & presentation av beslutsunderlag	13
AP008	Informationsspridning och marknadsföring	13
6.2	Reviderad tidplan och projektleverans	13
7	Genomförda aktiviteter.....	14
7.1	Etapp 1.....	14
7.2	Etapp 2a.....	14
7.3	Etapp 2b.....	14
7.4	Iakttagelser och behandlade frågeställningar.....	15
8	Finansiering och medverkande kommuner.....	16
8.1	Finansieringsbehov för projektet	16
8.2	Redovisning av kostnader.....	16
8.3	Projektgrupp och medfinansiärer	17
8.4	Styrgrupp	17
9	En lösning på problemet, inte visa problemets omfattning.....	17
9.1	Saknas lösning på marknaden	17
9.2	Saknar stöd för den dagliga verksamheten.....	17
9.3	Strukturerad respektive ostrukturerad data	17
9.4	Ingen tydlig systemleverantör	18
9.5	Tar för mycket tid och kräver juridisk spetskompetens.....	18
9.6	Befintliga licensmodeller	18
10	Kravställning för önskad lösning.....	18
10.1	Ostrukturerad data.....	18
10.2	Strukturerad data (omfattas ej av projektets förslag på lösning)	18
10.3	Identifiering av datakällor	18
10.4	Arbetsflöde för inventering.....	19
10.5	Angivande av metadata.....	19

10.6 Angivande av laglig grund	19
10.7 Koppling till externa källor för stöd i beslut	19
10.8 Minskat behov av juridisk spetskompetens	19
10.9 Icke tidigare identifierade personuppgifter	19
10.10 Kognitivt lärande (lokalt installerad programvara)	19
10.11 AI motor (centralt placerad programvara, i "molnet")	19
10.12 Användargenererat innehåll.....	19
10.13 Ingen data får lämna den lokala installationen i en organisations IT-miljö	20
10.14 Ett stöd för att lösa problemet.....	20
10.15 Processpaketering (del av)	20

Bilaga - beskrivning av lösningen

Inledning.....	1
1 Sammanfattning	1
2 Funktions- och teknisk beskrivning av Aigine.....	2
2.1 Funktionsbeskrivning.....	2
2.2 Installation och konfigurerings	2
2.3 Installation	3
2.4 Koppla datakällor.....	3
2.5 Filtrera	3
2.6 Gemensamma definitioner från centralt driftad AI-motor (i "molnet")	3
2.7 Utfiltrering av dokument som innehåller presumtiva personuppgifter.....	4
2.8 Tidsbesparing	4
2.9 Granska.....	4
2.10 Workflow	4
2.11 Grafiskt användargränssnitt	5
2.11.1 Hittade personuppgifter visas markerade.....	5
2.11.2 Inbyggt stöd för att bekräfta eller avvisa förslag på personuppgift.....	5
2.11.3 Tidsbesparing.....	5
2.12 Avgöra.....	6
2.13 Kontextuell databas.....	6
2.14 Tidsbesparing och möjlighet att sprida på fler resurser.....	6
2.15 Dokumentera.....	7
2.15.1 Inbyggt smart digitalt formulär för direkt dokumentation	7
2.15.2 Förslag på laglig grund från AI	7
2.15.3 Tidsbesparing och möjlighet att sprida på fler resurser.....	7

2.16 Kognitiva lärandet	7
2.17 Annoteringskodning lokalt	8
2.17.1 Exempel på annoteringskod innan Aigines anonymisering.....	8
2.17.2 Exempel på annoteringskod efter Aigines anonymisering.....	9
2.18 Kognitivt lärande i molnet	9
2.19 Kognitiva förmågor förbättras.....	9
2.20 Alla får del av de förbättrade förmågorna	9
2.21 Tidsbesparing	9
2.22 Exempel på arbetsinsats utan stöd	10
2.23 Exempel på potentiell effektivisering med stöd.....	10
3 Teknisk beskrivning	10
3.1 Licensbegränsning	10
3.2 Ingående komponenter	11
3.2.1 Aigine Unstructured Data Inventory Engine.....	11
3.2.2 Användargränssnitt	11
3.2.3 Workflowstöd	11
3.2.4 Metadatalager	11
3.2.5 StoredIQ.....	11
3.2.6 Watson AI och Watson NLU	11
3.2.7 Kognitivt Lärande.....	11
3.2.8 Elinar AI miner	12
4 Systemskiss.....	12
4.1 Kryptering och certifikathantering	12
4.2 Stöd för webbläsare	12
4.3 Stöd för filformat och protokoll	13
4.4 Uppdateringar	13
5 Leverans av tjänsten.....	13
5.1 Leverans.....	13
5.2 Prismatris för Kommuner	13



Projektet och utmaningen

Bakgrund, syfte och mål

1 GDPR-Projektet

1.1 Bakgrund till pilotprojektet

[Föreningen Sambruk](#), som är en medlemsorganisation med c:a 100 kommuner och organisationer, har uppmärksammat utmaningen med den stora arbetsbelastning för medlemmarna som den nya dataskyddsförordningen medför. Särskilt kopplat till att inventera organisationernas alla datamängder som idag behandlas och lagras i ett otal former och i en stor mängd informationssystem. Att kartlägga detta är ett mycket stort och komplext arbete, som för alla organisationer innebär en icke hanterbar mängd arbetstimmar, för att granska och bedöma laglig grund för hantering av personuppgifter. Därför har föreningen Sambruk, tillsammans med 10 medlemskommuner och ett regionförbund samt SKL, genomfört ett pilotprojekt med målsättningen att ta fram underlag och kravställning för en lösning som kan medföra väsentliga effektiviseringsvinster i det omfattande kartläggningsarbetet, som ingår i förberedelser för att säkra organisationers efterlevnad av den nya dataskyddsförordningens krav.

2 GDPR / Dataskyddsförordningen

Från och med den 25 maj 2018 gäller [GDPR](#), General Data Protection Regulation, på svenska Data-skyddsförordningen. Förordningen är direkt applicerbar i samtliga EUs medlemsländer, och i Sverige har den kompletterats av en ny [Dataskyddslag](#) som kommer ersätta PUL.

Syftet med GDPR-förordningen är att skydda den personliga integriteten och personliga data för varje medborgare i ett EU-land. GDPR kommer att påverka och ha direkt effekt på samtliga organisationer inom EU samt alla juridiska personer som på något sätt hanterar personliga data gällande en EU-medborgare, oavsett var dessa befinner sig i världen.

GDPR kan därför anses vara en global lagstiftning.

2.1 Om PUL

I Sverige ersätter Dataskyddsförordningen och Dataskyddslagen den tidigare Personuppgifts-lagen, PUL, som gällt sedan 1998.

De förändringar och utmaningar som uppmärksammas i Sverige med anledning av GDPR är ofta i förhållande till PUL. Samtidigt så saknade PUL tydliga sanktioner, vilket har gjort att många organisationers efterföljande av PUL kan misstänkas vara eftersatt.

För många organisationer kräver därför GDPR inte bara att man lyfter sig från PUL, utan i många fall att man måste ta ett helhetsgrepp på sådant som tidigare varit eftersatt.

2.2 Utökad definition av personuppgift

GDPR förändrar definitionen av personuppgift. Där PUL tidigare definierat personuppgift som " All slags information som direkt eller indirekt kan hänföras till en fysisk person som är i livet räknas enligt personuppgiftslagen som personuppgifter." utökar GDPR detta med att även inkludera uppgifter som



inte i sig själva räcker för identifiering av en individ, men som kan göra det i kombination med andra uppgifter.

"Med personuppgifter avses varje upplysning som avser en identifierad eller identifierbar fysisk person. Avgörande är att uppgiften, enskilt eller i kombination med andra uppgifter, kan knytas till en levande person."

Vad som därför anses utgöra en personuppgift enligt GDPR blir därför en fråga av vilken övrig information som organisationen innehar, eftersom det är den teoretiska kombinationen som avgör huruvida det är en personuppgift eller inte.

2.3 Utökad definition av datakälla

Där PUL hade sin starkaste reglering på det man kallar register, vilket innebär databas, så är GDPR teknikneutral. GDPR gäller därför alla samlingar med information som kan sökas. Detta inkluderar givetvis databaser, men även analoga arkivskåp där informationen är sorterad på ett sådant sätt som möjliggör sökning.

2.4 Missbruksregeln försvinner

Anledningen till att PUL i princip inte reglerade andra källor än databaser var den så kallade Missbruksregeln. Denna innebar, i korthet, att man får hantera personuppgifter, så länge den enskilde inte kränks.

Det innebar att många av PUL:s hanteringsregler inte behövde tillämpas på viss hantering av personuppgifter i ett ostrukturerat material. Syftet var att underlätta vardaglig hantering av personuppgifter som inte medförde integritetsrisker. Lättnaderna gällde för vardaglig ostrukturerad behandling av personuppgifter som löpande text i ordbehandlingssystem, löpande text på internet, ljud- och bildupptagningar och korrespondens per e-post. Undantaget kunde också omfatta enkla strukturer som klasslistor och listor över anställda.

För att avgöra om en behandling är kränkande måste man göra en samlad bedömning av hur känsliga uppgifterna är, i vilket sammanhang de förekommer, för vilket syfte de behandlas, vilken spridning de har fått eller riskerar att få, samt vad behandlingen kan leda till. Man får således göra en avvägning i det enskilda fallet där den registrerades intresse av en fredad, privat sfär vägs mot andra motstående intressen.

Med GDPR [försvinner missbruksregeln](#), vilket innebär att all data, strukturerad som ostrukturerad, digital eller analog, måste behandlas enligt samma strikta regelverk.

2.5 Hantering av personuppgifter

GDPR innehåller inget förbud för hantering av personuppgifter. Istället ställer det specifika krav på hur och när personuppgifter får hanteras. Förenklat kräver man att:

1. Man vet var personuppgifter finns
2. Man har säkrat tillgången till personuppgifterna genom behörigheter och liknande
3. Man vet vilken laglig grund som gör att den specifika personuppgiften får sparas och hanteras

Uppfyller man dessa huvudregler finns inga hinder att spara och hantera personuppgifter. GDPR ställer dock kravet att detta skall vara utrett och dokumenterat redan den 25/5 2018.

2.6 Registerförteckning

Det är alltså inte möjligt att "snubbla" över en personuppgift, och i efterhand utreda huruvida man hade rätt att inneha den eller inte. Av denna anledning finns det i GDPR en skyldighet att upprätta en registerförteckning av sin behandling av personuppgifter. Dessa register ska upprättas skriftligen, vara tillgängliga i elektronisk format och hållas uppdaterade. På begäran ska registret göras tillgängligt för

Datainspektionen. Vad som ska finnas med i förteckningen beskrivs i artikel 30 i dataskyddsförordningen.

- Namn och kontaktuppgifter för den personuppgiftsansvarige, den personuppgifts-ansvariges företrädare samt dataskyddsombudet.
- Ändamålen med behandlingen.
- En beskrivning av kategorierna av registrerade och kategorierna av personuppgifter.
- De kategorier av mottagare till vilka personuppgifterna har lämnats eller ska lämnas ut.
- I tillämpliga fall, överföringar av personuppgifter till ett tredjeland eller en internationell organisation.
- De förutsedda tidsfristerna för radering av de olika kategorierna av uppgifter.
- En allmän beskrivning av tekniska och organisatoriska säkerhetsåtgärder.

2.7 Lagliga grunder

Ändamålen med behandlingen är av stor betydelse, då behandling endast får ske då ändamålet kan hänföras till någon av de lagliga grunderna som framgår av artikel 6 i GDPR. Dessa ändamål, och lagliga grunder, skall framgå av registerförteckningen.

2.7.1 Samtycke

Den registrerade har lämnat sitt samtycke till att dennes personuppgifter behandlas för ett eller flera specifika ändamål.

2.7.2 Avtal

Behandlingen är nödvändig för att fullgöra ett avtal i vilket den registrerade är part eller för att vidta åtgärder på begäran av den registrerade innan ett sådant avtal ingås.

2.7.3 Rättslig förpliktelse

Behandlingen är nödvändig för att fullgöra en rättslig förpliktelse som åvilar den personuppgiftsansvarige.

2.7.4 Skydd för grundläggande intresse

Behandlingen är nödvändig för att skydda intressen som är av grundläggande betydelse för den registrerade eller för en annan fysisk person.

2.7.5 Uppgift av allmänt intresse och myndighetsutövning

Behandlingen är nödvändig för att utföra en uppgift av allmänt intresse eller som ett led i den personuppgiftsansvariges myndighetsutövning.

2.7.6 Efter en intresseavvägning

Behandlingen är nödvändig för ändamål som rör den personuppgiftsansvariges eller en tredje parts berättigade intressen, om inte den registrerades intressen eller grundläggande rättigheter och friheter väger tyngre och kräver skydd av personuppgifter, särskilt när den registrerade är ett barn.

2.8 Eventuella konsekvenser

GDPR är en rättighetslagstiftning framtagen med syfte att skydda EU-medborgares personliga integritet och personliga data. Från lagstiftarens sida har en medvetenhet funnits om att efterlevnad av tidigare lagstiftning varit långt ifrån bra. I GDPR framgår därför tydligt att brott mot förordningens regleringar skall leda till "avskräckande" konsekvenser.

2.9 Viten

Privata företag och organisationer riskerar vite om upp till 20 Miljoner Euro eller 4 % av årsomsättningen på koncernnivå per överträdelse. Då detta handlar om vite, dvs. böter, så kan bot

skrivas ut av myndighet. Det är sedan upp till den bötfällde att föra process för att undkomma betalningsansvar.

För offentliga myndigheter i Sverige kommer vitesbeloppen vara 5 Mkr för en mindre överträdelse och 10 Mkr för en större överträdelse. Dessa lägre vitesbelopp motiveras av regeringen med att myndigheter inte kan skaffa sig konkurrensfördelar genom att bryta mot GDPR, samtidigt som 10Mkr är det hittills högsta vitesbeloppet mot offentlig sektor. Det är det maximala vitesbeloppet som kan utdelas för brott mot lagen om offentlig upphandling.

2.10 Skadestånd

I GDPR finns en rätt för skadedrabbad att få skadestånd. GDPR ger uttryckligen den drabbade rätt till ersättning för både direkt och indirekt skada. Med indirekt skada menas här sveda och värk, vilket inkluderar även psykiskt lidande. Brott mot GDPR kan därför leda till mycket stora skadeståndsanspråk från skadedrabbade, och här har lagstiftaren inte gjort skillnad på privata och offentliga aktörer.

2.11 Bad will

GDPR är, som nämnts, en lagstiftning för att säkra den personliga integriteten. Med tid görs därför bedömningen att förlorat good will, eller i värsta fall bad will, kommer utgöra de största kännbara effekterna av lagstiftningen. Seriösa organisationer tar personlig integritet på allvar, oavsett om det handlar om kunder eller om medborgare.

De långtgående effekterna av bad will på organisationer, som missar dessa frågor, har redan synliggjorts, genom inte minst Transportstyrelse-skandalen och kongressförhöret av Facebooks grundare och VD Mark Zuckerberg och allt tyder på att dessa negativa effekter kommer öka, särskilt med koppling till viten.

3 Utmaningen

3.1 Vad behöver göras med ostrukturerade data

Som redan konstaterats så har ostrukturerad data tidigare varit oreglerad i PUL genom missbruksregeln. Då denna regel nu försvinner måste varje organisation säkra sin ostrukturerade data med avseende på det nya regelverket i GDPR. Som organisation måste man utgå ifrån att det kan dölja sig personuppgifter i varje dokument och i varje datakälla.

Möjligheten att helt sonika radera sin ostrukturerade data är direkt olämpligt, dels eftersom den kan tänkas behövas i verksamheten, men framförallt för att de rättsliga förpliktelserna kan göra det direkt olagligt att radera data.

Den strukturerade datan, där varje databas har ett tydligt syfte, behöver bara beskrivas på en aggregeringsnivå av ingående tabeller. För den ostrukturerade datan kan varje dokument utgöra en egen datakälla, såtillvida att den lagliga grunden och behovet av lagring är olika för varje dokument, även om de råkar ligga i samma fillagringsyta, eller i samma katalog.

Som ett led i följsamhet mot GDPR bör man därför, i en förvaltningsfas, bestämma hur personuppgifter får lagras på olika ytor. Genom en sådan struktur behöver registerförteckning endast upprättas på en metanivå - under förutsättning att dessa fillagringtytor kontinuerligt bevakas så att inte felaktiga uppgifter sparas på dem.

3.1.1 Läs igenom allt

För att säkerställa om och var personuppgifter finns i mängden ostrukturerad data, måste samtliga dokument och e-postmeddelanden läsas igenom.

3.1.2 Avgöra om det finns personuppgifter

Vid genomläsningen måste en bedömning göras om dokumentet innehåller personuppgifter i förhållande till GDPR:s nya definition, där uppgifter i kombination med andra uppgifter, kan utgöra en uppgift som möjliggör identifiering av en enskild.

Då långt ifrån alla dokument de facto innehåller personuppgifter innebär detta att en stor del av granskningen kommer ske av dokument som inte behöver granskas.

3.1.3 Hitta laglig grund

I det fall dokumentet innehåller personuppgifter krävs en kvalificerad juridisk bedömning om det finns laglig grund att spara och hantera denna uppgift. Denna bedömning sker i förhållande till både GDPR och andra lagrum som reglerar verksamheten. Över tid kommer denna juridiska bedömning även inkludera bedömning av rättspraxis från Sverige och andra europeiska länder.

3.1.4 Dokumentera

I det fall den juridiska bedömningen ger vid hand att laglig grund saknas för att spara och hantera uppgifterna, måste dokumentet hanteras för maskning av uppgifterna, eller utplåning av hela dokumentet. Denna hantering måste säkerställas, eftersom sparande av personuppgiften i det fall det bedömts sakna laglig grund utgör en allvarig incident och därför innebär en direktkvalificering för de högre vitesbeloppen i GDPR.

I det fall laglig grund hittas måste denna anges, tillsammans med information om hur länge uppgiften ska sparas. I det fall den lagliga grunden utgörs av avtal eller samtycke bör en hänvisning finnas till vart samtycke eller avtal finns.

Denna dokumentation utgör registerförteckning, vilket är ett krav enligt GDPR och innan en fungerande lagringsstruktur och policy etablerats i organisationen, riskerar varje dokument som innehåller personuppgifter utgöra ett eget "register" med avseende på laglig grund, ändamål och gallringstid.

4 Projektets syfte

Pilotprojektets syfte har varit att verifiera konceptet och nyttoeffekter, samt att sätta upp ett långsiktigt hållbart koncept som kan erbjudas Föreningen Sambruks medlemmar (och även andra intressenter), för att kunna genomföra kartläggningsarbetet enligt föregående avsnitt.

5 Projektets mål

Målsättningen för projektet har varit att erbjuda en tjänst som medför väsentliga effektiviseringsvinster i det omfattande kartläggningsarbetet, som ingår i förberedelser för att säkra organisationens (och IT-miljöns) uppfyllelse av den nya Dataskyddsförordningens krav.

6 Projektets förväntade nytta/resultat

Den paketerade tjänsten ska tas fram med syfte att kunna användas av alla aktörer inom offentlig sektor (Statliga- och kommunala myndigheter och bolag samt landsting och regioner) med avsikt att effektivisera arbetet med inventering av datamängder för utsökning av misstänkt förekomst av persondata i komplexa digitaliserade miljöer omfattande såväl strukturerad som ostrukturerad information.

6.1 Projektplan (ursprunglig)

AP001 Projektledning/-koordinering

- Planering



- Koordinering
- Dokumentation
- Detaljerad projektplan

AP002 Etablering av labbmiljö/test

- Etablera teknisk plattform

AP003 Persondatautsökning & verifiering

- Identifiera och definiera berörda parametrar
- Identifiera och definiera berörda datamängder
- Genomförande av tester och utvärdering
- Kartläggning- och jämförelse av processer
- Teknik- och konceptverifiering
- Dokumentation och presentation

AP004 Etablering av labbmiljö/prod

- Identifiera konkreta aktörer/intressenter ("stake holders")
- etablering av expertteam

AP005 Persondatautsökning & analys

- Identifiera och definiera berörda standarder
- Identifiera och definiera berörda datamängder
- Juridiska bedömningar
- Genomförande av tester och utvärdering

AP006 Konkretisering av konceptet som erbjudande

- Paketering och produktifiering
- Framtagande av affärs- och finansieringsmodell
- Nyttjandeprinciper

AP007 Avrapportering & presentation av beslutsunderlag

- Sammanställning av dokumentation
- Avrapportering

AP008 Informationsspridning och marknadsföring

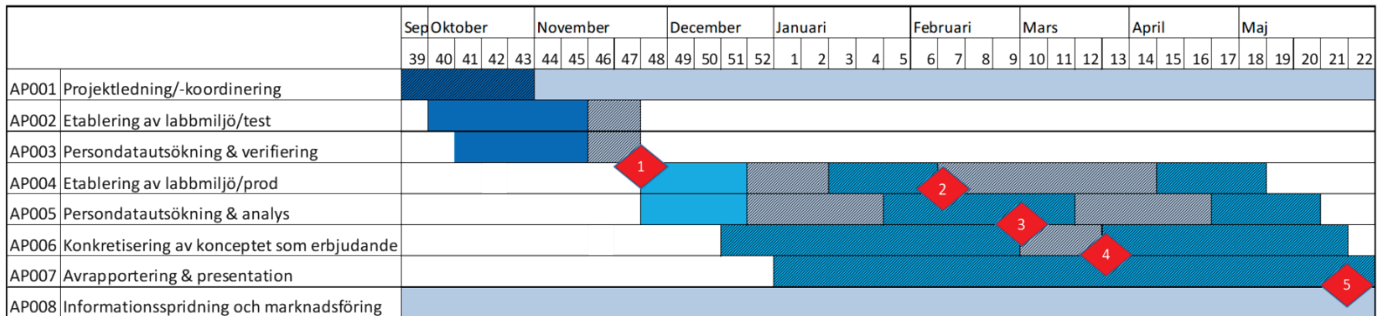
- Fortlöpande infospridning i Sambruks kanaler
- Intern info deltagande organisationer/kommuner
- deltagande i externa konferenser/seminarier

6.2 Reviderad tidplan och projektleverans

Projektet har delats upp i tre etapper som alla strävar till att ta fram ett kravunderlag för en tjänst som kan avropas genom Föreningen Sambruk. Vid varje etapps avslut har en eller flera beslutspunkter tagits fram för styrgruppen att ta ställning till. Efter beslut om fortsättning med nästa etapp har befintliga och nya medfinansierare bjudits in för att bidra med medel för att täcka de kostnader som hör till respektive etapp.

De styrgruppsmöten som genomförts för etapp 1, 2a och 2b, med tillhörande beslutspunkter, har kunnat genomföras med viss förskjutning i tid. Det är en av anledningarna till att hela tidplanen, och med det även slutleveransen, har flyttats framåt i tid i förhållande till den ursprungliga tidplanen. De

planerade aktiviteterna och leveranserna för respektive etapp har dock behållit sin ursprungliga omfattning och formulering. Styrgruppen har även tydliggjort behovet av att ha den leveransklara tjänsten tillgänglig vid projektets avslut.



7 Genomförda aktiviteter

7.1 Etapp 1

Arbetet inom etapp 1 var fokuserat till tekniska lösningar och juridiska förutsättningar samt finansiering av etapp 1 och 2a.

Tester har genomförts med delkomponenter av föreslagen arkitektur för lösning på verksamhetsdata hos en av de deltagande kommunerna.

Juridikreferensgrupp och teknikreferensgrupp har även etablerats för avdömning av juridiska och tekniska frågor som har uppstått under projektets gång.

7.2 Etapp 2a

Inom etapp 2a har arbetet fokuserat på juridiska frågeställningar kopplade till olika definitioner av personuppgifter. Utmaningen är att översätta definitionen i lagtext till exempel på data som kan utgöra personuppgift för att, i nästa steg, kunna översätta dessa exempel till maskinkod som kan användas i samband med genomlysning av stora och komplexa datamängder.

Vi har i de juridiska och tekniska referensgrupperna gått igenom och rätt ut ett antal frågeställningar.

7.3 Etapp 2b

- **Tjänstepaketering**
 - Utredning och fastställande av ingående tekniska komponenter som behövs för att uppnå arbetsflöde och funktioner enligt beskriven krav- och processbeskrivningar
 - Utredning av omfattning och fastställande av vilken part som står för utveckling av gränssnitt mellan de tekniska komponenterna och den mänskliga interaktionen
- **Affärsmodell**
 - Utredning och fastställande av licenskostnader för de delar av tekniska komponenter som ingår enligt beskriven dataarkitektur
 - Utredning och fastställande av prissättningsmodell som tar hänsyn till variabler kopplade till
 - Storlek på kommun (antal anställda)

- Medlemskap i Föreningen Sambruk
Medlemmar i föreningen Sambruk har möjlighet att utnyttja det s k Teckalundantaget i LOU, för avrop av tjänsten
- Deltagare som medfinansiär i projektet (om möjligt)
 - Utredning och förslag på prismodeller och leveransvillkor för organisationer som saknar medlemskap i Föreningen Sambruk
- **Go to market**
 - Utredning och fastställande av leverantörer som kan leverera tjänsten
 - Utredning av omfattning för förvaltningsåtaganden
- **Konkretisering av konceptet som erbjudande** (Arbetspaket 6)
 - Paketering och produktifiering
 - Framtagande av affärs- och finansieringsmodell
 - Definition av nyttjandeprinciper
- **Avrapportering & presentation av beslutsunderlag** (Arbetspaket 7)
 - Sammanställning av dokumentation
 - Förslag på förvaltning av tjänsten
 - Avrapportering
 - Resultat
 - Förslag till slutrapport

7.4 *lakttagelser och behandlade frågeställningar*

Det är en stor utmaning att översätta en definition från en lagtext (exempelvis data som kan utgöra personuppgift) till dagligt tal, för att, i nästa steg, översätta dessa exempel till maskinkod som kan användas i samband med utsökning och genomlysning av stora och komplexa datamängder. Därför är det lokala kognitiva lärandet för vidare hantering av en centralt placerad AI-motor för utveckling av bättre sökalgoritmer en mycket viktig delkomponent i den färdiga lösningen.

Vi har i de juridiska och tekniska referensgrupperna gått igenom och rätt ut ett antal frågeställningar.

Behandlade frågeställningar inom juristreferensgruppen

- Hantering och definition av personuppgift
- Hantering, definition och beskrivning av process för laga stöd
- Hantering och beskrivning av process för metadata (data om data)
- Möte med jurister på SKL:
 - Hantering av metadataindex/registerförteckning. (*juridiskt perspektiv*)
 - Hantering av fulltextindex (*juridiskt perspektiv*)
 - Definitioner och beskrivningar av personuppgifter
 - Mallar
 - Laga stöd
 - Checklistor
 - Datakällor
 - Ostrukturerad data
 - Strukturerad data

Behandlade frågeställningar inom teknikreferensgruppen

- Metadata
- Maskinkod/RegEx vs Annotationsteknik
- Crawlers vs molntjänster

- Ständigt uppdaterade gemensamma definitioner
- Mönsterigenkänning, kognitivt lärande och artificiell intelligens
- Teknisk paketering

Produkter medverkande i lösningsförslaget

- IBM® Information Governance Catalog
- IBM® StoredIQ Suite
- IBM® InfoSphere® Information Analyzer
- IBM® Watson Explorer
 - IBM® Watson Knowledge Studio
- IBM® Enterprise content management

8 Finansiering och medverkande kommuner

8.1 Finansieringsbehov för projektet

Varje etapp har budgeterades till 300 000 kr per etapp.

Totalbudgeten för projektet beräknades till 900 000 kr.

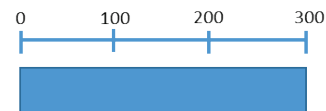
Etapp 2b är inte fullt finansierad vid tidpunkten för redovisning av slutrapporten.

➤ Total projektbudget 900.000 kr

- Samtliga intressenter svarar för egen arbetstid & -kostnader
- Konsulters resor och ev lokalkostnader tillkommer

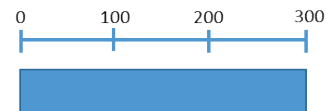
➤ Etapp 1: 300.000 kr

- IBM etablerar en första testmiljö – Watson
- Arbetsinsatser enl 'Etapp 1'



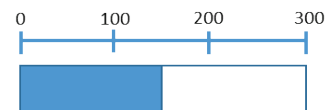
➤ Etapp 2a: 300 000 kr

- IBM etablerar produktionsmiljön – Watson
- Referensgrupper (juridik, teknik) knyts till projektet
- Definitioner och kodning enl 'Etapp 2a' (AP 4, 5 + del 8)



➤ Etapp 2b: 300 000 kr

- Paketering
- Affärsmodell
- Aktiviteter enl 'Etapp 2b' (AP 6, 7 + del 8)



8.2 Redovisning av kostnader

Kostnader har fördelats på följande kostnadsslag:

- Externa konsulter och projektledare: 450 000 kr
- Extern juridisk expertis 300 000 kr
- Resor och administrativa kostnader 150 000 kr

8.3 Projektgrupp och medfinansiärer

Deltagande organisationer från Föreningen Sambruks medlemmar som deltagit i och medfinansierat projektet har varit Eskilstuna kommun, Gävle kommun, Göteborgs stad, Helsingborgs stad, Nacka kommun, Skellefteå kommun, Sundsvalls kommun, Täby kommun, Västerås stad samt Västra Götalandsregionen, VGR.

Vid tidpunkten för rapportens redovisning saknas medfinansiering för stor del a etapp 2b.

8.4 Styrgrupp

Styrgruppen har bestått av representanter från

- Föreningen Sambruk
- SKL (etapp 1)
- Göteborgs stad
- Nacka kommun
- Västra Götalandsregionen

Styrgruppen har sammanträtt i anslutning till respektive beslutspunkt i projektplanen.

Fem möten har genomförts under projektiden.

Föredragande inför styrgruppen har varit projektledaren.

Resultat och slutsatser i projektet

9 En lösning på problemet, inte visa problemets omfattning

9.1 Saknas lösning på marknaden

En av de inledande aktiviteterna var att kartlägga potentiella, befintliga och tillgängliga tekniska lösningar. De olika systemlösningar som identifierades kunde förvisso visa problemets omfattning genom att utföra sökningar efter presumtiva personuppgifter. Dock kan de inte lösa problemet då det saknas stöd för genomläsning, analys, juridisk bedömning och dokumentation.

9.2 Saknar stöd för den dagliga verksamheten

Detta medför att dessa verktyg lämpar sig till enstaka inventeringar kopplade till projekt med målet att kartlägga omfattning av förekomster av personuppgifter snarare än ett kontinuerligt och iterativt angreppssätt som direkt stödjer den dagliga verksamheten.

9.3 Strukturerad respektive ostrukturerad data

En annan iakttagelse som gjordes tidigt i projektet var att ostrukturerad data utgör den största och mest akuta utmaningen, när det handlar om genomlysning av datamängder.

Många organisationer som har påbörjat det inledande GDPR-arbetet anser att de personuppgifter som hanteras i verksamhetssystem, som klassificeras som strukturerad data, går att söka fram inom systemen. Angivande av lagligt stöd för hantering av personuppgifter ingår mestadels i definitionen av de stödprocesser som utgör grund för verksamhetsstödet. Dessutom har verksamhetssystemen en uttalad och befintlig leverantör som kan erbjuda tjänster för att förändra de processer och funktioner i systemen som är erforderliga.

Med bakgrund i den iakttagelsen gjordes avgränsningen att, i pilotprojektet, endast kravställa mot ostrukturerad data.

9.4 Ingen tydlig systemleverantör

Ostrukturerad data har inte en uttalad och befintlig systemleverantör som adresserar problemet eller för den delen erbjuder en lösning. De kan heller inte tvingas att ta ansvar för innehållet i en gemensam lagringsyta eller filhanteringssystem. Problemet har därför inte kunnat adresseras tidigare och har av den anledningen prioriterats ner eller negligerats.

9.5 Tar för mycket tid och kräver juridisk spetskompetens

Det är tydligt och uppenbart att det krävs omfattande resurser i arbetet med att genomlysa, filtrera, granska, genomföra juridisk bedömning och dokumentera beslut. Att bedöma den exakta tidsåtgången är mycket svårt att avgöra, utan att ha genomfört hela arbetet. För att synliggöra den tid som varje organisation behöver lägga på dessa aktiviteter har projektet tagit fram en beräkningsmodell och genomfört ett antal simuleringar som bygger på följande antaganden.

- *Antagande 1; ett genomsnittligt dokument är 2 MB stort*
- *Antagande 2; det tar i genomsnitt 2 minuter att läsa igenom ett dokument*
- *Antagande 3; 10% av den totala mängden datafiler innehåller någon form av personuppgift*
- *Antagande 4; genomsnittlig tid för juridiskt beslutsfattande om laglig grund är 3 minuter*
- *Antagande 5; genomsnittlig tid för dokumentation av beslut är 2 minuter*

Sammantaget, resulterar ovanstående simulering i ett scenario, där den totala arbetstiden blir 2.604 arbetsdagar, motsvarande knappt tolv årsarbete, per Terabyte data.

Med viss empirisk kunskap kan vi dessutom uppskatta att en genomsnittskommun, med 35.000 invånare, behandlar 15 - 20 Terabyte data. Detta scenario för manuell bearbetning av en organisations samlade, ostrukturerade data, visar med stor tydlighet att arbetsbördan blir ofantligt stor.

9.6 Befintliga licensmodeller

De flesta tillgängliga licensmodeller för hantering av information är baserade på nyttjandegrad eller lagringsutrymme, vilket är en stor utmaning inom kommunens verksamhet, eftersom det då krävs ett omfattande inventeringsarbete innan beslut om val av systemlösning kan genomföras.

10 Kravställning för önskad lösning

10.1 Ostrukturerad data

Systemet ska kunna hantera olika former av ostrukturerad, data även de som lagras utanför den lokala miljön till exempel OneDrive och Office 365. Även e-postsystem räknas till ostrukturerad data och behöver kunna hanteras enligt samma regelverk som lagrade filer.

10.2 Strukturerad data (omfattas ej av förstudieprojektets förslag på lösning)

Data som är lagrad i databaser inom verksamhetssystem, s.k. strukturerad data, ska kunna hanteras på databastabellsnivå. Inom databaser och strukturerad data återfinns ofta ostrukturerad data i form av fritextfält och bifogade filer. Även denna form av data behöver kunna hanteras inom systemet.

10.3 Identifiering av datakällor

Det behövs stöd för att kunna koppla genomlysningssystemet till respektive identifierad datakälla med erforderlig behörighet att läsa filer och data som lagras. Det behöver även finnas utrymme för att ange e-postadress till ansvarig för respektive datamängd. Detta för att kunna distribuera sökresultat för vidare granskning.

10.4 Arbetsflöde för inventering

Varje handläggare med ansvar för en datamängd ska få notifiering efter genomlyst datamängd genom e-post. Denna e-post ska länka vidare till den bruttolista som ska granskas. I respektive granskning av dokument ska de misstänkta förekomsterna av persondata visas genom färgmarkeringar som visualiserar olika typer av presumtiva personuppgifter som systemet har identifierat. Detta för att effektivisera granskningen av dokument.

10.5 Angivande av metadata

Metadata är lagrad data om data. Varje granskning ska resultera i ett antal metadataangivelser. Det metadatalager som skapas ska innehålla information som exempelvis avvisande av personuppgift, när ett dokument som systemet misstänker innehåller personuppgifter i själva verket inte gör det. Sådana dokument ska undantas i kommande genomlysningar enligt principen att människans bedömning alltid trumfar maskinens bedömning. Annan typ av metadata som ska lagras är typ av personuppgift som identifierats och även angivande av laglig grund.

10.6 Angivande av laglig grund

De olika typer av laglig grund som systemet inledningsvis ska ha stöd för är avtal, samtycke, myndighetsutövning, rättslig förpliktelse, skydd av grundläggande intresse samt berättigat intresse.

10.7 Koppling till externa källor för stöd i beslut

I samband med granskning och angivande av laglig grund ska systemet kunna visa stöd för beslut i form av befintliga checklistor och stödtexter som hämtas från branschorganisationer.

10.8 Minskat behov av juridisk spetskompetens

Med hjälp av beslutsstödet som kan hämtas från branschorganisationer underlättas arbetet med angivande av laglig grund så att behovet av juridisk spetskompetens minskar vilket ytterligare skapar möjlighet att effektivisera arbetet.

10.9 Icke tidigare identifierade personuppgifter

Den som granskar datamängderna kommer med stor sannolikhet identifiera definitioner av personuppgifter som inte har identifierats av systemet i samband med genomlysning. Denna nya definition är tänkt att användas i kommande genomlysningar.

10.10 Kognitivt lärande (lokalt installerad programvara)

Genom kognitivt lärande och annotationsteknik ska systemet analysera hur definitionen är grammatiskt konstruerad och sammansatt. Denna annoterade beskrivning av den nya definitionen av personuppgift ska skickas till en centralt placerad AI-motor.

10.11 AI motor (centralt placerad programvara, i "molnet")

Den centralt placerade AI-motorn ska lagra nya och befintliga definitioner av personuppgifter i en databas med ständigt uppdaterade gemensamma definitioner som ska används som sökalgoritm av alla förekommande installationer av systemet.

10.12 Användargenererat innehåll

Systemet ska bygga på samverkan genom att alla som använder systemet också kommer att tillföra nya och skarpare definitioner vilket medför att sökresultaten, och systemet, blir bättre över tid. Detta gäller:

- Ständigt uppdaterade gemensamma definitioner
- Definitioner av olika personuppgifter

- Metadata om befintliga och nya mönster
- Annotationer - Kognitivt lärande (AI)

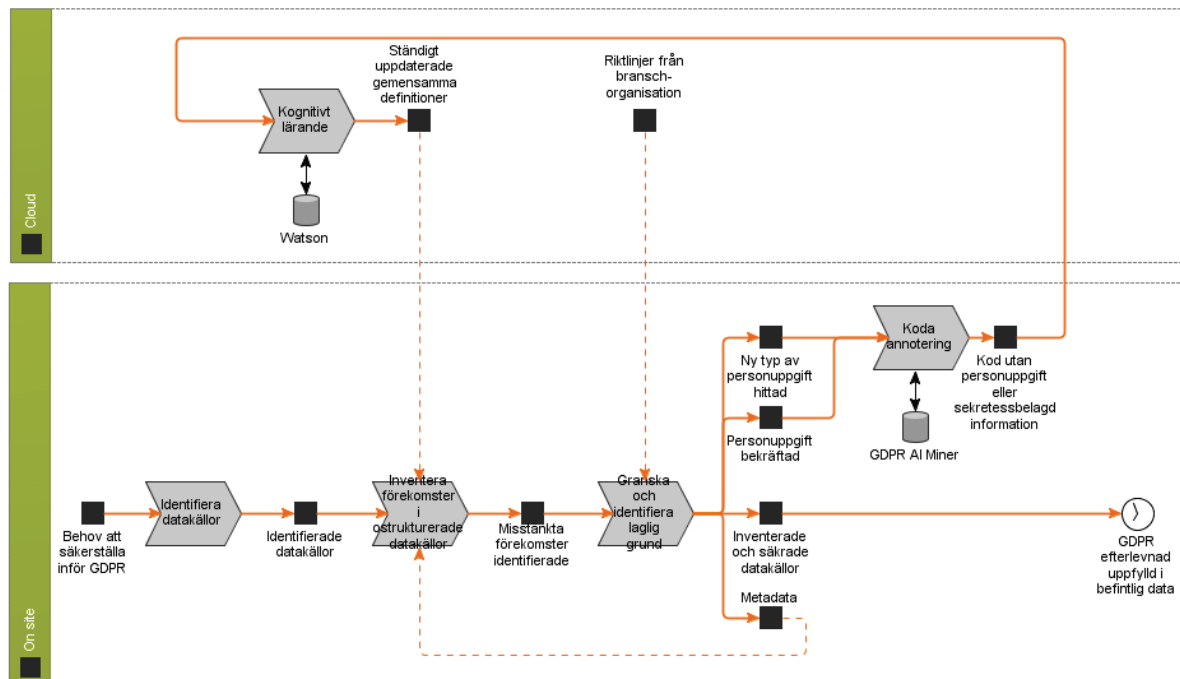
10.13 Ingen data får lämna den lokala installationen i en organisations IT-miljö

En mycket viktig del i kravställningen är att lösningen blir juridiskt oklanderlig, även utifrån GDPR. Det vill säga att inga personuppgifter eller känslig information under några omständigheter får lämna kommunens IT-miljö. Om så vore fallet, skulle ett sådant förfarande strida mot regelverket i GDPR. All genomlysning, listning av sökfiltrering, resultat av granskning eller angivande av lagligt stöd ska genomföras lokalt och ska stanna i den lokala installationen.

10.14 Ett stöd för att lösa problemet

Med dessa krav på systemet som ska levereras, skapas förutsättningar för ett verktyg som dels underlättar arbetet med kartläggning inför GDPR, men även ett stöd för respektive medarbete i det dagliga arbetet.

10.15 Processpaketering (del av)



Detaljer av processbeskrivningen ovan finns att tillgå via Förening Sambruks webbplats.

Lösningen – Aigine

Inledning

Föreningen Sambruks förstudieprojekt och dess resultat och leveranser har dokumenterats i projektets slutrapport.

Med underlaget från förstudieprojektet initierades och genomfördes ett agilt utvecklingsprojekt, i samverkan mellan företagen Aigine AB, IBM Svenska AB och Elinar OY Ltd (Finland).

Utvecklingsprojektet har finansierats av de deltagande företagen, utanför Föreningen Sambruks projektbudget.

Avstämning mellan utvecklingsgruppen och projektets kravställning har skett kontinuerligt.

I denna Bilaga beskrivs den framtagna, leveransklara lösningen, vilken mer än väl motsvarar förstudieprojektet kravställningar.

1 Sammanfattning

Resultatet av utvecklingssamarbetet mellan företagen Aigine AB, IBM Svenska AB och Elinar OY Ltd (Finland) är en komplett, leveransklar lösning med följande egenskaper:

- ✓ *Analys och filtrering av alla datakällor - därigenom sorteras bort dokument som inte innehåller personuppgifter. På så sätt elimineras behov av manuell genomläsning av dokument som inte behöver granskas.*
- ✓ *Presentation av dokumenten med presumtiva personuppgifter markerade. Detta medför att genomläsningen går väsentligt mycket fortare.*
- ✓ *Återanvändning av kognitiva förmågor*
- ✓ *Hjälp för handläggare att hitta personuppgifter och föreslå den lagliga grunden för behandlingen, genom en kontextuell kunskapsdatabas och Artificiell Intelligens. Detta gör dels att analysen går fortare, men även att analysen kan ske av personal som inte har kvalificerad juridisk utbildning*
- ✓ *I applikationens användargränssnitt anges direkt den lagliga grunden för dokumentet. Detta sparar tid, eftersom man inte behöver dokumentera beslutet separat*
- ✓ *Lösningen innebär en väsentlig förändring av hur Artificiell Intelligens används. Genom att skapa koden som krävs för upplärning av Artificiell Intelligens **lokalt**, säkerställs att ingen information lämnar den lokala IT-miljön. På så sätt används det faktiska arbetet för att hela tiden göra den AI-verktyget bättre, dvs mer effektivt. Dessa vässade kognitiva förmågor kommer sedan alla som använder lösningen till gagn; genom förbättrad filtrering, markering och förslag till laglig grund*

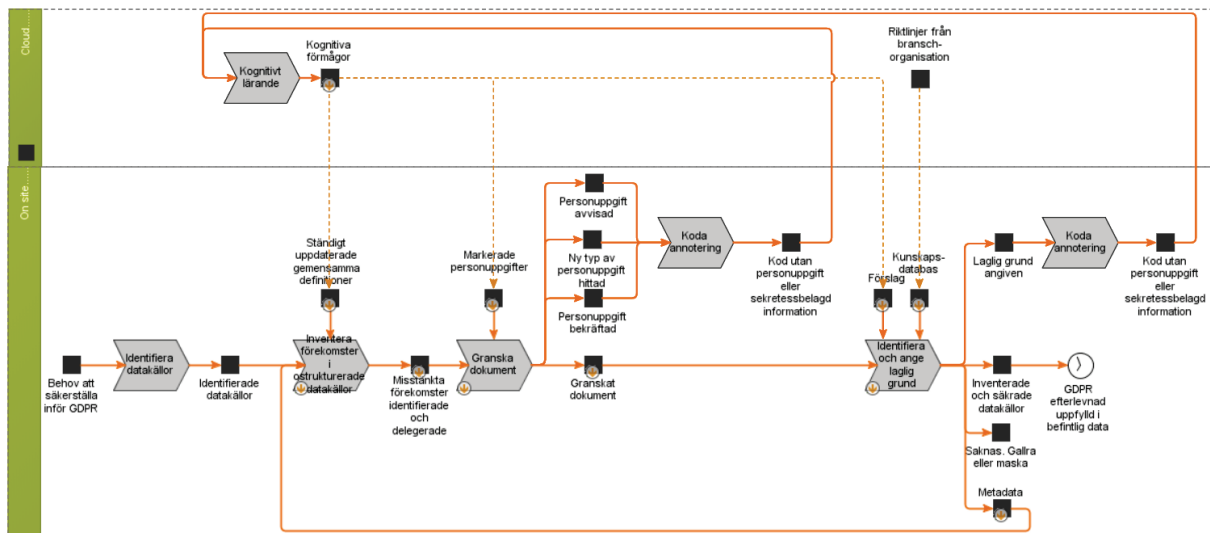
Den utvecklade lösningen heter **Aigine** och levereras av IBMs återförsäljare (ex. Atea, CGI, Evry, Pulsen och Tieto) som en lösning för hantering av personuppgifter i ostrukturerad data. Lösningen är paketerad på sådant sätt att endast en till två timmars arbetstid krävs för produktionssättning, under förutsättning att åtkomst finns till relevanta datakällor. Installation sker lokalt hos respektive organisation och läser kognitiva förmågor från molnet.

Alla bidrar till AI-verktygets förmåga/kunskap - alla får ta del av de nya gemensamma förmågorna

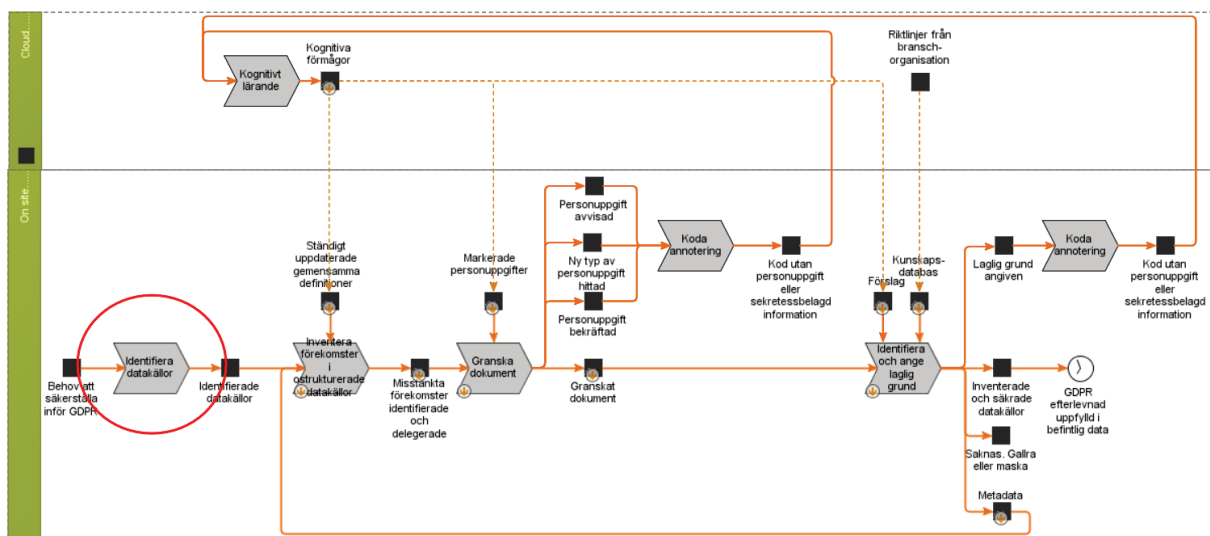
2 Funktions- och teknisk beskrivning av Aigine

Utvecklingen av den leveransklara tjänsten, som beskrivs i målformuleringen av förstudieprojektets projektbeskrivning (se också kapitel 1 till 10 i projektets slutrapport), har skett löpande från januari 2018. Utvecklingsteamet har bemannats och finansierats av Aigine, IBM och Elinar. Utvecklingsteamet har utgått från de kravkomponenter, riktlinjer och processbeskrivningar som har tagits fram av Föreningen Sambruks förstudieprojekt och genomfört ett agilt utvecklingsprojekt med stort fokus på användbarhet och funktion. Resultatet av utvecklingsprojektet är en tjänst som levereras genom IBM:s återförsäljarnät och genom Föreningen Sambruks kanaler. En absolut målsättning vid framtagandet av Aigine har varit att minimera tjänstebehovet för att få lösningen produktionssatt. GDPR generellt, och framförallt i förhållande till ostrukturerad data, utgör en tillräckligt stor utmaning. Lösningen är därför skapad för att minimera kostnader och tid för sådant som inte tillför värde till GDPR-arbetet. Målsättningen har varit att så långt som möjligt eliminerat sådana funktioner och sådan funktionalitet som kräver en tjänsteinsats.

2.1 Funktionsbeskrivning



2.2 Installation och konfiguration



För genomsökningen laddar Aigine ner de senaste definitionerna från den del av Aigine som drifas som molntjänst. Detta gör att prestandan i sökningen ökar vid varje genomsökning.

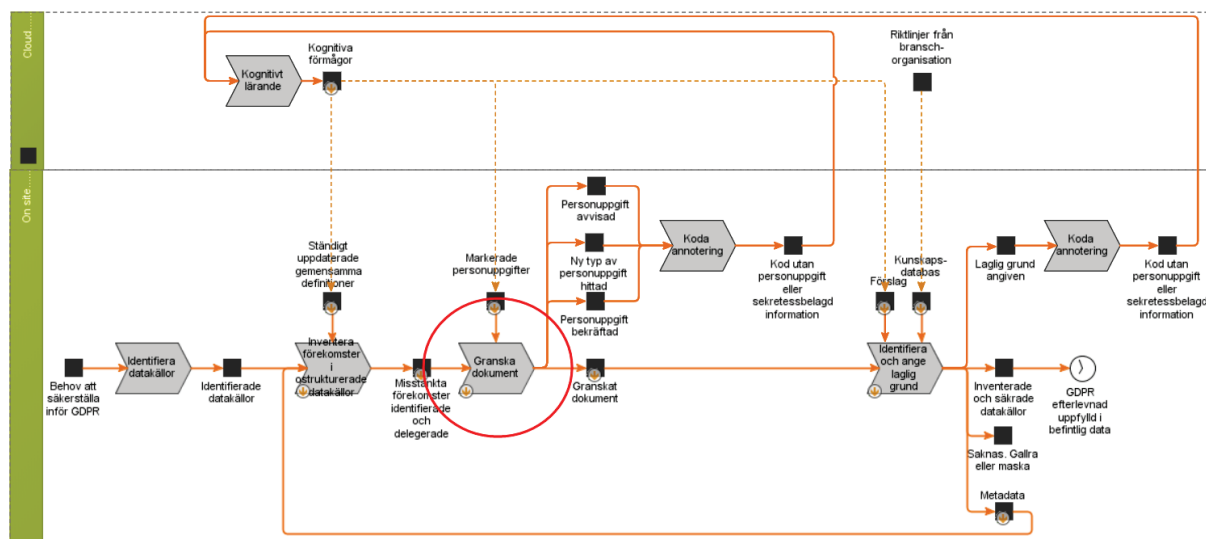
2.7 Utfiltrering av dokument som innehåller presumtiva personuppgifter

Aigine använder de nedladdade definitionerna för att avgöra om dokumentet innehåller presumtiva personuppgifter. Enbart misstänkta dokument tas med i underlaget för granskning och bedömning. Vid förnyad sökning undantas dokument som redan granskats genom användandet av metadatalagret.

2.8 Tidsbesparing

Aigines filtrering av dokument som inte innehåller personuppgifter gör att antalet dokument som måste granskas minskar med 85-95%.

2.9 Granska

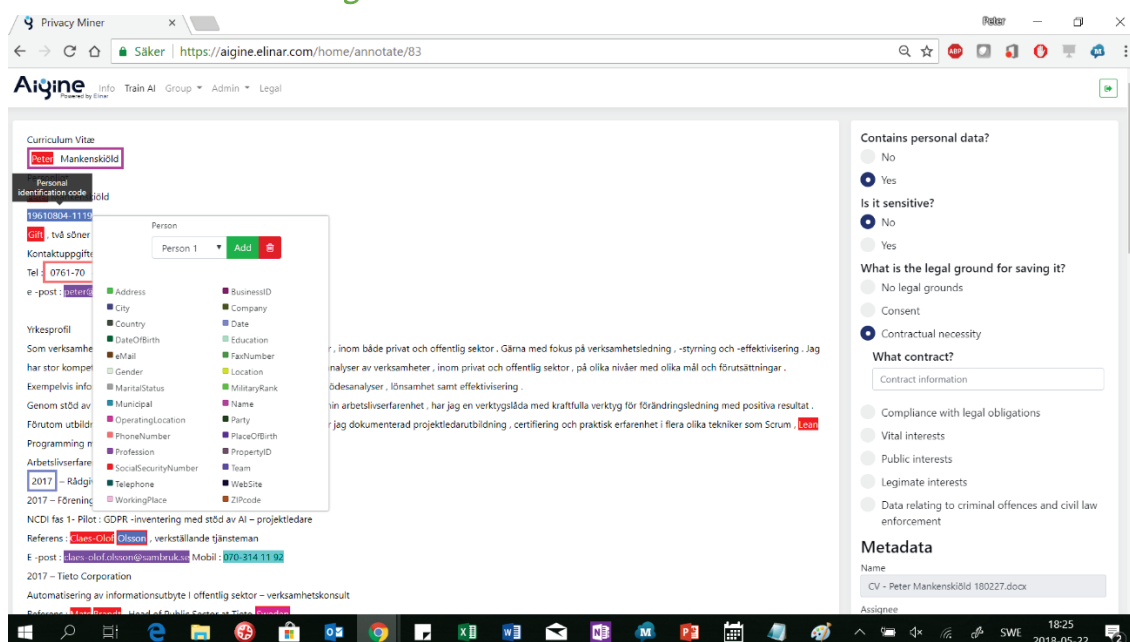


Aigine har skapats för att erbjuda en lösning på utmaningen med ostrukturerad data inför GDPR. Av denna anledning har vi tagit funktionalitet från BPM – Business Process Management och byggt in dessa i lösningen. Aigine har därför en digital process som automatiserar visst arbete, ökar effektiviteten i det manuella arbetet och tillser att det verkligen blir gjort.

2.10 Workflow

Aigines workflowmotor delar upp träfflistan per datakälla och skickar en unik lista till varje datakällansvarig via e-post. En datakälla kan definieras ner på mappnivå. Listan innehåller de dokument som innehåller presumtiva personuppgifter, och som inte har granskats av en människa tidigare. Varje dokument i listan kan klickas för att öppna det webbaserade, grafiska gränssnittet.

2.11 Grafiskt användargränssnitt



Aigine innehåller ett grafiskt gränssnitt för att underlätta granskningen av dokumentet.

2.11.1 Hittade personuppgifter visas markerade

De presumtiva personuppgifterna visas markerade vilket snabbar på genomläsningen och granskningen.

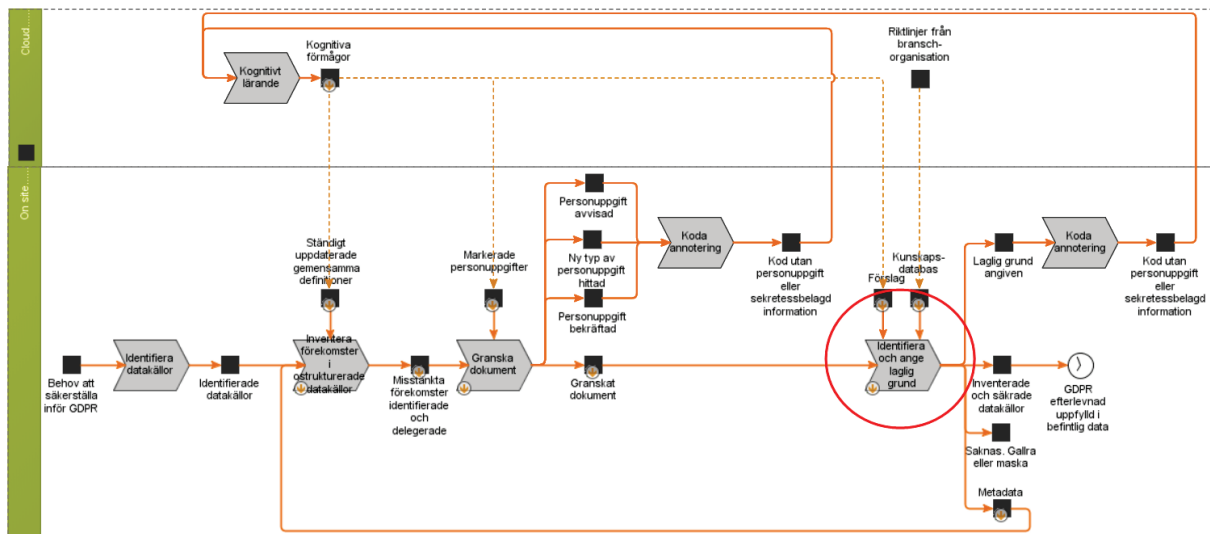
2.11.2 Inbyggt stöd för att bekräfta eller avvisa förslag på personuppgift

Verktøget innehåller intuitiva funktioner för att markera, avvisa eller bekräfta personuppgifter. Här anger man även vilken typ av personuppgift det handlar om. Denna information krävs i registerförteckningen.

2.11.3 Tidsbesparing

Genom att visa presumtiva personuppgifter markerade direkt i dokumentet snabbas granskningen på med mellan 80-95%.

2.12 Avgöra



Kraven på registerförteckning säger inte enbart att man måste veta vilken typ av personuppgifter man sparar var, krav finns även att ange ändamålet med behandlingen och därmed angivande av laglig grund för hanteringen.

2.13 Kontextuell databas

Aigine innehåller en kontextuell databas som visar relevant information för de olika juridiska ställningstaganden som måste tas i förhållande till dokumentet.

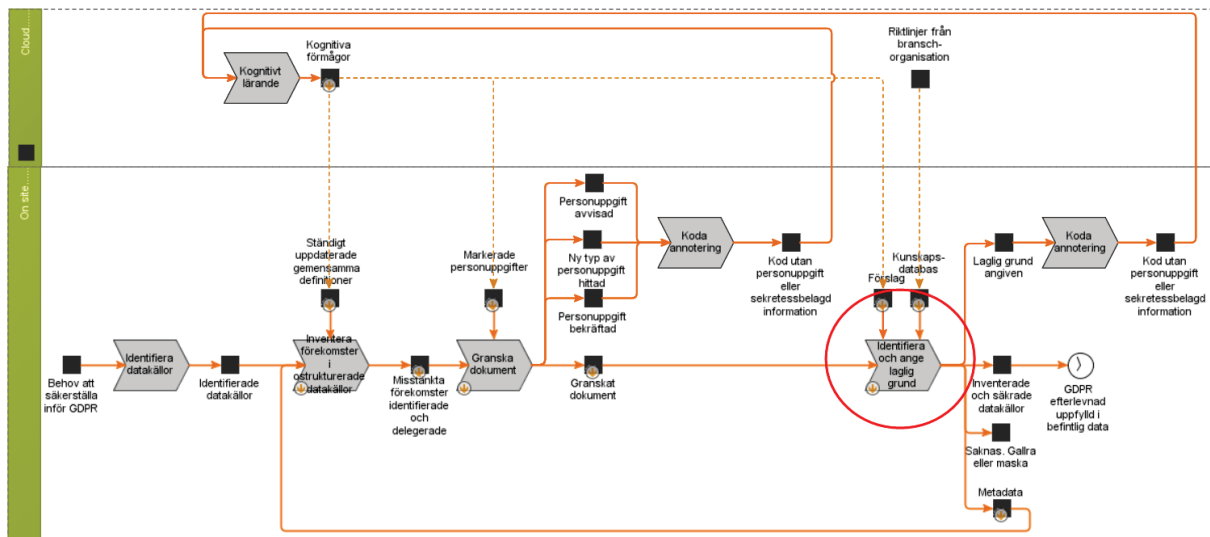
Som standard visas riktlinjer från före detta Datainspektionen, nu Integritetsskyddsmyndigheten, som är tillsynsmyndighet för GDPR.

Aigine kan även konfigureras för att visa information från andra källor.

2.14 Tidsbesparing och möjlighet att sprida på fler resurser

Tillgång till en kontextuell kunskapsdatabas direkt i användargränssnittet effektiviserar bedömningsarbetet. Den som gör bedömningen behöver inte själv söka kontext och information från spridda informationskällor. Den kontextuella kunskapsdatabasen fungerar även som kompetensförstärkning i realtid, vilket innebär att uppgiften kan utföras av personal med lägre juridisk skolning. Detta gör att en effektivitetsökning på 80-90 % kan uppnås, samtidigt som spridningen på andra resurser gör att eventuellt anställda jurister inte blir flaskhalsar i processen.

2.15 Dokumentera



2.15.1 Inbyggt smart digitalt formulär för direkt dokumentation

Gränssnittet i Aigine innehåller ett smart, digitalt, formulär som direkt fylls i enligt kraven för registerförteckning. Genom att använda modern teknik för informationsinsamling kan vi säkerställa att dokumentationen verkligen sker, samtidigt som arbetet underlättas för den som utför dokumentationen. Då varje dokument, presumtivt, är en egen datakälla, och därför måste utgöra ett eget objekt i registerförteckningen, är det oklart hur detta ska lösas på annat sätt. Kunder är hänvisade till dokumentation i excel, vilket varken säkerställer kvalitet eller att arbetet verkligen genomförs.

2.15.2 Förslag på laglig grund från AI

De kognitiva förmågorna i Aigine föreslår även lämpligt lagligt stöd för hanteringen av personuppgifter i det enskilda dokumentet. Beslutet tas dock alltid av en människa. Dessa förslag snabbar också på det juridiska bedömningsarbetet genom att peka granskaren i rätt riktning. Detta innebär, tillsammans med den kontextuella databasen, att kompetenskraven på granskaren minskar.

2.15.3 Tidsbesparing och möjlighet att sprida på fler resurser

Att de kognitiva förmågorna föreslår laglig grund effektiviserar det juridiska bedömningsarbetet. Den faktiska effektivitetsvinsten är beroende av många faktorer, där den viktigaste är granskarens tilltro till Als förmåga. Träffsäkerheten i förslagen kommer öka med tid, allt eftersom de kognitiva förmågorna vässas. En effektivitetsökning på 50-90 % är möjlig, och i en förlängning är det möjligt att även beslutsfattandet sker av AI. En sådan utveckling innebär att inget mänskligt arbete återstår i aktiviteten.

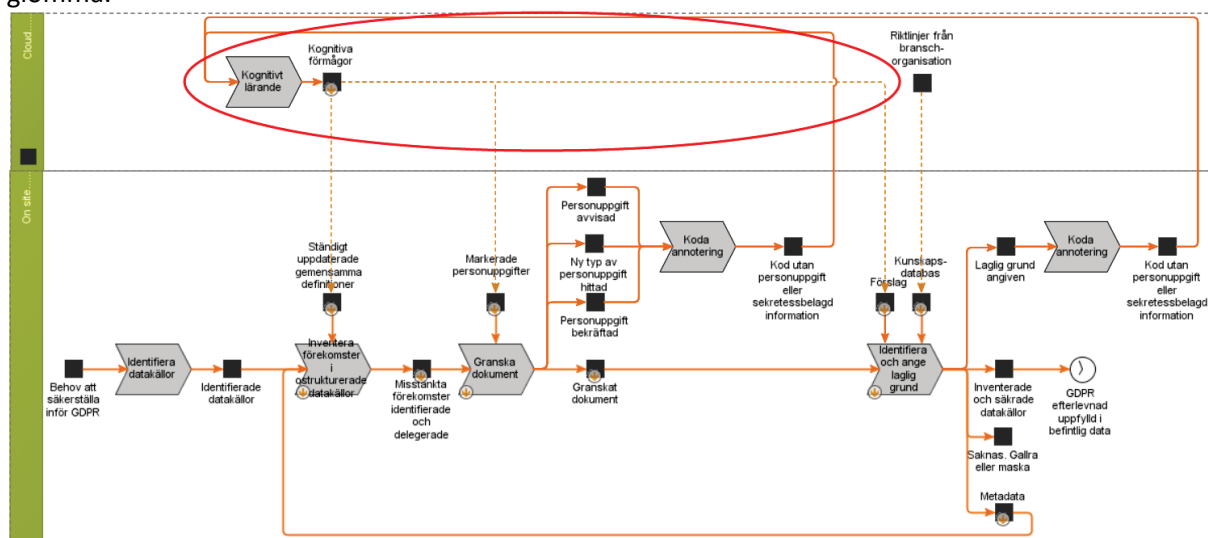
Aigine spårar förslag och faktisk bedömning vilket innebär att det existerar realtidsdata på de kognitiva förmågornas träffsäkerhet. Eftersom inte heller människor gör rätt i 100 % av fallen gör denna data det möjligt att avgöra när den Artificiella Intelligensens träffsäkerhet överstiger människans. Först då kan det faktiska beslutsfattandet överföras.

2.16 Kognitiva lärandet

Den funktionalitet som erbjuds med Aigine har som syfte att minska den behövliga resurstiden för att uppnå ett visst resultat. För att säkerställa ökad prestanda över tid innehåller därför Aigine kognitivt lärande med Artificiell Intelligens.

En förutsättning för kognitivt lärande är att tillräcklig data finns tillhands för mönsterigenkänning och s k deep learning.

Målsättning är också att Aigine levererar användarvärde direkt efter installation, med minimal arbetsinsats, varför tidskrävande traditionella upplärningsprojekt passar mindre bra. Aigine använder därför verkligt mänskligt arbete som grund för sin upplärning och förbättring av sina kognitiva förmågor. Med Aigine är AI inte längre en elev med individuell kursplan och speciallärare, utan en trainee som får hänga med på allt som händer och som bidrar efter, en hela tiden ökande, förmåga. Aigine verkar inom ett nytt rättsområde, där lagändringar och tillkommande rättspraxis innebär att AI snabbt måste läras om. Aigine använder sig därför inte av en fullständigt självlärande AI-motor utan har valt programmeringsbar AI, Watson. Detta gör det möjligt att snabbt få AI att lära om och också glömma.



2.17 Annoteringskodning lokalt

För att säkerställa följsamhet mot både GDPR och den svenska offentlighets och sekretesslagen är det en förutsättning att aldrig personuppgifter eller presumtivet sekretessbelagd information lämnar kunden. Arkitekturen i Aigine är därför utformad på ett sådant sätt att kunddata aldrig skickas från kunden. Istället är en del av AI, specifikt den del som skapar den annoteringskod som krävs för kognitivt lärande, placerad lokalt hos kund. Denna kod utgör ren maskinkod och kan inte på något sätt återskapas till förståelig information.

Annoteringskod skapas vid varje granskning och angivande av laglig grund.

Observera att detta inte är en krypteringslösning, utan framtagande av anonymiserad programkod.

Kryptering och certifikathantering hanteras senare i dokumentet

2.17.1 Exempel på annoteringskod innan Aigines anonymisering

Annoteringskod som traditionellt används av AI innehåller kod med personuppgifter i klartext:

```
4 | I Knudsen | N , | u Elise | F <nl> | x 5 | I Sletteveien | T 10 | I 1617 | Z FREDRIKSTAD | N <nl> | x 6 | I
30073347897 | S <nl> | x 7 | u 410%2032%20911 | O <nl> | x 9 | I EliseKnudsen@teleworm.us | E <nl> | x 28 |
I Lima | N Ellen | F <nl> | x 29 | I Kvalamarka | T 30 | I 5514 | Z HAUGESUND | T <nl> | x 30 | I 11094847615 | S
<nl> | x 31%20979%2031%20963 | O <nl> | x 111 | I Sara | F Dimmen | N <nl> | x 112 | I 14077548891 | S <nl>
| x 116 | I Even | C Pedersen | C <nl> | x 117 | I 05026548367 | S <nl> | x 121 | I Filip | F Berg | N <nl> | x 122 | I
22117547918 | S <nl> | x 126 | I Leonard | F Hult | N <nl> | x 127 | I 22117548388 | S <nl> | x 131 | I Victoria | F
Høydal | N <nl> | x 132 | I 14049847875 | S <nl> | x
```

2.17.2 Exempel på annoteringskod efter Aigines anonymisering

För att säkerställa att inga personuppgifter eller sekretessbelagd data lämnar kund, sker kodutveckling lokalt. Denna gör koden oförståelig i förhållande till den bakomliggande datan. Nedanstående är ovanstående kod tvättad.

```
Integer1 | I LastName001 | N , | u FirstName001 | F <nl> | x Integer2 | I Location001 | T Integer3 | I  
ZipCode001 | Z LastName002 | N <nl> | x Integer4 | I SocSecNum001 | S <nl> | x 7 | u PhoNum001 | O <nl> | x  
Integer5 | I Email001 | E <nl> | x Integer6 | I LastName003 | N FirstName002 | F <nl> | x Integer7 | I  
Location002 | T Integer8 | I ZipCode002 | Z Location003 | T <nl> | x Integer8 | I SocSecNum002 | S <nl> | x  
PhoNum002 | O <nl> | x Integer10 | I FirstName003 | F LastName004 | N <nl> | x Integer11 | I  
SocSecNum003 | S <nl> | x Integer12 | C <nl> | x Integer13 | I SocSecNum004 | S <nl> | x Integer14 | I  
FirstName004 | F LastName006 | N <nl> | x Integer15 | I SocSecNum005 | S <nl> | x Integer16 | I  
FirstName005 | F LastName007 | N <nl> | x Integer17 | I SocSecNum006 | S <nl> | x Integer18 | I  
FirstName006 | F LastName008 | N <nl> | x Integer19 | I SocSecNum007 | S <nl> | x
```

2.18 Kognitivt lärande i molnet

Det kognitiva lärandet sker i molnet, genom användandet av anonymiserad annoteringskod från varje granskning och juridisk bedömning. AI i molnet ligger installerad på en server som kontrolleras av Aigine AB.

2.19 Kognitiva förmågor förbättras

Genom denna lösning kan vi säkerställa att AI får tillräcklig data för att snabbt förbättra sina kognitiva förmågor. Detta gör att Aigine tämligen omgående förbättrar förmågan att förstå vad som utgör personuppgifter, hur personuppgifter ska markeras, samt, utifrån dokumentets struktur och disposition, vilken laglig grund som kan finnas för att spara det.

2.20 Alla får del av de förbättrade förmågorna

Aigines vässade kognitiva förmågor kommer alla kunder till del, genom att de laddas ner från molnet till kundens lokala installation. Detta gör att träffsäkerheten i filtreringen hela tiden ökar, vilket kan innebära både fler eller färre träffar, beroende på dokumentens innehåll.

Genom en kontinuerligt förbättrad förståelse för sammanhang ökar också förmågan att markera relevanta personuppgifter i varje dokument, vilket ytterligare effektiviserar den mänskliga granskningen.

Aigines förslag till laglig grund kommer också förbättras med högre relevans och träffsäkerhet. Inom en överskådlig tid kommer fortfarande människor fatta det juridiska beslutet, men Aigines förbättrade förslag kommer snabba på detta beslutsfattande ytterligare.

Alla bidrar till AI-verktygets förmåga/kunskap - alla får ta del av de nya gemensamma förmågorna

2.21 Tidsbesparing

Aigine har kombinerat erfarenheter från Business Process Management, användarcentrerad utveckling och juridiskt bedömningsarbete i syfte att minska den tid det tar att inventera, granska och dokumentera innehållet i ostrukturerade datakällor. Minskningen måste vara så stor att det blir både praktiskt och teoretiskt möjligt för organisationer att ens närma sig utmaningen.

Aigines kunskapsdatabas och förslag på lagligt stöd innebär även att kraven på juridisk skolning hos de resurser som genomför arbetet minskar. Även detta bidrar till att göra arbetsinsatsen både teoretiskt och praktiskt möjligt då arbetet kan fördelas på fler resurser i kundens organisation.

Det är alltid svårt att bedöma den verkliga tidsbesparingen då denna är beroende av ett stort antal faktorer. Samtidigt kommer prestandan, och därmed effektiviseringen, öka med skärpta kognitiva förmågor.

2.22 Exempel på arbetsinsats utan stöd

Mängd ostrukturerad data	1	terabyte
Genomsnittlig dokumentstorlek	2	mb/dokument
Genomsnittlig tid för genomläsning	2	minuter/dokument
Procentandel som innehåller personuppgifter	10%	procent
Genomsnittlig tid för juridiskt beslutsfattande om laglig grund	3	minuter/dokument
Genomsnittlig tid för dokumentation i registerförteckning	2	minuter/dokument
Total arbetstid	11,84	manår

2.23 Exempel på potentiell effektivisering med stöd

Effektivisering filtrering	90%	procent
Effektivisering genomläsning %	90%	procent
Effektivisering beslutsfattande %	90%	procent
Effektivisering dokumentation %	90%	procent
Total effektivisering	97%	procent
Återstående arbetstid	0,33	manår

Beräkningen avser 1 Terabyte data. Antagandet har gjorts att varje dokument är 2mb stort i genomsnitt. Att läsa igenom varje dokument för att utröna huruvida det innehåller personuppgifter antas ta 2 minuter per dokument.

Här antas att 10% av dokumenten faktiskt innehåller personuppgifter, vilket ligger i linje med gjorda inventeringar under lösningens framtagande.

Tiden att göra den juridiska bedömningen antas ta 3 minuter. Detta inkluderar informationsinsamling från källor som Integritetsskyddsmyndigheten. Dokumentationen av varje dokument antas ta 2 minuter. Här måste typ av personuppgifter, ändamål, laglig grund, gallringstid etc. anges.

Aigine effektiviserar arbetet genom att endast presentera dokument som faktiskt innehåller presumtiva personuppgifter genom filtrering. Detta effektiviserar hanteringen med 90%.

Aigine effektiviserar genomläsningen av dokumentet med 90% genom att visa hittade personuppgifter markerade.

Aigine effektiviserar arbetet med beslutsfattandet att hitta laglig grund med 90% genom att dels föreslå detta med sina kognitiva förmågor, dels tillhandahålla kontextuell hjälp direkt från relevanta källor.

Aigine effektiviserar dokumentationen med 90% genom att tillhandahålla en dokumentationsmöjlighet direkt i granskningsvyn. Informationen avges i ett smart, digitalt, formulär där vara relevanta fält visas.

3 Teknisk beskrivning

3.1 Licensbegränsning

Aigine består av ett stort antal mycket kompetenta programvaror, paketerade för att lösa utmaningen med personuppgifter i ostrukturerad data och erbjuda en helhetslösning till en rimlig kostnad. Nyttjanderätten till samtliga ingående programvaror är därför strikt begränsad till funktionsbeskrivningen i detta dokument och det användaravtal som kunden tecknar. Kund som önskar nyttja ytterligare förmågor i ingående programvaror måste således teckna separat licens för dessa funktioner.

3.2 Ingående komponenter

3.2.1 Aigine Unstructured Data Inventory Engine

Lokalt installerad mjukvara som använder förmågor från StoredIQ, Watson NLU och Elinar AI Miner samt egna funktioner.

3.2.2 Användargränssnitt

Mjukvaran innehåller också enklare GUI för systemkonfiguration samt GUI för slutanvändare. Detta GUI består i sin tur av flera delar, bland annat markering av presumtiva personuppgifter, enkel konfirmering av uppgifter, kunskapsdatabas samt formulär för datainsamling.

3.2.3 Workflowstöd

Mjukvaran innehåller en enklare workflowfunktion som delegerar ansvar via e-post till ansvariga.

3.2.4 Metadatalager

Mjukvaran innehåller ett metadatalager och enklare rapporteringsfunktionalitet via e-post. Basmetadata och arbetsflödesrelaterade metadata sparas om respektive dokument i en databas (*postgresql*).

3.2.5 StoredIQ

Tillför automatiserade policys för hantering av ostrukturerad data. Ökar förmågan att ta underbyggda beslut baserat på stora mängder information. Skapar mönster och hanterar såväl ostrukturerad som strukturerad data i samma system.

3.2.6 Watson AI och Watson NLU

Artificiell intelligens (AI) och Annotationstekniken Natural Language Understanding (NLU) kallas även för machine intelligence, MI, är intelligens som uppvisas och ses hos upplärda maskiner. Artificiell intelligens är mycket användbart i iterativa processer, där teknologin självständigt skapar information som behövs i nästa iterativa sekvens. I vissa fall kan uppgifter med stora mängder data där iterativa processer med inslag av rekursion (iterativa beräkningar med resultatåtermatningar och krav på erfarenhetslärande) vara så pass komplicerad och omfångsrika att en människa inte klarar av att komma till ett motsvarande resultat inom rimlig tid. Artificiell intelligens innebär i dessa fall en direkt och omfattande tidsbesparing. Dessutom ges stora kvalitetsökningar i och med att en uppgift kan flyttas från en människa till en dator som enligt en lärande process imiterar och förfinar det mänskliga arbetet i den lärande processen, och gör som den ska och inte kan slarva.

3.2.7 Kognitivt Lärande

Kognitiv IT och artificiell intelligens AI är begrepp inom kognitiv it (cognitive computing). Uttrycket myntades av IBM, under arbetet med Watsondatorn. Kognitiv it och AI används omväxlande för att beskriva samma saker. Beroende på om olika typerna av algoritmer, plattformar, maskinläring och neurala djuplärande (deep learning) nätverk menas. AI används oftast som en benämning på tekniska lösningar som kan utföra uppgifter som det vanligtvis krävs mänsklig intelligens för. Kognitiv it definieras av maskinintelligens, som kan beskrivas som algoritmisk funktionalitet som används för att förbättra människors prestationer, automatisera komplexa arbetsuppgifter och skapa så kallade kognitiva agenter som kan simulera människors tänkande och agerande.

Tre huvudsakliga områden för kognitiv it:

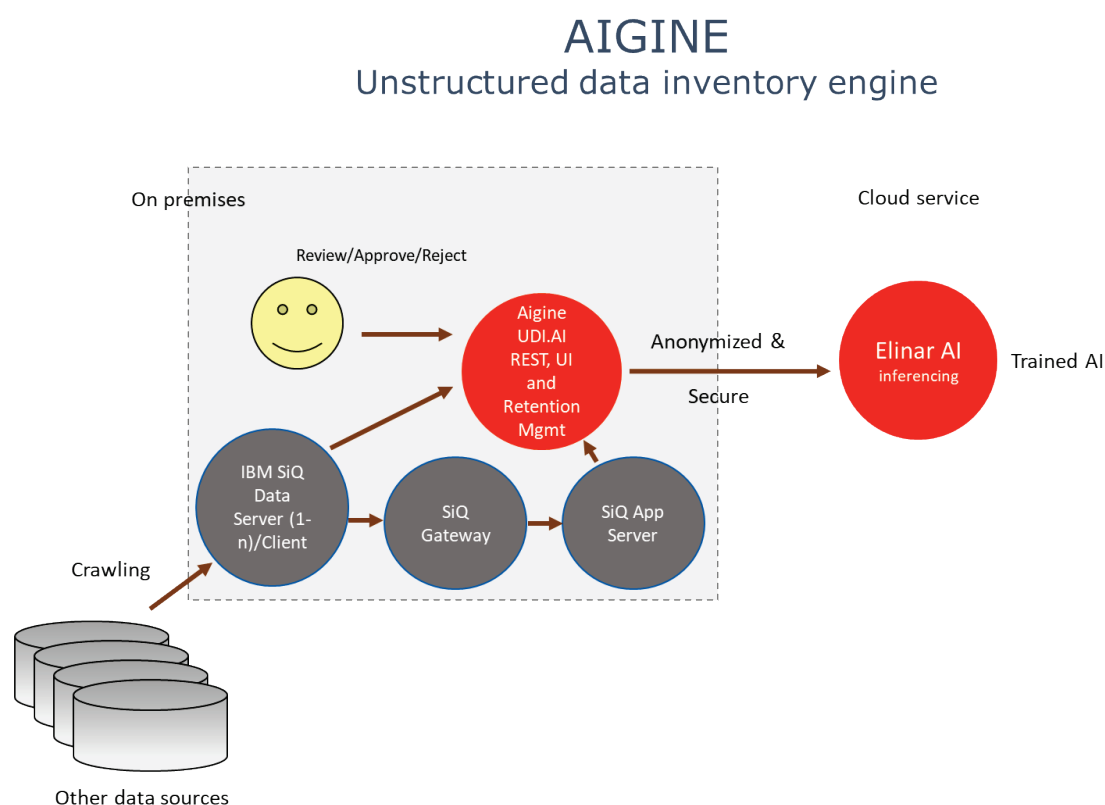
- Automatisering av upprepningsbara uppgifter, med avsikt att öka effektivitet, kvalitet och noggrannhet.
- Skapa insikter och upptäcka, identifiera dolda mönster och relationer.
- För att skapa utökade lösningar för personalisering (hyperpersonalization).

Alla tillämpningar för avancerad klassificering och prediktiv analys och att presentera lösningsförslag.

3.2.8 Elinar AI miner

System som ger annoteringsinformationen från de olika lokala systemen till den centrala molnlösningens artificiella intelligens (AI). AI miner använder en teknik som kallas flerskiktade artificiella neurala nätverk (ANN) för att skapa maskinförståelse och omvandlar den lokala informationen till personoberoende annoteringskod i processer som ligger till grund för lärande processer i molnet genom Natural Language Understanding (NLU).

4 Systemskiss



Aigine kombinerar IBM StoredIQ, IBM Watson NLU annotation module och Elinar AI-miner till en omfattande lösning för filtrering, genomläsning, beslutsfattande rörande laglig grund och dokumentation av genomförda beslut för kommuner och organisationer.

Systemet installeras och lagrar alla resultat av genomläsning, identifiering av potentiella personuppgifter, beslutsfattande, dokumentation lokalt och kommunicerar endast anonymiserad annotationsdata med den centralt baserade AI-minern

4.1 Kryptering och certifikathantering

Aigine använder HTTPS (SSL) Certifikat utfärdade av [GlobalSign](#) I denna lösning används så kallade wild card certifikat.

För kommunikation mellan Aigine UDI och AI-miner används en metod med förinstallerade rotcertifikat vilket innebär att båda komponenterna i kommunikationskedjan har var sin del av certifikatstrukturen som möjliggör kommunikation.

4.2 Stöd för webbläsare

Vi rekommenderar senaste versionen av någon av följande webbläsare:

- Mozilla Firefox

- Google Chrome
- Apple Safari
- Microsoft Edge
- Internet Explorer

Alla ovanstående webbläsare stöds av Aigine. Prestanda kan variera mellan olika webbläsare. Bäst prestanda uppnås med Chrome och Firefox (quantum).

4.3 Stöd för filformat och protokoll

För scanning och inventering av datakällor används StoredIQ från IBM. Denna mjukvara har ett mycket stort stöd för olika filtyper och protokoll. Antalet filtyper och protokoll som stöds utökas dessutom kontinuerligt, och nya versioner av StoredIQ, med stöd för nya filtyper, kommer ingå i uppdateringarna av lösningen. Ytterligare information om filformat som stöds i lösningen återfinns på följande länk: <http://www.redbooks.ibm.com/redpapers/pdfs/redp5316.pdf>

4.4 Uppdateringar

Aigine uppdaterar löpande lösningen med felrättningar, säkerhetsförbättringar, prestandaoptimeringar, förbättrade gränssnitt och eventuellt tillkommande funktionalitet. Aigine har inte full följsamhet mot versionsuppdateringar av ingående komponenter, utan en kvalificering görs huruvida uppdateringar av ingående komponenter innehåller relevanta förändringar som bör införlivas i lösningen. Nya versioner kommer därför släppas två till fyra gånger per år.

5 Leverans av tjänsten

5.1 Leverans

Aigine levereras IBMs återförsäljare (exempelvis Atea, CGI, Evry, Pulsen och Tieto) som en lösning för hantering av personuppgifter i ostrukturerad data. Lösningen är paketerad på sådant sätt att endast 1-2 timmars arbetsinsats krävs för produktionssättning, under förutsättning att åtkomst finns till relevanta datakällor. Installation sker lokalt hos respektive organisation och Aigine läser kognitiva förmågor från molnet.

5.2 Prismatris för Kommuner

Prissättningen av tjänsten Aigine är baserad på antal anställda i kommunen. Prismatrisen är även differentierad beroende på om kommunen är medlem i föreningen Sambruk eller inte.

Kommuner	tjänstekostnad/månad	
Antal anställda	Sambruk	Övriga
< 1 000	15 000	16 500
1 000 - 1 500	18 500	20 350
1 500 - 2 000	22 500	24 750
2 000 - 5 000	26 000	28 600
5 000 - 10 000	30 000	33 000
10 000 >	36 000	39 600

Aigine tillhandahålls som abonnemang med 12 månaders bindningstid.