Bacheloroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for datateknologi og informatikk

Ingvild Stølen

Praktisk håndtering av ustrukturerte data i samsvar med personopplysningsloven

Bacheloroppgave i Informatikk - Informasjonsbehandling
Veileder: Torstein Elias Løland Hjelle
Mai 2019

NTNU
Kunnskap for en bedre verden

Ingvild Stølen

Praktisk håndtering av ustrukturerte data i samsvar med personopplysningsloven

Bacheloroppgave i Informatikk - Informasjonsbehandling
Veileder: Torstein Elias Løland Hjelle
Mai 2019

Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for datateknologi og informatikk

NTNU
Kunnskap for en bedre verden

Praktisk håndtering av ustrukturerte data i samsvar med personopplysningsloven

Forord

Denne oppgaven er skrevet som et avsluttende arbeid i graden bachelor i informatikk – informasjonsbehandling ved NTNU. Problemstillingen dukket opp da forfatteren fikk i oppdrag av sin arbeidsgiver å se på bedriftens mulige problemområder i forbindelse med GDPR og personvern generelt. Det kom fram at det er lite skrevet om utfordringene rundt ustrukturerte data og hvordan best håndtere disse for å sikre et best mulig personvern.

Oppgavens resultat er en strukturert og lettlest framstilling av foreslåtte tiltak til bedrifter som har behov for å sikre at deres drift er i samsvar med GDPR.

Takk til Torstein Elias Løland Hjelle for meget god og konstruktiv veiledning i prosessen. Takk også til Ivar Løkken for hjelp med korrekturlesing.

Sammendrag

Målet for denne oppgaven er å finne ut hva en virksomhet kan gjøre for å få kontroll på ustrukturerte data, spesielt med hensyn på små- og mellomstore bedrifter som ikke har mulighet til å bruke store midler. For å finne ut dette er det gjort søk i litteratur og andre kilder. I tillegg er det gjort et forsøk med å benytte Cognitive Services (maskinlæring) fra Azure som et verktøy for å få oversikt over ustrukturerte data. Sluttresultatet av arbeidet er veileder som gir råd om hvordan en virksomhet kan gå fram i arbeidet med ustrukturerte data og GDPR. Denne er publisert på ustrukturert.no.

Abstract

The purpose of this thesis is to find out what a business can do to gain control over unstructured data, with focus mainly on small and medium sized businesses with limited funds. A study of literature and other sources has been conducted. Additionally, a proof-of-concept has been made using Cognitive Services (machine learning) from Azure as a tool for getting an overview of unstructured data. The result of this work is a guide that advises on how to get a business's handling of unstructured data compliant with GDPR. The guide is published at unstrukturert.no.

Problemstilling

I 2018 tredde en ny personvernforordning i kraft i EU og EØS, The General Data Protection Regulation (GDPR). Den gjelder både for bedrifter innenfor området hvor forordningen gjelder, og for bedrifter som selger sine varer og tjenester til forbrukere innenfor området, selv om de måtte være plassert et annet sted.

GDPR gir forbruker rett til å få vite hva en bedrift har av personopplysninger, hvor det er lagret og hvordan deres persondata blir brukt. Med personopplysninger menes all informasjon som kan knyttes til en person, som navn, adresse, kjøpshistorikk og så videre. I tillegg skal en enkeltperson kunne få utlevert all data som er lagret om seg i elektronisk format. Personen har også krav på at informasjonen skal slettes, dersom det ikke er gode grunner som taler for å ikke slette.

I motsetning til tidligere lover og regler gis det under GDPR anledning til å straffe bedrifter som bryter med forordningen hardt økonomisk, med bøter på opptil 4% av den samlede globale omsetningen til selskapet. Dette gir både små og store selskaper incentiver til å følge forordningen.

Når en bedrift skal sørge for at den kan oppfylle GDPR vil det være relativt enkelt å få oversikt over data som er lagret i strukturert form, som for eksempel i kundedatabasen, HR-systemet eller et elevregister. Langt vanskeligere er det å skaffe en oversikt og sikre at man er i samsvar når det gjelder ustrukturerte data som eposter, filer, kvitteringer og lignende. Det er lite skrevet om denne problemstillingen, og det som finnes er i stor grad skrevet av bedrifter som vil selge inn sine løsninger. For en liten virksomhet som ikke har mulighet til å leie inn ressurser på dette feltet er det vanskelig å finne ut hva man kan gjøre og hvordan.

Denne oppgaven tar sikte på å skaffe en oversikt over tiltak en bedrift kan gjøre for å sikre samsvar med GDPR også for ustrukturerte data. Dette gjøres ved hjelp av en studie av mulige tiltak som kan gjøres og teknologi som kan benyttes, samt et forsøk med å innføre teknologi for en prøve-organisasjon.

Ønsket er at oversikten over fremgangsmetode og tiltak skal kunne være til nytte for små og mellomstore bedrifter som trenger denne informasjonen men ikke har råd til å kjøpe konsulenttjenester for å få oversikt. Resultatet vil derfor publiseres på ustrukturert.no presentert som en veileder i tillegg til vedlegget i oppgaven.

Innhold

Viktige definisjoner

Strukturerte data

Med strukturerte data menes data som er organisert i klart definerte datatyper/skjema. Dette vil for eksempel være data man har i en kundedatabase, i et HR-system eller lignende.

Ustrukturerte data
Med ustrukturerte data menes opplysninger som ikke er organisert i en forhåndsbestemt struktur. Dette kan finnes for eksempel som del av en tekst, en lydfil, et medieoppslag, dokument og så videre.

Forordning

I EU er forordning betegnelsen på en lov som i helhet får bindende virkning i alle medlemsstater uten at det trengs ytterlige vedtak eller godkjenninger i de enkelte medlemsland. En forordning i EU går over nasjonale lover i medlemslandene.

Behandlingsansvarlig

Den som bestemmer formålet med behandlingen av personopplysninger er definert som behandlingsansvarlig etter GDPR. Denne er ansvarlig selv om selve behandlingen er satt ut til en underleverandør.

Databehandler

Underleverandør som behandler data på vegne av en behandlingsansvarlig.

GDPR – Personvernforordningen

GDPR (the General Data Protection Regulation) ble godkjent I EU-parlamentet den 14. april 2016 og hadde ikrafttredelse den 25. mai 2018. I Norge trådte personvernforordningen i kraft 20. juli samtidig med den nye norske personvernvernloven (Lovdata, 2018). Forordningen har til hensikt å styrke og harmonisere personvernet ved behandling av personopplysninger i EU.

Personopplysninger

I forbindelse med GDPR tolkes personopplysninger som en hvilken som helst opplysning som gjelder en privatperson som er identifisert eller kan identifiseres. I personvernforordningens artikkel 4 heter det:

> *«enhver opplysning om en identifisert eller identifiserbar fysisk person («den registrerte»); en identifiserbar fysisk person er en person som direkte eller indirekte kan identifiseres, særlig ved hjelp av en identifikator, f.eks. et navn, et identifikasjonsnummer, lokaliseringsopplysninger, en nettidentifikator eller ett eller flere elementer som er spesifikke for nevnte fysiske persons fysiske, fysiologiske, genetiske, psykiske, økonomiske, kulturelle eller sosiale identitet»*

Dersom opplysningen kan knyttes til for eksempel et fingeravtrykk eller en IP-adresse gjelder også forordningen, selv om navn ikke er registrert. Det samme gjelder bilder av personer. Alt som kan knyttes til privatperson omfattes og dermed vil de aller fleste som tilbyr tjenester eller varer av noen art være forpliktet til å følge forordningen.

Sensitive personopplysninger

I artikkel 9 og 10 I personvernforordningen er det listet opp opplysninger som regnes som sensitive og dermed må behandles ekstra varsomt. Disse er:

- Opplysninger om rasemessig eller etnisk opprinnelse
- Opplysninger om politisk oppfatning
- Opplysninger om religion
- Opplysninger om filosofisk overbevisning
- Opplysninger om fagforeningsmedlemskap
- Genetiske opplysninger
- Biometriske opplysninger med det formål å entydig identifisere noen
- Helseopplysninger
- Opplysninger om seksuelle forhold
- Opplysninger om seksuell legning
- Opplysninger om straffedommer
- Opplysninger om lovovertredelser

Disse er det i utgangspunktet forbudt å behandle, men det er en rekke unntak og forbehold. I hovedtrekk må de som skal behandle slike data kunne argumentere godt for hvorfor dette er nødvendig og de skal hente inn samtykke fra personene opplysningene tilhører. Det er også en rekke unntak for medisinsk forskning, medlemslister hos trossamfunn og lignende.

Å gjennomgå de personopplysninger som finnes i en organisasjon for å avdekke om noe av det er i kategorien sensitivt er en del av arbeidet som bør gjøres i forbindelse med å sikre samsvar med personvernforordningen.

Virksomhetenes plikter etter personvernforordningen

GDPR medfører en rekke rettigheter for enkeltpersoner med tilhørende plikter for virksomhetene. I dette kapittelet følger en kort oppsummering.

Sette seg inn i regelverket

Virksomheten har ansvar for å til enhver tid ha kunnskap om hvilke lover og regler som gjelder. Datatilsynet har gitt ut en veileder som beskriver de grunnleggende personvernsprinsippene (fra artikkel 5, 6 og 9 i personvernforordningen).

Offentlige virksomheter, og en del andre virksomheter har plikt til å utnevne et personvernombud som skal ha som oppgave å følge med at virksomheten til enhver tid opererer i samsvar med personvernforordningen.

Kartlegge personopplysninger og formålet med de (behandlingsgrunnlag)

Virksomheten skal ha oversikt over hvilke personopplysninger som oppbevares og hvordan de oppbevares. En person har krav på å alltid få vite hva som finnes av personopplysninger knyttet til en selv. Dermed må virksomheten også vite dette.

For både sensitive og ikke-sensitive personopplysninger er det definert klare regler for hva som er godkjente grunner for at en virksomhet skal kunne lagre personopplysninger. Derfor må en virksomhet ikke bare vite hva de oppbevarer, men også gå gjennom hvorfor de skal oppbevare de. Personopplysninger skal kun brukes til spesifikke, uttrykkelige, angitte og legitime formål (personvernforordningen artikkel 5).

Man kan ikke oppbevare personopplysninger fordi de kan være greie å ha (datatilsynet, 2018). Det er heller ikke godt nok med en rund formulering som for eksempel “til analyseformål”. En slik kartlegging skal også brukes til å informere personer hvis personopplysninger oppbevares. Disse har nemlig krav på å få vite nøyaktig hva opplysningene brukes til og virksomheten plikter å tilgjengeliggjøre dette på en lettfattelig måte. Om virksomheten senere endrer på hva de skal bruke opplysningene til vil de måtte innhente nytt samtykke fra den enkelte.

Formålet med lagringen vil også ha innvirkning på hvor lang tid det vil være rimelig å lagre data før de slettes. Virksomheten skal ikke lagre personopplysninger lengre enn nødvendig, altså plikter virksomheten å ha en rutine for sletting av data som ikke lengre er relevant.

I tillegg til at virksomheten skal ha oversikt over personopplysninger og behandlingsgrunnlag skal det også føres protokoll over kategorier av behandlingsaktiviteter som utføres. Protokollen skal blant annet inneholde informasjon kategorier av personopplysninger, kategorier av registrerte, formål og eventuell utveksling av informasjon.

Virksomheter har ofte underleverandører som behandler personopplysninger på vegne av seg. Dette kan være for eksempel skytjenester der data er lagret. For slike tilfeller er den som er behandlingsansvarlig pliktig å påse at databehandleren etterlever lovverket gjennom å ha en databehandleravtale. Det er krav i forordningen til hva en slik avtale skal inneholde. Mange leverandører av IT-tjenester har standardavtaler de presenterer for sine kunder. Selv om det er leverandøren (databehandler) som presenterer avtalen er det likevel her kunden (behandlingsansvarlig) som har ansvar for å påse at avtalen er dekkende og innenfor lovverket.

Dersom opplysninger overføres til land utenfor EU/EØS gjelder spesielle regler.

I noen tilfeller der det er sannsynlig at en behandling medfører høy risiko for personers rettigheter og friheter skal det foretas en særlig vurdering av personvernkonsekvenser (DPIA). Det er ganske sammensatte regler for når dette er påkrevd, blant annet omfattes en del ny teknologi som behandler personopplysninger om barn, kameraovervåkning, biometriske opplysninger med mer.

Informere og innhente samtykke

Virksomheten skal sørge for at alle som det oppbevares personinformasjon om skal være informert om hvilke opplysninger som behandles og formålet. Informasjonen skal være formulert på en lettfattelig måte og det skal komme tydelig fram hvilke opplysninger som behandles og hva formålet er. Informasjonen skal også være lett tilgjengelig slik at de registrerte ikke skal være nødt til å lete for å finne den, og den skal ikke være innbakt i andre dokumenter som for eksempel brukervillkår.

Samtykke skal også kunne trekkes tilbake. Virksomheten kan ikke etter innhentet samtykke utvide behandlingen til å omfatte andre formål enn det som det ble gitt samtykke til, og det er heller ikke tillatt å formulere samtykkeerklæringer på en så vag og uspesifisert måte at det meste av databehandling kan omfattes.

Ivareta den registrertes rettighet til å bli rettet og glemt

Personopplysninger skal slettes når virksomheten ikke lengre kan hevde saklig grunn for å oppbevare de. De skal også rettes dersom de er ukorrekte eller utdaterte.

En registrert person har rett til:

- At personopplysninger som er feil blir rettet
- At personopplysninger som er lagret blir slettet
- å få en oversikt over lagrede personopplysninger på et lesbart elektronisk format

For virksomheten betyr dette at systemer og rutiner må på plass for å kunne ivareta disse rettighetene. Dette vil omfatte at man har rutiner både for å etterkomme forespørsler fra registrerte, og at man er proaktiv og sørger for at sletting skjer der det ikke lengre er behov for data. Det kan også påhvile virksomheten ansvar for å aktivt oppdatere data slik at disse er korrekte.

Hva som er tilstrekkelig vil være kontekst-avhengig. For en mindre virksomhet kan det være tilstrekkelig at man går gjennom manuelt og sletter ved forespørsel, mens større virksomheter trenger et mer automatisert system om de skal kunne oppfylle pliktene sine. Uansett skal virksomheten sørge for å kontrollere identiteten til den som forespør, og det skal være gratis å benytte seg av disse rettighetene.

Innebygd personvern

Personvernforordningens artikkel 25 angir krav til at personvern skal ta hensyn til når man utvikler nye systemer. Dette gjeldet i alle faser av utviklingen og man må tenke personvern i ting som standardinnstillinger, tilgangsstyring, skjemautforming, muligheter for sletting og så videre. Plikten omfatter også brukergrensesnittet og informasjonen ut til brukere.

Forsvarlig sikkerhet

Personvernforordningen stiller også krav til at personopplysninger skal være forsvarlig sikret slik at uvedkommende ikke kan få se eller endre data. I forordningens artikkel 39 heter det:

> «Personopplysninger bør behandles på en måte som gir tilstrekkelig sikkerhet og konfidensialitet, herunder for å hindre ulovlig tilgang til eller bruk av personopplysninger og utstyret som brukes i forbindelse med behandlingen."

Det er altså ikke spesifikke krav til sikkerheten, men det framgår at virksomhetene likevel er forpliktet til å ha sikkerhetstiltak. Datatilsynet anbefaler tiltak som sterk autentisering, regelmessige gjennomganger, kryptering og anonymisering.

Internkontroll/oppdatering

Virksomhetene plikter å ha et system for internkontroll, slik at man sikrer at virksomhetens plikter til enhver tid er oppfylt (personvernforordningens artikkel 24). Det gis ikke konkrete krav til hvordan internkontrollen skal være, kravet består i at man skal ha en systematisk tilnærming slik at de tiltakene og retningslinjene virksomheten har oppdateres ved behov og avvik blir fulgt opp.

Rapportere om avvik

Det vil alltid være en risiko for avvik. Virksomheten har plikt til å melde til datatilsynet så snart som mulig dersom det har vært et brudd på personopplysningssikkerheten. Et brudd på personopplysningssikkerheten defineres etter personvernforordningens artikkel 4 slik:

> *Et brudd på personopplysningssikkerheten, er et brudd på sikkerheten som fører til utilsiktet eller ulovlig tilintetgjøring, tap, endring, ulovlig spredning av eller tilgang til personopplysninger som er overført, lagret eller på annen måte behandlet.*

I tillegg til at datatilsynet skal varsles skal også de berørte personene bli informert dersom det er høy risiko for at deres opplysninger er på avveie. Hva som anses som høy risiko er det gitt retningslinjer for som man kan finne på datatilsynet sine nettsider (Datatilsynet, 2018).

Virksomheten skal snarest gjøre nødvendige tiltak for å få lukket avviket, og evaluere tiltak for å forhindre at det gjentar seg.

Hvordan kan en virksomhet gå fram for å sikre at de er i samsvar med personvernforordningen?

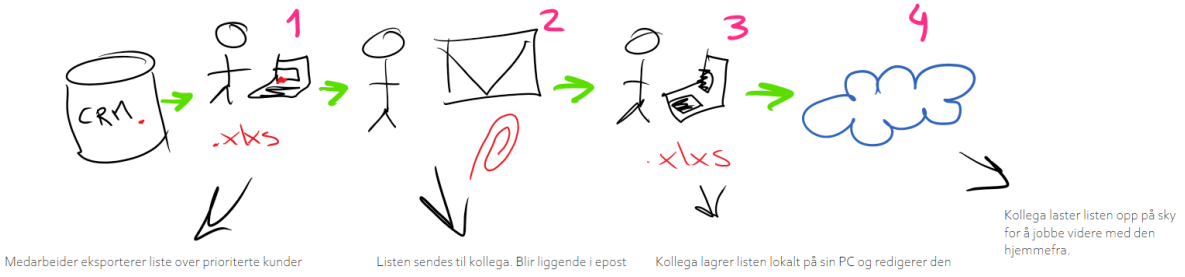
Å sikre samsvar med personvernforordningen er en prosess som vil pågå kontinuerlig i en virksomhet i og med at virksomheten og dens omgivelser vil være i stadig endring. Det vil likevel være naturlig at det kjøres et litt mer omfattende prosjekt til å begynne med for å skaffe en oversikt og sette gode rutiner for å opprettholde samsvar.

Kartlegging

Det første man blir anbefalt å gjøre for å sikre at man er i samsvar med personvernforordningen er å kartlegge hva virksomheten har av personopplysninger og hvordan de håndteres i dag. Dette gjelder like mye for de ustrukturerte dataene som for de strukturerte.

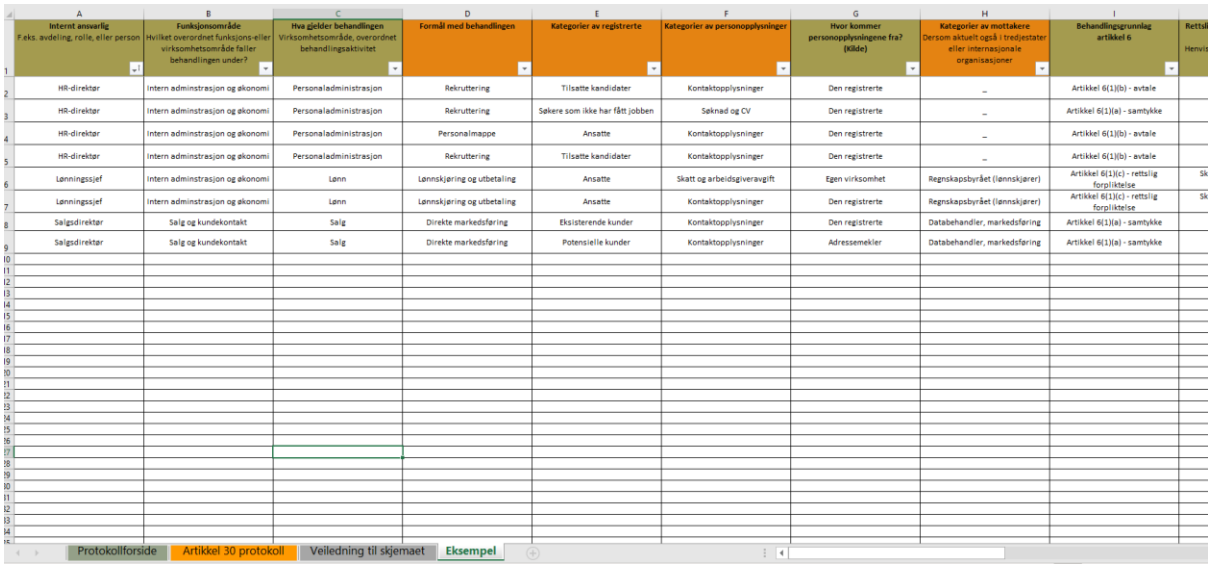
Når det gjelder de ustrukturerte kan det være vanskeligere å få oversikt, og det må gjerne litt mer detektivarbeid til. Her kan man se for seg ulike fremgangsmetoder. En metode kan være at man inviterer mellomlederne til et møte/workshop der man går gjennom hva ustrukturerte data er og ber de komme med innspill om hva de har. En annen er at man intervjuer viktige personer i de ulike avdelingene og forhører seg om hva de har av personopplysninger og hvordan de blir håndtert i dag. Det vil være forskjellig hva som passer best i hver enkelt virksomhet. Listen over vanlige former for ustrukturerte data kan brukes som støtte.

I kartleggingen anbefales det at man ikke bare ser på hvilke data som finnes, men også hvordan disse håndteres. Det er svært vanlig at data på strukturert form finner veien ut av sine databaser og ender opp som ustrukturerte data. Dette skjer typisk ved at for eksempel noen i salgsavdelingen eksporterer en liste over kunder til en excel-fil for å sende på epost til en kollega. Det er også vanlig at det settes opp periodiske eksporter fra CRM og ERP-systemer (Kuppinger, 2018). Dette gjør det svært vanskelig å ha kontroll på hvor slike data ender opp og hvordan de blir håndtert.



Figur 1 Klassisk flyt av personinformasjon i ustrukturert form.

For å organisere den informasjonen man finner i kartleggingen kan man bruke datatilsynets mal for behandlingsansvarliges protokoll (datatilsynet, 2018). Denne kan man laste ned gratis på datatilsynets sider. Malen er et skjema i excel-format og er enkel å fylle ut. I og med at personvernforordningen stiller krav til protokollføring kan man her slå to fluer i en smekk. Hvis man allerede har en slik protokoll kan denne fungere som et utgangspunkt.



Internt ansvarlig F.eks. avdeling, rolle, eller person	Funksjonsområde Hvilket overordnet funksjons-eller virksomhetsområde faller behandlingen under?	Hva gjelder behandlingen Virksomhetsområde, overordnet behandlingsaktivitet	Formål med behandlingen	Kategorier av registrerte	Kategorier av personopplysninger	Hvor kommer personopplysningene fra? (Kilde)	Kategorier av mottakere Dersom aktuelt også i tredjestater eller internasjonale organisasjoner	Behandlingsgrunnlag artikkel 6	Rettsl Henvis
HR-direktør	Intern adminstrasjon og økonomi	Personaladministrasjon	Rekruttering	Tilsatte kandidater	Kontaktopplysninger	Den registrerte	_	Artikkel 6(1)(b) - avtale	
HR-direktør	Intern adminstrasjon og økonomi	Personaladministrasjon	Rekruttering	Søkere som ikke har fått jobben	Søknad og CV	Den registrerte	_	Artikkel 6(1)(a) - samtykke	
HR-direktør	Intern adminstrasjon og økonomi	Personaladministrasjon	Personalmappe	Ansatte	Kontaktopplysninger	Den registrerte	_	Artikkel 6(1)(b) - avtale	
HR-direktør	Intern adminstrasjon og økonomi	Personaladministrasjon	Rekruttering	Tilsatte kandidater	Kontaktopplysninger	Den registrerte	_	Artikkel 6(1)(b) - avtale	
Lønningssjef	Intern adminstrasjon og økonomi	Lønn	Lønnskjøring og utbetaling	Ansatte	Skatt og arbeidsgiveravgift	Egen virksomhet	Regnskapsbyrået (lønnskjører)	Artikkel 6(1)(c) - rettslig forpliktelse	Sk
Lønningssjef	Intern adminstrasjon og økonomi	Lønn	Lønnskjøring og utbetaling	Ansatte	Kontaktopplysninger	Den registrerte	Regnskapsbyrået (lønnskjører)	Artikkel 6(1)(c) - rettslig forpliktelse	Sk
Salgsdirektør	Salg og kundekontakt	Salg	Direkte markedsføring	Eksisterende kunder	Kontaktopplysninger	Den registrerte	Databehandler, markedsføring	Artikkel 6(1)(a) - samtykke	
Salgsdirektør	Salg og kundekontakt	Salg	Direkte markedsføring	Potensielle kunder	Kontaktopplysninger	Adressemekler	Databehandler, markedsføring	Artikkel 6(1)(a) - samtykke	

Protokollforside | Artikkel 30 protokoll | Veiledning til skjemaet | Eksempel

Figur 2: skjermbilde fra datatilsynets mal for protokoll

Kartleggingen skal ikke bare avdekke hvilke personopplysninger som finnes, men også hvordan de oppbevares og behandles. Her vil det kunne komme underleverandører inn i bildet. Underleverandører/databehandlere er en egen kolonne i protokoll-malen, så det vil også avdekkes dersom man går gjennom denne. Dersom man avdekker at det brukes underleverandører må det gås gjennom at man har databehandler-avtaler med disse.

I mange tilfeller kan det være vanskelig å vite hva som finnes av personopplysninger. Dette gjelder spesielt i der data er lagret på tradisjonelt ikke-søkbart format, som tekst i pdf-er, dokumenter med håndskrift, bilder og lignende. Personvernforordningen omfatter alle personopplysninger, også disse, så det er viktig at de kommer med.

Uansett hvilken metode som velges vil det være avgjørende for arbeidets suksess at en person får ansvar for utførelsen og at det blir signalisert ut i organisasjonen at denne personen har ledelsens støtte og at dette er prioritert (Hjertø, 2018).

Vanlige former for ustrukturerte data

Digitale kilder

E-post

De fleste virksomheter benytter epost. Vanligvis har de to ulike former for epostkontoer: de som er felles som flere har tilgang til og personlige e-post kontoer. Slike e-post kontoer faller inn under virksomhetens ansvar.

E-post kontoer som flere har tilgang til er typisk adresser som er eksponert utover mot omverden, som adresser for generelle henvendelser, support, salg og så videre. Personlige epost-kontoer har adressen til en ansatt og vil kun være tilgjengelig for den enkelte ansatte. Disse adressene er det vanligvis bare de som har vært I kontakt med den ansatte av en eller annen grunn som har. Disse e-post kontoene har virksomheten normal ikke tilgang på. Av hensyn til den ansattes personvern skal ikke arbeidsgiver ha innsyn med mindre spesielle omstendigheter inntreffer (Lovdata, 2018). Dette kan være for eksempel at en arbeidstaker plutselig dør eller blir fraværende og det er nødvendig for driften at man går inn i eposten eller det kan være mistanke om grov korrupsjon eller lignende.

Elektroniske filer

De fleste virksomheter har en hel rekke filer liggende. Dette vil være i mange formater, som bilder, tekstfiler og regneark. Noen filer vil ligge på fellesområder som alle i en avdeling eller hele bedriften har tilgang til, mens andre vil ligge I hjemmeområder der kun en ansatt har tilgang.

Levende media

Noen virksomheter kan ha både lyd- og bildefiler liggende. Noen eksempler på dette er opptak av telefonsamtaler, video fra overvåkningsutstyr og små filmsnutter brukt I intern eller ekstern markedsføring. Disse er i samme grad som skriftlige kilder og bilder underlagt personvernforordningen.

Analoge kilder

Dokumenter I papirformat

Mange dokumenter finnes kun i fysisk format og kan bli liggende i lang tid i arkiver eller mindre organisert på diverse skrivebord og i skuffer rundt om i bedriften. De samme reglene gjelder for personopplysninger her som for opplysninger lagret i elektronisk format, og det kreves at virksomheten har oversikt. Slike dokumenter inneholder ofte personopplysninger slik som navn på kvitteringer, bilder av folk og møtereferat, eller enda mer detaljert som er tilfellet når søknader og relaterte dokumenter er innlevert eller skrevet ut på papir. Virksomheten skal slette disse når det ikke lengre er behov for å oppbevare de, og det skal foreligge en rutine for dette.

Andre analoge media

På samme måte som med dokumenter i papirformat kan en virksomhet ha liggende lyd- og videoopptak på ulike fysiske medier. Opptak fra overvåkningskameraer skal man være ekstra påpasselig med oppbevaring av, datatilsynet har laget en egen veileder for hva som er tillat og ikke på dette området (Datatilsynet, 2016). Andre analoge media kan være filmer fra fester, lanseringer og lignende feiringer en bedrift har hatt.

Maskinlæring som verktøy

Det finnes verktøy basert på maskinlæring som kan være til hjelp i arbeidet med å håndtere ustrukturerte data. Med slike verktøy kan man finne personopplysninger i det som ellers ville ha vært ikke søkbare filer, som skannede kvitteringer eller dokumenter med håndskrift, filer på pdf-format, bilder med tekst på og lydfiler.

Det finnes selskaper som spesialiserer seg på å tilpasse maskinlærings-teknologi til å kaffe oversikt over personopplysninger. De beste systemene greier å skanne gjennom dokumenter og identifisere tagge alle forekomster av en type data, som for eksempel alle navn på personer eller alle IP-adresser. Dette kan kombineres med sikkerhetsmekanismer, for eksempel kan man sette kryptering på alle filer der det finnes et kredittkort-nummer. Et eksempel på dette er Microsoft Azure Information Protection (MSIP). SailPoint er en annen aktør som spesialiserer seg i å sette tilganger til forskjellige roller av brukere basert på kategorier (Kuppinger, 2018). I tillegg finnes løsninger som analyserer bilder og produserer søkbar tekst. Eksempel på detter er IBMs Watson og Microsoft Cognitive services.

Det som vil være en fallgruve her er at samtidig som man får oversikt, også skaper mer data med personinformasjon. Alle taggene på dokumentene vil utgjøre enda en registrering av personinformasjonen, og må også behandles i henhold til GDPR. Her blir det viktig å ha et bevisst forhold til hvordan man håndterer slike metadata.

GAP-analyse

Når man har fått oversikt over hvilke personopplysninger man har i dag er neste steg å gjøre en gap-analyse som viser avvikene mellom dagens håndtering av personopplysninger i ustrukturerte data og samsvar med personvernforordningen.

For å finne ut hva avvikene er kan man starte med å se på om man har behandlingsgrunnlag. Her er reglene helt klare: har man ikke behandlingsgrunnlag kan man heller ikke oppbevare opplysningene og de skal slettes. Datatilsynet har laget en veileder om behandlingsgrunnlag som kan brukes i dette arbeidet. Denne finner man på datatilsynet sine nettsider tilgjengelig også på utskriftsvennlig format (Datatilsynet, 2018) .

Lage personvernerklæring

For å oppfylle de kravene som stilles til informasjon kan det være hensiktsmessig å utforme en personvernerklæring. Man kan finne mye på nettet, men man må huske på kravene til at denne skal være lettfattelig og grei – og at den skal tilgjengeliggjøres separat fra andre bestemmelser og dokumentasjon.

Mulige tiltak

Rutiner for fysisk sikring

Dersom man i kartleggingen oppdager at data er oppbevart på måter som innebærer risiko for at personopplysninger kommer på avveie må det iverksettes tiltak for fysisk sikring. Dette kan være enkle ting som for eksempel låsing av arkivskap, låser på skuffeseksjoner og oppbevaringsmøbler på kontorer og ulike sektorer i kontorlandskapet med ulik tilgang.

En gjennomgang kan dessverre ofte resultere i at man finner ut at det personopplysninger lagret i permer som står fritt tilgjengelig på hyller rundt om, i kontorer som står åpne eller i landskap. En annen typisk feil er at papirer med personopplysninger blir leggende uavhentet på printer. Her finnes det mange løsninger med at ansatte må taste en kode eller bruke nøkkelkort med RFID for at utskriften skal starte.

Alle som kan ha befatning med personopplysninger bør bevisstgjøres dette og det vil være fordelaktig om det utarbeides klare rutiner for hvordan papirer og utskrifter skal håndteres.

Rutinemessig sletting av tilganger og data ved avslutning av stilling

Manuelle rutiner

I sin enkleste form kan en manuell rutine gå ut på at en administrator fjerner tilgangene til en ansatt i det de slutter samtidig som hjemmeområde, mailboks og annet relevant innhold slettes. Dersom dette skal ivaretas med manuelle rutiner fordrer det et godt samarbeid mellom IT, lønn- og personal og mellomledere. Mellomledere må gi beskjed når en ansatt sier opp eller bytter stilling, både til lønn- og personal og til IT-avdelingen, eventuelt må ansvaret for å varsle IT-avdelingen tilfalle HR-avdelingen. Det må i begge tilfeller tegnes opp helt klare rutiner og ansvaret må plasseres tydelig for å sikre seg at avslutninger blir fulgt opp skikkelig. Det vil ikke være i henhold til personopplysningsforordningen å oppbevare filer og epost-arkiv fra personell som ikke lengre jobber i virksomheten.

Automatiske og halvautomatiske rutiner

Man kan kjøre sletting ved manuelle rutiner, men det er antageligvis tryggere å sette opp løsninger som automatisk utfører denne jobben for å sikre at ikke det blir glemt. Det er ulike tilbyder på

markedet som lager alt fra enkle løsninger med sjekklister som sendes til ledere og IT-ansvarlig til mer avanserte løsninger.

Automatisk sletting av filer basert på utløpsdato

Ved hjelp av Powershell og Task Scheduler kan man selv uten store budsjetter sette opp slik at filer i en mappe slettes hvis de ikke er endret etter et visst antall dager. Dette er rimelig enkelt og kan settes opp selv om man ikke har de helt store programmeringskunnskapene. Det vil fungere både på Windows Server og på klientene.

Det vil for de fleste ikke være aktuelt å sette på en slik task på alle mappene, men det kan være nyttig for å hindre at det ligger gamle ting i nedlastings-mappen for eksempel. I alle tilfeller bør man gå over søppel-folderen og sikre at det er automatisk sletting på den.

Automatisk sletting av e-post basert på utløpsdato

E-post klienten inneholder ofte mye data man har liten oversikt over, både i e-poster og som vedlegg. Det kan derfor være nyttig å sette opp sletting av mail som er eldre enn for eksempel 30 dager. De fleste e-post klienter har muligheter for dette, i Outlook er det auto-archive funksjonen som vil brukes til dette formålet.

Det vil nok ikke være ønskelig for de fleste å slette all e-post etter en viss dato, men man kan innføre rutiner på at man arkiverer alt som skal tas vare på og auto-sletter det som blir liggende i innboksen. På denne måten tvinger man fram en mer bevisst holdning til hva man trenger å beholde og ikke. Det er også anbefalt å gå over innstillingene for sletting av området for slettede elementer, så man er sikker på at også disse blir slettet skikkelig etter en viss tid.

Fullautomatisert livssyklus-håndtering av brukere (IDM-system)

For større virksomheter kan det være aktuelt med et system for livssyklus-håndtering av brukere.

Et avansert IDM-system (Identity Management System) vil hver dag hente uttrekk fra HRM-systemet (Human Resource Management). Det er her alle ansatte registreres når de starter, slutter eller endrer stilling. Å knytte tilganger, grupper og brukerkontoer opp mot HRM er svært nyttig fordi det reduserer sjansen for menneskelige feil (data punches ett sted) og man knytter tilgangene opp mot det mest oppdaterte systemet. I og med at lønn og kostnader knyttes direkte til avdeling og stillings-status i HRM-systemet vil dette normalt være mer oppdatert enn de fleste andre systemer en virksomhet har.

Når IDM-systemet finner en ny ansatt i uttrekket vil det se på hvilken stillingstype personen har, og dersom det er en type stilling som skal ha databruker vil det opprette AD-konto, e-postkonto og et hjemmeområde som blir aktivert på angitt startdato. De mest avanserte IDM—systemene har også integrasjoner mot fagsystemer som for eksempel saksbehandlingssystem slik at tilgang opprettes også her for de stillingstypene som skal ha slik adgang. Dette kan også kobles mot systemer for fysiske tilganger gjennom API mot nøkkelkort-systemer.

De beste systemene holder også orden på om en bruker skifter stilling, og avslutter tilgangene på tidligere stilling og innvilger tilganger som er relevante for ny stilling ved skifte. Dette vil hindre at en ansatt som har jobbet i en stor organisasjon i mange år har opparbeidet seg mange tilganger til ulike systemer i løpet av årene som egentlig skulle ha vært avsluttet (James A Martin, 2018).

IDM-systemet vil plukke opp når det er satt slutt-dato eller at en ansatt ikke lengre finnes i uttrekket. Da legges brukeren i karantene, og AD-brukeren med tilhørende stillinger settes til inaktiv så den ikke har tilgang på systemer og e-post lengre. Dersom det har vært en feil i HRM-systemet vil dette

raskt avdekkes da den ansatte ikke kommer seg på e-post eller noen systemer, og brukeren kan enkelt hentes tilbake. Etter en periode (vanligvis 30 dager) vil hjemmeområdene og e-postboksen til brukeren slettes permanent.

Disse systemene er ganske kostbare og vanligvis ikke aktuelle for små og mellomstore bedrifter. De fleste som benytter slike systemer er store bedrifter med mange ansatte og virksomheter som skoler og universitet som har stor gjennomstrømming av brukere som skal opprettes og avsluttes. En hyggelig bieffekt av et slikt system er at det også bidrar til å holde lisenskostnader i sjakk.

Rutine for avhending av utstyr og papirer

Å rutinemessig kvitte seg med data og papirer med personopplysninger man ikke har grunn til å holde på er en naturlig konsekvens av personvernforordningen. Man må ikke glemme at denne avhendingen bør skje på en fornuftig måte. Som hovedregel bør alle papirer som kan inneholde personopplysninger makuleres. Når det gjelder lagringsmedier som USB-minnepinner, harddisker, CD-rom og så videre kan være litt mer komplisert, og norSiS foreslår følgende metoder (norSIS, 2019):

- bruke programvare som skriver over data slik at den ikke kan hentes fram
- avmagnetisering. Dette vil kun fungere på tradisjonelle harddisker, disketter, magnettaper og andre magnetiske lagringsmedier.
- Fysisk destruksjon som knusing slik at det blir svært vanskelig å rekonstruere lagringsmediet.
- En kombinasjon av de tre metodene som er nevnt.

Samskriving – å dele lenker i stedet for filer

I dag er det både billig og enkelt å benytte seg av ulike samhandlingsplattformer som gjør at ansatte kan samarbeide på et dokument uten at det må sendes fram og tilbake. Det finnes både gratis produkter som kan tas i bruk av små virksomheter, og mer avanserte og kostbare løsninger for større virksomheter. Å samskrive på sentralt lokaliserte dokumenter vil føre til en bedre kontroll på hvor filer med personopplysninger befinner seg, og det er ofte også mer effektivt for brukerne. Det gir også mulighet til å trekke tilbake tilgang til filer eller områder med umiddelbar virkning.

Noen eksempler på slike systemer er:

- **Teams**: et relativt nytt produkt fra Microsoft som samler felles filområder, chattefunksjoner, videomøter og en rekke andre funksjoner i ett produkt. Teams fungerer som en hub for Office 365 og bygger på sharepoint. Finnes i gratisversjon med basisfunksjoner og som abonnementsprodukt med pris per bruker for en versjon med mer funksjonalitet.
- **Google**: en kombinasjon av googles fildelingstjeneste Google Drive og deres web-applikasjoner Sheets, Docs og slides kan brukes gratis for små volum, eller som betalingstjeneste ved større lagringsbehov.
- **Bitrix24**: full-service tjeneste med CRM, prosjektstyringsverktøy, timeregistrering, chat, videomøter og wiki-løsning i tillegg til fildeling med samskrivingsfunksjonalitet. Leveres både som on-premise (kunden kjøper programvaren og installerer på sine servere) og SaaS (kunden kjøper et abonnement på programvareløsningen med tilgang til en instans av løsningen i sky). Gratis for opptil 12 brukere.

Tjenestene er generelt enkle å ta i bruk. Den største vanskeligheten med dette tiltaket vil være at man må endre folk sine vaner. Når folk er vant til å sende filer i stedet for lenker på e-post kan det være en tung jobb å endre på dette. Det er derfor viktig at ledelsen tar i bruk disse verktøyene aktivt og at implementeringen blir med i kulturprogrammet.

Kjøper man tjenestene levert på sky vil det også være viktig å forsikre seg om at leverandøren har en databehandleravtale som er i henhold til regelverket.

GDPR og sikkerhetskultur i sammenheng

GDPR og informasjonssikkerhet er tett knyttet (informasjonssikkerhet og kontinuerlig arbeid er spesifikt nevnt i personvernforordningen), og det blir derfor naturlig å innlemme bevissthet om personvernforordningen i arbeidet med sikkerhetskultur.

IBM har beregnet at så mye som 95% av uønskede hendelser i forbindelse med informasjonssikkerhet hadde menneskelig feil som årsak eller medvirkende faktor (IBM, 2014). Det er altså ingen grunn til at arbeidet med å bevisstgjøre og skape god kultur rundt sikkerhet og personvern skal få mindre plass enn de tekniske tiltakene.

For å lykkes med å skape en varig sikkerhetskultur holder det ikke med et kurs eller en felles epost som sier at fra nå av må alle passe på. Man må involvere alle ledd i virksomheten i et kontinuerlig arbeid. Spesielt viktig er det at ledelsen forstår at dette er arbeid som er av høyeste viktighet for virksomheten, og at de støtter arbeidet på en synlig måte ut til alle involverte (Hjertø, 2018).

Mange som jobber med sikkerhet har uttrykt bekymring for at arbeidet med å være i samsvar med personvernforordningen vil forverre forholdene for arbeidet med tradisjonell informasjonssikkerhet. Dette begrunnes med tre ting i hovedsak (Thales eSecurity, 2017):

- Personvernforordningen er kompleks og virksomhetene er usikre på hvordan de skal håndtere arbeidet med personvern. Det er en utbredt oppfatning at personvernforordningen har ført til økt kompleksitet i sikkerhetsarbeidet.
- Personvernforordningen skaper distraksjon som trekker oppmerksomheten til sikkerhetsteamet over på mindre viktige områder, som kan føre til at sårbarheter blir liggende åpne.
- De store bøtene man kan få for brudd på sikkerheten som medfører personopplysninger på avveie gjør at det blir mer attraktivt for hackere som driver med utpressing å gjøre datainnbrudd.

Periodiske gjennomganger av virksomhetens personvern

Det kan anbefales å ha periodiske gjennomganger av personvernet i en virksomhet. I likhet med sikkerhetsarbeid er arbeidet med personvern en kontinuerlig innsats. Hvor ofte det er behov for gjennomgang vil avhenge av hvilken type virksomhet det er og virksomhetens omfang. Slike gjennomganger vil bidra til å oppfylle kravet om internkontroll.

Det anbefales å kalle inn til gjennomgang med de nøkkelpersoner og gå gjennom:

- protokollen
- om noe i virksomheten har endret seg siden sist gjennomgang
- nye avgjørelser/dommer. Lovtolkningen kan endre seg.
- nye ting man må ta hensyn til i omgivelsene, for eksempel nye former for trusler mot informasjonssikkerheten
- avvik

I disse gjennomgangene kan en også drøfte tiltak for holdningsskapende arbeid sett i lys av hva som er kommet fram i gjennomgangen. Tiltak kan for eksempel være nyhetsbrev, foredrag, møter, oppslag på intranett eller interne konkurranser. Hva avvikene består i kan være en god pekepinn på hvor det er nyttig å rette innsatsen.

For å følge med på nyheter av interesse for dette arbeidet kan det være smart for alle som er involvert å abonnere på nyhetsbrevene til norSiS og datatilsynet. Dette er gratis, og man får god og oppdatert informasjon direkte i sin epost-boks. I tillegg er nettsiden til European Data Protection Board en god ressurs, her publiseres nye retningslinjer og presiseringer angående hvordan bestemmelsene i personvernforordningen skal tolkes. Hvis man er spesielt interessert kan man også følge med på datatilsynets blogg: personvernbloggen.no.

Forsøk – JBS Borettslag og Azure CS

Det praktiske arbeidet med denne oppgaven består av to deler: å prøve ut om maskinlæringsteknologi kan benyttes til å få oversikt over ustrukturerte data også for små og mellomstore virksomheter, og en veileder/oversikt over tiltak. Kommende kapittel beskriver den første delen av oppgaven: testa av maskinlæringsteknologi.

For denne delen av oppgaven ble det behandlet personopplysninger, og det måtte derfor søkes til Norsk Senter for Forskningsdata (NSD) om godkjenning for behandlingen. NTNU hadde levert en felles søknad for årets bacheloroppgaver, men denne kunne ikke benyttes i dette tilfellet da det ikke er mulig å hente inn samtykke fra de registrerte. Det ble derfor sendt inn et eget meldeskjema, og søknaden ble godkjent etter en skjønnsvurdering av at fordelene var større enn ulempene. I vedlegg 1 finnes meldeskjema og vurdering av dette fra NSD.

JBS Borettslag

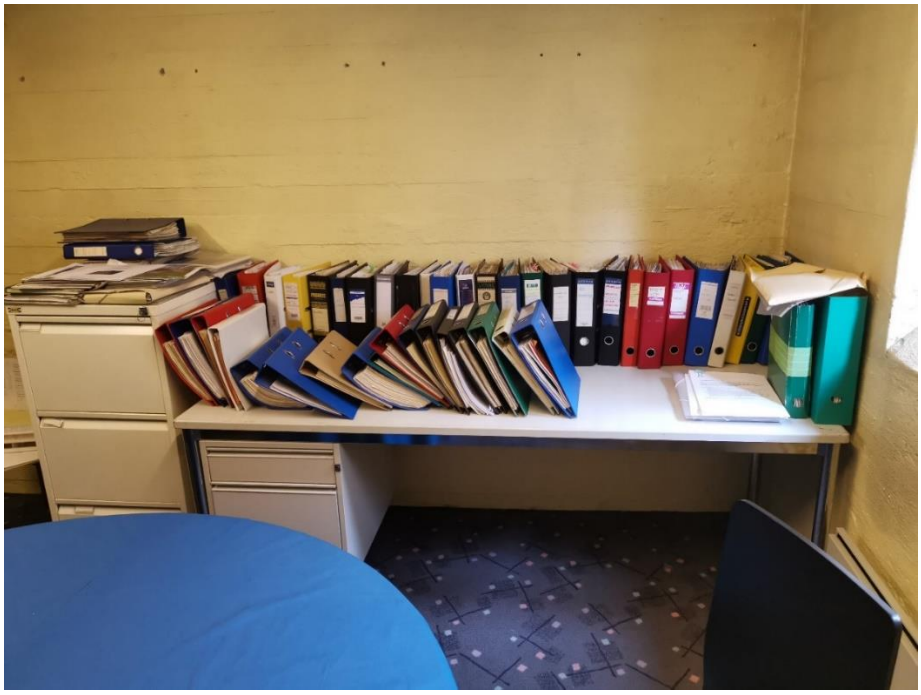
JBS Borettslag er et borettslag på Tøyen i Oslo opprettet i 1933 (JBS Borettslag, 2013). Borettslaget har 79 leiligheter og har et styre på 5 personer som blir valgt av andelseierne i generalforsamling. Medlemmene i styret er vanligvis andelseiere i borettslaget som ikke kan forventes å ha utstrakt kompetanse om personvernforordningen. De fleste leilighetene i borettslaget er ett- og toroms leiligheter. Med en slik sammensetning av leiligheter og sentral beliggenhet bor det mange førstegangskjøpere der som etter en tid flytter ut når de vil ha noe større. Dette betyr at beboere og dermed også styre blir byttet ut relativt ofte.

Som alle virksomheter forplikter borettslaget å oppfylle bestemmelsene i personvernforordningen. Men med et styre som skiftes ut relativt ofte og begrenset budsjett er det ikke rom for å bruke konsulenter for å få foretatt kartlegging og sikre at man er i samsvar med personvernforordningen.

Dagens situasjon for borettslaget

Borettslaget benytter seg av OBOS som forretningsfører og har på denne måten god oversikt over de strukturerte dataene som oppbevares. OBOS går god for at deres systemer ivaretar de registrertes rettigheter og at sikkerheten er god nok.

Når det gjelder ustrukturerte data finnes disse i hovedsak som eposter på styreleders adresse og filer på tilhørende g-drive (lagringsplass på sky). Disse dataene er i søkbart format, og det er lett å finne opplysninger om en person dersom noen vil ha sin informasjon slettet. I tillegg til dette finnes det et gammelt arkiv på borettslagets styrerom. Her er det alle typer papirer datert tilbake til borettslagets stiftelsesår i 1930. E-post konto og tilhørende g-drive går i arv fra styreleder til styreleder.



Figur 3: styrerommet hos borettslaget

Man kunne ha valgt å makulere alt av eldre dato, men dette er risikabelt i og med at det kan finnes materiale man trenger senere til dokumentasjonsformål. Dette kan være tegninger over tekniske installasjoner, samsvarserklæringer, søknader til plan- og bygningsetaten, overdragelser og så videre. Det er derfor et behov for å få kategorisert innholdet i disse permene slik at man kan få slettet det borettslaget ikke har god grunn til å beholde. I tillegg vil det være en fordel å få en viss oversikt over innholdet.

Fremgangsmetode for forsøket

På grunn av den dynamiske prismodellen er systematisering med Azure Cognitive Search innenfor det borettslaget kan ta seg råd til. Som en test av å bruke AI som verktøy for å få kontroll på ustrukturerte data ble forsøket å skanne inn et sett med dokumenter fra styrerommet og indeksere disse i Azure Cognitive Search sammen andre filtyper for å teste om det er mulig få alle data inkludert eposter og bilder inn under samme løsning. Etter indekseringen vil det bli gjort forsøk med forskjellige typer søk for å se om løsningen fungerer i situasjoner man kan se for seg blir nyttige:

- å bruke søkefunksjonalitet som finner alle registreringer av personer
- å bruke søkefunksjonalitet som finner alle organisasjoner
- å søke på en enkelt person og se om man finner alle oppføringer av denne personen i datasettet
- å gjøre tekst fra bilder og dokumenter søkbar

Målet er å se om dette vil være en hensiktsmessig fremgangsmetode for små og mellomstore bedrifter som har samme type utfordringer som JBS Borettslag. Dermed blir det nødvendig å vurdere resultatene ikke kun etter om det lar seg gjøre, men også hvor komplisert/vanskelig prosessen var og om det er en gjennomførbar metode for en virksomhet med begrensede ressurser og IT-kompetanse.

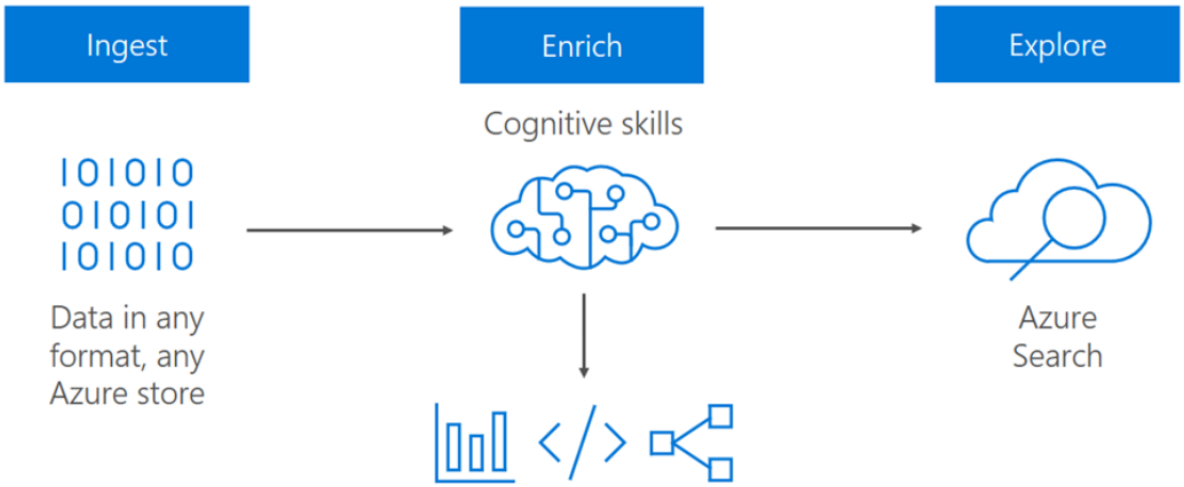
Om Cognitive search

I mai 2018 lanserte Microsoft produktet Cognitive search på sin sky-plattform Azure (Microsoft, 2018). Dette er et sammensatt produkt som kan gjøre det mulig for små og mellomstore bedrifter til å kunne bruke kunstig intelligens til å kartlegge og få kontroll på ustrukturerte data.

Teknologi

Cognitive search er satt sammen av flere ulike produkter som finnes under Azure-paraplyen. I seg selv er ikke produktet revolusjonerende, men det som er nytt er at man uten å ha stor kunnskap om søketeknologi og maskinlæring kan sette opp et slikt system til en overkommelig pris for en mindre bedrift.

Cognitive search bygger på tre prosesser som settes sammen:



Innta: å få matet data inn i Azure. Hvis man skal ha mulighet til å søke effektivt i dataene bør de befinne seg i skyen. Man kan dytte data opp i skyen på forskjellige måter. For ustrukturerte data er Azure Blob Storage det som vil fungere, men man kan også kombinere dette med strukturerte data for eksempel i CosmosDB, MySQL eller SQL DB.

Berike: dette er den spennende delen der maskinlæringsalgoritmer og kraftfull indeksering kommer inn i bildet. Tjenesten klarer å trekke ut tekst fra bilder og skannede dokumenter, til og med når teksten er i form av håndskrift. Man kan også legge til tjenester som går gjennom bilder og merker de på bakgrunn av hva som er i bildene. Noen av de tilgjengelige teknologiene er:

Named Entities Recognition (NER): en underform av språkteknologi/datalingvistikk som går ut på å bruke maskinlæring til å gjenkjenne kategorier av ord. Dette kan brukes til å sette merker på alle forekomster innen en kategori, som for eksempel alle firmanavn eller alle stedsnavn. Denne typen jobb kan virke enkel, men den kan være vanskelig for en maskin. Man kunne satt et menneske som nettopp har lært seg til å lese til oppgaven «strek under alle navn på personer du ser i denne teksten», men å programmere en maskin til å gjøre denne jobben er ganske komplisert. Selv om dette ikke er enkel teknologi har den eksistert i lang tid. De beste systemene klarte allerede i 1998 å produsere resultater med 93% treffsikkerhet for tekster på engelsk (Marsh, 1998).

Optical Character Recognition (OCR): er en teknologi som går ut på å tolke tekst ut fra bilder – enten fra skannede dokumenter med trykket tekst, eller fra bilder med håndskrift eller tekst på. Teknologien brukes i tilfeller der informasjon skal være lesbar både for mennesker og maskiner, og man ikke ha sekundære inndata-kilder (Eikivil, 1993). Teknologien kan anvendes på utallige områder.

Ett eksempel er bomstasjoner som fotograferer nummerskiltene på passerende biler og tolker bilnummeret for å vite hvem som skal faktureres for passeringen. Et annet er at man kan oversette skrift fra bilder til punktskrift. Dette er nyttig for personer med synshemming.

Key Phrase Extraction: funksjonalitet som går ut på å trekke ut hovedpoengene fra en tekst I ustrukturert form (Microsoft, 2019). Hos Azure er det ikke er støtte for dette på norsk enda.

Languange detection: Teknologi som finner ut hvilket språk en tekst er skrevet på. Mange sluttbrukere er godt kjent med denne typen teknologi fra for eksempel søkemotorer eller tekstbehandlere som bruker den til å automatisk korrigere skrivefeil.

Image analysis: funksjon i Azure som kaller på metoder som analyserer bilder og trekker ut relevant informasjon. Dette kan være for eksempel «finn alle bilder med en hund i» eller ansiktsgjenkjenning av mennesker.

Uforske: resultatet av berikings-prosessen blir satt i en Azure search-index. Det vil si at man etter denne prosessen har en systematisert data-store som gjør det mulig å raskt trekke ut den informasjonen man ser etter.

Prising

Azure har en prismodell som bygger på pay-as-you go prinsippet, det vil si at man ikke betaler for mer enn man bruker. De ulike tjenestene har hver sin pris som gjør at man betaler for lagring, søkefunksjonalitet, web-servicer og så videre hver for seg. Fordelen med dette er at borettslaget ikke må forplikte seg til en kostbar og omfattende løsning for å kunne benytte seg av den funksjonaliteten som kan være tilgjengelig. Ulempen er at det kan være vanskelig å få oversikt og forutse kostnadene.

Personvern i løsningen

Microsoft tar på seg ansvaret for sikkerheten i sine løsninger, men kunden er fortsatt ansvarlig for bruken av løsningene. For eksempel garanterer Microsoft for sikkerheten i sin løsning for sterk autentisering, men det er kundens ansvar skru på sånn at det kreves sterk autentisering for tilgang til sine data.

Azure har de fleste sertifiseringer som garanterer for kvaliteten i løsningen som ISO 27001, og de tilgjengeliggjør revisjonsrapportene på sertifiseringen på sine nettsider (Microsoft Trust Portal, 2019).

Gjennomføring av forsøket

Datasett

Det ble tatt ut tre tilfeldige ringpermer fra styrerommet til bruk som datasett i forsøket, alle fra perioden 1980 - 2000.

I dette forsøket ble det benyttet en effektiv skanner som ble lånt fra Deichmanske bibliotek. Dette gjorde jobben lettere, men det var likevel et relativt stort arbeid å få skannet inn de utvalgte permene med dokumenter. Stifter måtte fjernes og mange papirer var ikke på standard format så de måtte mates inn i skanneren manuelt en for en. Det vil anslagsvis ta 54 timer[1] med dugnadsinnsats fra styret hvis man vil ønske å skanne alt inn.

[1] Det er 96 permer på styrerommet. Det er tatt utgangspunkt i 0,5 time per perm + 6 timer for å frakte alt fram og tilbake.

Underveis i skanningen kom det fram at de permene som var utvalgt passet godt til formålet. Det var håndskrevne papirer, maskinskrevne papirer, bilder og tabeller i materialet.

I tillegg til papirer som ble scannet til pdf-format ble det også lagt inn noen outlook-elementer (e-poster) og bilder (.jpg) i datasettet for å teste på.

Oppsett av Azure Cognitive Search

Det første som ble gjort var bestilling av services i Azure portalen. Man starter med å bestille Azure search. Dette er enkelt og krever lite av brukeren, men man må ta stilling til en ganske kompleks prismodell underveis. Azure opererer ikke med pakkepriser, så man må se på både antall dokumenter man ønsker å søke på, og på størrelsen på lagringsplassen for å finne total pris. En annen ting man må være påpasselig med er at søketjenesten man bestiller er i et område som støtter Cognitive Search. Når man senere skal legge til AI-funksjonalitet i søkemotoren man lager her må den ligge på samme geografi innen Azure sky-systemet.

Det neste steget var å dytte dataene man vil søke på opp i skyen. For cognitive search brukes Blob Storage, en lagringstype som kan holde på de fleste filtyper. Dette er en jobb som kan gjøres enten via konsoll eller i det grafiske brukergrensesnittet i Azure portal. Når man bestiller servicen må man ta stilling til en del ting, som for eksempel hvor hurtig tilgjengelig lagringen skal være og hvor mye redundans man ønsker priset inn. Man må også velge de riktige innstillingene for tilgjengelighet slik at søkefunksjonen skal kunne få tak i dataene samtidig som de er utilgjengelige for omverdenen.

Etter å ha registrert dataene må man knytte søketjenesten til datalageret og behandle dataene med de ønskede tjenestene. I dette tilfellet ble det valgt OCR og gjenkjenning på navn (people) og organisasjoner (organizations). Dette utvalget utgjør et *Skillset* som igjen knyttes til en indexer som traverserer gjennom datasettet og merker dataene med hva den finner.

Figur 4: skjermbilde fra Azure-portal av detaljer om søketjenesten

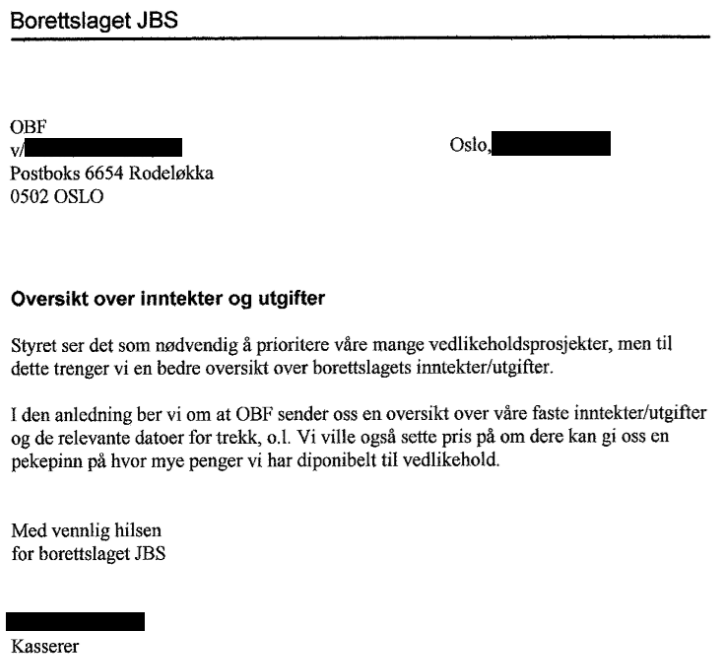
Når man har satt opp en søke-service på denne måten og kjørt indekseren får man en url som kan brukes mot Cognitive services sammen med kall fra tilhørende REST API. Man kjøre kall fra konsoll, vi a search explorer i portalen eller man kan bygge inn funksjonaliteten i en egen web-applikasjon. Resultatene av søk kommer tilbake som JSON dersom man ikke angir noe annet.

Utfall av forsøket med cognitive search

Det ble gjort forskjellige typer søk for å avdekke om Cognitive services slik det tilbys i dag er en farbar vei for små bedrifter som JBS-borettslag for å få oversikt over ustrukturerte data.

Forsøk med digitalisering av maskinskrevne sider

Det ble gjort forsøk på søk i materialet med utgangspunktet i navn på nøkkelpersoner i perioden. For eksempel ble det gjort søk på navnet til forretningsfører. Følgende dokument er et eksempel:



Borettslaget JBS

OBF
v/[redacted]
Postboks 6654 Rodeløkka
0502 OSLO

Oslo,[redacted]

Oversikt over inntekter og utgifter

Styret ser det som nødvendig å prioritere våre mange vedlikeholdsprosjekter, men til dette trenger vi en bedre oversikt over borettslagets inntekter/utgifter.

I den anledning ber vi om at OBF sender oss en oversikt over våre faste inntekter/utgifter og de relevante datoer for trekk, o.l. Vi ville også sette pris på om dere kan gi oss en pekepinn på hvor mye penger vi har diponibelt til vedlikehold.

Med vennlig hilsen
for borettslaget JBS

[redacted]
Kasserer

Figur 5: scannet dokument

Dokumentet ligger i datasettet som en pdf-fil skannet fra en av permene funnet på styrerommet.

Når man søker på navnet til mottaker av brevet får man opp dokumentet og følgende tolkning:


```
"merged_content": "\n[image: image0.tif] Borettslaget JBS OBF v/[redacted] Oslo,[redacted] Postboks
6654 Rodelokka 0502 OSLO Oversikt over inntekter og utgifter Styret ser det som nodvendig a prioritere vare mange
vedlikeholdsprosjekter, men til dette trenger vi en bedre oversikt over borettslagets inntekter/utgifter. I den anledning
ber vi om at OBF sender oss en oversikt over vare faste inntekter/utgifter og de relevante datoer for trekk, o.1. Vi
ville ogsa sette pris pa om dere kan gi oss en pekepinn pa hvor mye penger vi har diponibelt til vedlikehold. Med vennlig
hilsen for borettslaget JBS[redacted]Kasserer Borettslaget Jens Bjelkesgt. 43/Sorligt. 10, Jens Bjelkesgt. 43,
0578 Oslo \n\n\n",
```


Figur 6 Resultat av Cognitive Services tolkning av dokumentet

Man får altså en søkbar og ganske presis tekst på digitalt format ut av behandlingen.

Forsøk med å trekke ut alle personnavn i datasettet

Først ble det gjort forsøk på å trekke ut en liste over personer av datamaterialet. Det ble raskt klart at dette var en oppgave som AI-en ikke helt hadde kontroll på. Følgende resultat kom ut av en fil som inneholdt en plan for beplantning etter oppussing av bakgården:

```
"aHR0cHM6Ly9oZW5yaWsxMi5ibG9iLmNvcmUud2luZG93cy5uZXQvaGVucmlrMTIvUGxhbnRlbGlzdGUlMjBKQkwlMjAlMjBib3JldHRzbGFnLnBkZg2",
        "people": [
            "Prunus",
            "Syringa",
            "Syringa v.",
            "Philadelphus",
            "Ribes",
            "Hinnomaen Keltainen",
            "Aronia",
            "Kolkwitzia",
            "Buddeleja",
            "Spiraea",
            "Rosa",
            "Partenocissus t.",
            "Lonicera",
            "Aristolochia",
            "Petiolaris",
            "Rosa helena",
            "Hekker",
            "Gul",
            "Nepeta",
            "Mia Ruys'      Hvit",
            "Kransveronika",
            "Hemorocallis",
            "Perry"
        ],
```


Figur 7: utsnitt av søkeresultat

Det er grunn til å tro at dette problemet har sammenheng med språk. Å gjøre tilsvarende forsøk mot engelske navn i en database med dokumenter på engelsk ble gjort ved å laste ned en kopi av Azure-teamets applikasjon som kjører tilsvarende søk på et bibliotek med data fra JFK-etterforskningen. Denne ligger tilgjengelig på GitHub (Microsoft, 2018) og som web-applikasjon (Azure Team, 2018). I denne var det ikke noe problem å få ut en fornuftig liste over navn fra et dokument. Dette på tross av at det skal være støtte for Norsk og at det er meningen at Cognitiv search skal klare å identifisere hvilket språk teksten er på (Microsoft Azure, 2019).

Forsøk på å trekke ut alle navn på organisasjoner

Tilsvarende ble det gjort søk på organisasjoner i samme dokument. Det som er interessant er at CS returnerer mange opplistinger som ikke er organisasjoner, samtidig som de organisasjonene som var nevnt i dette dokumentet ikke dukker opp (JBS Borettslag, arkitektfirmaet og landskapsarkitektene).

```
89          "organizations": [
90              "JBL",
91              "Katja E         Eple",
92              "Mme Lemoine'      Hvit",
93              "Fagerbusk",
94              "Tove E      Sommerfuglbusk",
95              "Rubinspirea",
96              "Salix",
97              "Pipeholurt",
98              "CM",
99              "Geranium Johnsons Blue       Praktstorkenebb",
100             "Rabarber",
101             "Orientvalmue",
102             "Kattehale",
103             "Stella Yellow"
104         ]
```

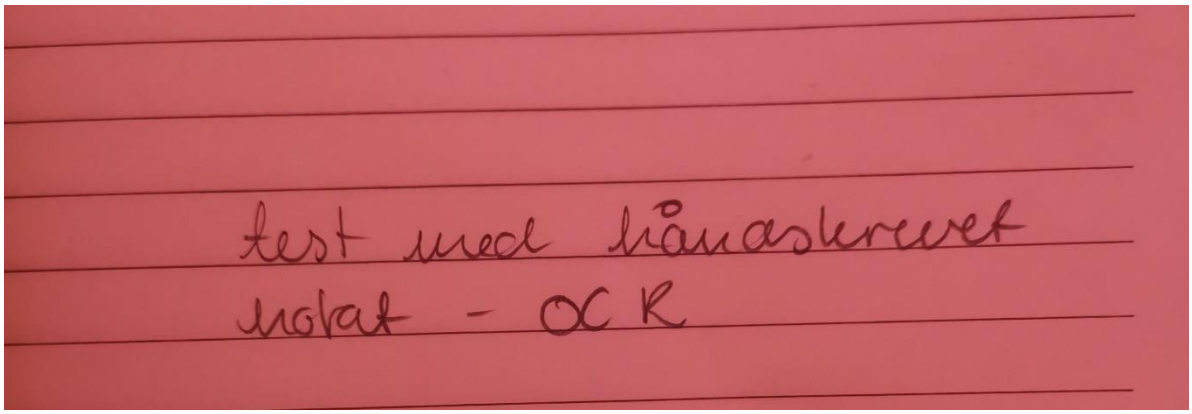
Figur 8 utsnitt av søkeresultat

Forsøk med søk på personnavn

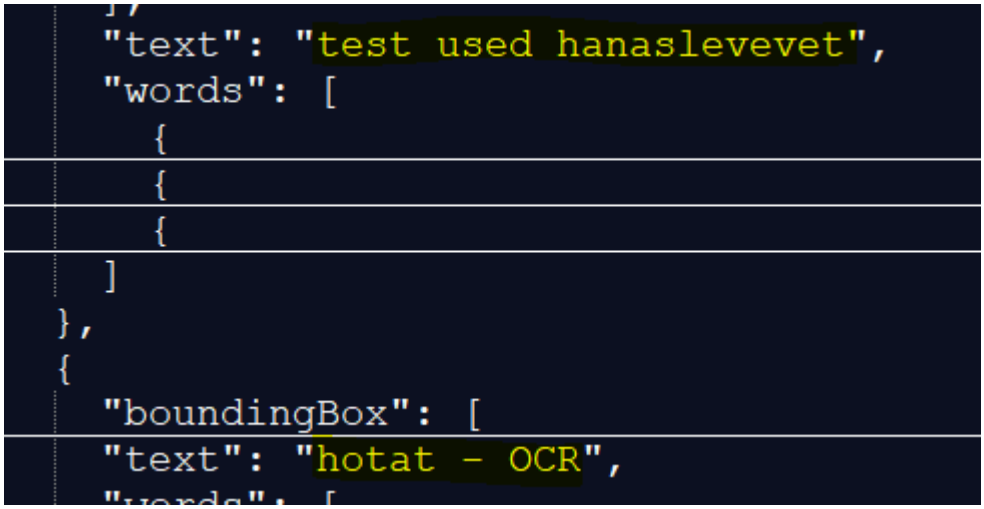
Det ble også gjort søk for å se om det var mulig å finne alle forekomster av en person sitt navn i datasettet. Ved søk på «Ingvild» (forfatters fornavn), ble det listet ut de filene som inneholdt navnet, men det ble ikke definert som en person. Tilsvarende søk ble gjort på andre personer med lignende resultat.

Forsøk med tekst fra håndskrift

Test med å bruke OCR-funksjonaliteten på bilder av håndskrevne notater ble også gjort. Av personvernhensyn er ikke eksempel fra borettslaget tatt med her, men det ble lagt inn et annet forsøksbilde som ikke er fra borettslagets datasett for illustrasjon:



Figur 9: eksempel på input til OCR



Figur 10: Utdrag fra output fra OCR

Her ser man at resultatet kommer i form av de nærmeste engelske ordene den kunne finne.

Oppsummering av forsøket

Av dette forsøket er det erfart at det er relativt enkelt å gjøre data søkbare ved hjelp av Azure Cognitive search. Det krevde ikke veldig stor innsats å få satt opp systemet, og resultatene på å gjøre skannede dokumenter med maskinskrevet tekst søkbar var gode.

Det er derimot ikke sikkert at dette er en god måte for små og mellomstore bedrifter å få oversikt over sine ustrukturerte data på. Dersom borettslaget skulle ha tatt i bruk verktøyene fra dette forsøket i full skala for å få oversikt over papirene på styrerommet ville følgene ha vært nødvendig:

1: dugnadsinnsats på skanning

2: utgifter til Azure

3: styremedlem med kunnskaper til å opprette lagring med de nødvendige sikkerhetsmekanismer og tilhørende søk

4: Produksjon av et brukergrensesnitt som vil gjøre det mulig å utføre søk og slette dokumenter på en enkel måte

For å få fullt utbytte av dette må man ha mer enn gjennomsnittlig kunnskap om de verktøyene som skal brukes. Å gjøre papirene søkbare er rimelig enkelt, men dersom indexene skal brukes av styremedlemmer må det også lages et brukergrensesnitt som gjør det enkelt i bruk. Funksjonaliteten som skal identifisere personer og organisasjoner er ikke brukbar i nåværende tilstand. Dette kan endre seg ettersom Microsoft stadig trener opp algoritmen og den vil nok bli bedre til å forstå norsk med tiden.

Selv om man ikke får til å liste ut personer kan borettslaget likevel få mye ut av å benytte de tjenestene som fungerte godt ved at det blir mulig å søke på navn og nøkkelord i arkivet. For eksempel kan de raskt finne saker som omhandler eiendommen ved behov. De kan også få bedre kontroll på hvem som har tilgang hvis arkivet ligger med passordbeskyttelse i skyen, i stedet for i papirform på styrerommet. I tillegg blir de bedre sikret mot å miste de i vannlekkasjer og slike ting.

Når det gjelder utgifter ble der erfart at prisen for å holde søketjenesten aktiv var på kr. 19 per dag ved 1 index. I forsøket ble det scannet permer med dokumenter som resulterte i et datasett på 31,7MB. Til sammen finnes papirer tilsvarende 96 permer på styrerommet, så man ville ha hatt

behov for omtrent 1GB lagring[2]. Ved å plotte inn estimatene i priskalkulatoren på nettsiden til Azure ser man at man vil få en kostnad på ca kr 7444 per år (AZure, 2019)[3]. Man vil kunne omgås det meste av disse kostnadene ved å kjøre et wildcard-søk på indexen og laste ned resultatet. Man får da en stor JSON-fil man kan gjøre tekstsøk i, men dette blir en lite brukervennlig løsning som også medfører de risikoene som er med lokal lagring.

Forutsatt at man har det styremedlemmet som nevnt i pkt 3 og 4, vil det fortsatt kreve litt av borettslaget både økonomisk og med tanke på innsats å få en slik løsning opp. I tillegg har man ingen garanti for at man vil ha en person som kan vedlikeholde systemet i styret i framtiden, noe som kan føre til kostnader dersom man må benytte konsulenter. Det er dermed mulig, men ikke sikkert, at det vil være mer gunstig å benytte Azure Cognitive Services enn å manuelt gå gjennom arkivet.

[2] (31,7MB/3)*96 = 1014,4MB ≈ 1GB

[3] 898US Dollars multiplisert med dagens kurs 17/04: 8,4921 – se Vedlegg 2

Veileder – en systematisert framstilling av mulige tiltak

Den andre delen av oppgaven er en presentasjon av hvordan en virksomhet kan gå fram for å best mulig bli i samsvar med personvernforordningen. Hensikten var å få laget en oversikt som er lett å lese og som kan benyttes av mindre bedrifter, foreninger og andre virksomheter som har problemer med å få kontroll på ustrukturerte data.

Fremgangsmetode for veilederen

For å få laget en slik oversikt ble det først vurdert å lage en brosjyre i pdf-format som kan skrives ut hos brukeren. Etter å ha tenkt litt på det ble det konkludert med at det vil være mer i tråd med hensikten bak prosjektet å lage en veileder i form av en webside, da dette er mer dynamisk og lettere å nå ut med.

For å gjøre veilederen mest mulig mobilvennlig og lesbar ble det valgt å bruke rammeverket Bootstrap sammen med enkel CSS. Det ble kjøpt inn et billig domene (ustrukturert.no) og for å holde kostnadene med webhotell nede ble det valgt å ikke benytte noe særlig serverside-funksjonalitet. Nettsiden kan ikke gjøre noe særlig annet enn å vise tekst og bilder. Det er brukt litt javascript for enkle popovers og menyer i tillegg til Bootstrap, og ellers går det i simpel html. Fordelene med dette er at det er enkelt å vedlikeholde og kan hostes billig, og man lagrer heller ingenting om de besøkende.

Sluttprodukt: ustrukturert.no

Innholdet på siden består av de rådene og tipsene som kom fram under arbeidet med denne oppgaven. Det ble lagt til lenker til viktige ressurser og popovers med definisjoner. Siden er bygd opp som en steg-for-steg-veiviser, der man starter med kartlegging som første steg og avslutter med tips om gjennomføring og kontinuerlig endring. Det ble også laget en seksjon for vedlegg med to enkle steg-for-steg oppskrifter på hvordan to av tiltakene kan gjøres.

I og med at veilederen er publisert og tilgjengelig for alle er det lagt vekt på at dette er et studentarbeid og at man ikke kan anse rådene som juridiske råd.

Avslutningsvis finnes det en e-post adresse der brukere kan gi tilbakemeldinger og forslag til innhold. Dersom websiden skal drives videre vil det kunne komme til nytte. En utskrift av nettsiden finnes i vedlegg 3.

Oppsummering

Hensikten med denne oppgaven har vært å finne ut hva en virksomhet kan gjøre for å få kontroll på ustrukturerte data og å presentere dette på en ryddig og tilgjengelig måte som kan komme til nytte for virksomheter som jobber med dette.

Arbeidet med å finne ut hvorvidt Azure Cognitive Services kan brukes til dette formålet ga ikke et entydig svar, men det er tydelig at teknologien virker og er tilgjengelig selv om det kunne ha vært enda enklere å ta i bruk for små virksomheter. Testen som ble utført med data på engelsk indikerer at språk er avgjørende for hvor godt tjenesten fungerer. Det er derfor grunn til å tro at tjenesten vil bli bedre til å tolke og gjenkjenne norsk etter hvert som den blir mer brukt og får mer trening. Dette kan gjøre den mer anvendelig i framtiden.

Nye spørsmål har dukket opp, spesielt interessant er problemstillingen med metadata som skal brukes til å bli mer i samsvar med GDR, men som samtidig utgjør ytterligere personopplysninger. Dette var det lite tilgjengelig faglitteratur om og det kan være interessant å se videre på dette både i et teknologisk og rettsvitenskapelig perspektiv.

Alt i alt ble det funnet mange tiltak som kan gjennomføres, og det har vært mulig å få presentert disse i en oversiktlig form. I arbeidet med oppgaven har det vært gledelig å oppdage at det er svært mye det går an å gjøre, selv med begrensede midler.

Referanser

Azure Team. (2018, 11). *JFK-demo*. Hentet fra JFK-demo: https://jfk-demo.azurewebsites.net/

AZure, M. (2019, 04 04). *azure.micrisoft*. Hentet fra pricing calculator: https://azure.microsoft.com/en-us/pricing/calculator/?service=storage

Datatilsynet. (2016, 06 2). *datatilsynet.no*. Hentet fra datatilsynet.no: https://www.datatilsynet.no/regelverk-og-verktoy/veiledere/kameraovervaking/

datatilsynet. (2018, 05 28). *Datatilsynet.no*. Hentet fra Datatilsynet.no/rettigheter-og-plikter: https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/behandlingsgrunnlag/

datatilsynet. (2018). *Datatilsynet.no*. Hentet fra datatilsynet: https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/protokoll-over-behandlingsaktiviteter/

Datatilsynet. (2018). *datatilsynet.no*. Hentet fra https://www.datatilsynet.no/regelverk-og-verktoy/: https://www.datatilsynet.no/regelverk-og-verktoy/veiledere/veileder-om-behandlingsgrunnlag/

Datatilsynet. (2018, 08 07). *datatilsynet.no*. Hentet fra datatilsynet.no: https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/avvikshandtering/nar-skal-jeg-melde-avvik/

Eikivil, L. (1993). *OCR - optical character recognition.* Oslo: Norsk regnesentral.

Hjertø, G. (2018). Innføring av ISMS - oppstart og forankring. Institutt for datateknologi og informatikk, NTNU.

IBM. (2014). *IBM Security Services 2014 Cyber Security Intelligence Index.* Somers NY: IBM Global Technology Services.

James A Martin, J. K. (2018, 10 09). *CSO*. Hentet fra CSOonline: https://www.csoonline.com/article/2120384/what-is-iam-identity-and-access-management-explained.html

JBS Borettslag. (2013, 6 17). *JBS Borettslag*. Hentet fra jbsborettslag.no: http://jbsborettslag.no/5272/om-oss/om-oss

Kuppinger, M. (2018). *Governance for all data: Get a grip on unstructured data.* Kuppinger Cole Analytics.

Lovdata. (2018, 7 2). *Forskrift om arbeidsgivers innsyn i e-postkasse og annet elektronisk materiale*. Hentet fra lovdata.no: https://lovdata.no/dokument/SF/forskrift/2018-07-02-1108

Lovdata. (2018). *Lovdata.no*. Hentet fra Lov om behandling av personopplysninger (personopplysningsloven): https://lovdata.no/lov/2018-06-15-38/§32

Marsh, E. (1998, 04 29). MUC7-evluation of IE-technology: overview of results. *MUC7-evluation of IE-technology: overview of results*. MUC-7.

Microsoft. (2018, 05 14). *GitHub*. Hentet fra Github: https://github.com/Microsoft/AzureSearch_JFK_Files

Microsoft. (2018, 05 07). *Microsoft Azure*. Hentet fra Microsoft Azure: https://azure.microsoft.com/en-us/blog/announcing-cognitive-search-azure-search-cognitive-capabilities/

Microsoft. (2019, 02 13). *docs.microsoft.com*. Hentet fra Microsoft Azure: https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-keyword-extraction

Microsoft Azure. (2019, 04 11). *docs.microsoft.com*. Hentet fra docs.microsoft.com: https://docs.microsoft.com/en-us/azure/cognitive-services/language-support

Microsoft Trust Portal. (2019). *Microsoft Trust Portal*. Hentet fra servicetrust.microsoft.com: https://servicetrust.microsoft.com/Documents/ComplianceReports

norSIS. (2019, 04 01). *norsis.no*. Hentet fra norsis.no: norsis.no/sikker-sletting

Thales eSecurity. (2017). *Why there's more to the GDPR than just fines.* Plantation FL: Thales eSecurity.

Vedlegg 1 – NSD Skjema og vurdering

NSD NORSK SENTER FOR FORSKNINGSDATA

Meldeskjema 664854

Sist oppdatert

18.02.2019

Hvilke personopplysninger skal du behandle?

- Navn (også ved signatur/samtykke)
- Fødselsnummer eller andre nasjonale identifikasjonsnumre
- Fødselsdato
- Adresse eller telefonnummer
- Bakgrunnsopplysninger som vil kunne identifisere en person
- Andre opplysninger som vil kunne identifisere en fysisk person

Type opplysninger

Du har svart ja til at du skal behandle bakgrunnsopplysninger, beskriv hvilke

Det kan være kvtteringer fra håndtverkerfirma for eksempel med navn og arbeidssted til de som har utført en jobb i utvalget

Du har svart ja til at du behandler andre opplysninger som vil kunne identifisere en person, beskriv hvilke

Det kan være henvisninger til leilighetsnummer/oppgang og så videre som kan utledes til hvem noen er uten at navnet står der.

Skal du behandle særlige kategorier personopplysninger eller personopplysninger om straffedommer eller lovovertredelser?

Nei

Prosjektinformasjon

Prosjekttittel

Bruke cognitive services for å kartlegge ustrukturerte data

Prosjektbeskrivelse

Det er vanskelig for en bedrift å vite hva de egentlig har av ustrukturerte data, for eksempel i papirform. Prosjektet går ut på å scanne inn papirer fra arkivet til JBS borettslag og se om det er mulig å bruke Azure Cognitive services til å lese håndskrift for å kartlegge hva man har av personopplysninger, hva som bør slettes og hva man må ta vare på.

Fagfelt

Teknologi

Begrunn behovet for å behandle personopplysningene

Bakgrunnen for at jeg må behandle disse personopplysningene er at det uten et eksempel-datasett ikke vil være mulig å finne ut om borettslaget (og andre i samme situasjon) kan benytte azure cognitive services for å oppfylle sine forpliktelser i henhold til gdpr.

Ekstern finansiering

Type prosjekt

Studentprosjekt, bachelorstudium

Gjelder innmeldingen for flere studentprosjekter (felles vurdering)?

Nei

Kontaktinformasjon, student

Ingvild Stølen, ingvilst@stud.ntnu.no, tlf: 40499180

Behandlingsansvar

Behandlingsansvarlig institusjon

NTNU Norges teknisk-naturvitenskapelige universitet / Fakultet for informasjonsteknologi og elektroteknikk (IE) / Institutt for datateknologi og informatikk

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Torstein Elias L. Hjelle, torstein.hjelle@ntnu.no, tlf: 91702358

Skal behandlingsansvaret deles med andre institusjoner (felles behandlingsansvarlige)?

Nei

Utvalg 1

Beskriv utvalget

Personer som har vært i kontakt med JBS borettslag

Rekruttering eller trekking av utvalget

Tilfeldig valgte permer fra gammelt arkiv

Alder

18 - 100

Inngår det voksne (18 år +) i utvalget som ikke kan samtykke selv?

Nei

Personopplysninger for utvalg 1

- Navn (også ved signatur/samtykke)
- Fødselsnummer eller andre nasjonale identifikasjonsnumre
- Fødselsdato
- Adresse eller telefonnummer
- Bakgrunnsopplysninger som vil kunne identifisere en person
- Andre opplysninger som vil kunne identifisere en fysisk person

Hvordan samler du inn data fra utvalg 1?

Annet

Beskriv

Grunnlag for å behandle alminnelige kategorier av personopplysninger

Allmenn interesse eller offentlig myndighet (art. 6 nr. 1 bokstav e)

Redegjør for valget av behandlingsgrunnlag

Behandlingen vil bedre personvernet for de som omfattes. Målet med behandlingen er at borettslaget ikke lengre skal oppbevare papirer med personopplysninger i permer i styrerommet dersom det ikke er nødvendig. Tilgang til informasjonen er kun til person som allerede har denne tilgangen (studenten som gjør behandlingen er styremedlem med tilgang). Personopplysninger vil sladdes i rapporten/oppgaven.

Informasjon for utvalg 1

Informerer du utvalget om behandlingen av opplysningene?

Nei

Begrunn hvorfor du ikke informerer utvalget om behandlingen.

Bakgrunnen for prosjektet er at man ikke på forhånd vet hvem man har personopplysninger om i gamle papirer. Det blir derfor vanskelig å varsle i forkant til de man i etterkant finner ut at det er personopplysninger om.

Tredjepersoner

Skal du behandle personopplysninger om tredjepersoner?

Nei

Dokumentasjon

Hvordan dokumenteres samtykkene?

Ikke utfyllt

Hvordan kan de registrerte få innsyn, rettet eller slettet opplysninger om seg selv?

Dersom noen som har vært i kontakt med borettslaget kontakter styret via mail vil borettslagets styre gå gjennom og slette det de kan finne av informasjon om den enkelte.

Totalt antall registrerte i prosjektet

1-99

Tillatelser

Skal du innhente følgende godkjenninger eller tillatelser for prosjektet?

Behandling

Hvor behandles opplysningene?

- Ekstern tjeneste eller nettverk (databehandler)

Hvem behandler/har tilgang til opplysningene?

- Student (studentprosjekt)
- Databehandler

Hvilken databehandler har tilgang til opplysningene?

Microsoft Azure

Tilgjengeliggjøres opplysningene utenfor EU/EØS til en tredjestat eller internasjonal organisasjon?

Nei

Sikkerhet

Oppbevares personopplysningene atskilt fra øvrige data (kodenøkkel)?

Nei

Begrunn hvorfor personopplysningene oppbevares sammen med de øvrige opplysningene

Det er ikke to forskjellige ting (personopplysningene er dataene)

Hvilke tekniske og fysiske tiltak sikrer personopplysningene?

- Opplysningene krypteres under forsendelse
- Flerfaktorautentisering
- Adgangsbegrensning
- Adgangslogg
- Endringslogg

Varighet

Prosjektperiode

11.02.2019 - 16.05.2019

Skal data med personopplysninger oppbevares utover prosjektperioden?

Nei, alle data slettes innen prosjektslutt

Vil de registrerte kunne identifiseres (direkte eller indirekte) i oppgave/avhandling/øvrige publikasjoner fra prosjektet?

Nei

Tilleggsopplysninger

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

Bruke cognitive services for å kartlegge ustrukturerte data

Referansenummer

664854

Registrert

18.02.2019 av Ingvild Stølen - ingvilst@stud.ntnu.no

Behandlingsansvarlig institusjon

NTNU Norges teknisk-naturvitenskapelige universitet / Fakultet for informasjonsteknologi og elektroteknikk (IE) / Institutt for datateknologi og informatikk

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Torstein Elias L. Hjelle, torstein.hjelle@ntnu.no, tlf: 91702358

Type prosjekt

Studentprosjekt, bachelorstudium

Kontaktinformasjon, student

Ingvild Stølen, ingvilst@stud.ntnu.no, tlf: 40499180

Prosjektperiode

11.02.2019 - 16.05.2019

Status

21.03.2019 - Vurdert

Vurdering (1)

21.03.2019 - Vurdert

Det er vår vurdering at behandlingen vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet den 21.03.2019 med vedlegg. Behandlingen kan starte.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:

https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html

Du må vente på svar fra NSD før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 16.05.2019.

LOVLIG GRUNNLAG
Prosjektet vil behandle personopplysninger med grunnlag i en oppgave av allmenn interesse.

Vår vurdering er at behandlingen oppfyller vilkåret om vitenskapelig forskning, jf. personopplysningsloven § 8, og dermed utfører en oppgave i allmenhetens interesse.
Lovlig grunnlag for behandlingen vil dermed være utførelse av en oppgave i allmenhetens interesse, jf. personvernforordningen art. 6 nr. 1 bokstav e, jf. art. 6 nr. 3 bokstav b), jf. personopplysningsloven § 8.

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen:

- om lovlighet, rettferdighet og åpenhet (art. 5.1 a)
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles til nye uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), protest (art 21).

Det gjøres unntak fra retten til informasjon (jf. art. 14 nr. 5, bokstav b) med begrunnelse i at datamaterialet som skal analyseres er av historisk art og at det vil være umulig eller uforholdsmessig vanskelig å finne de registrerte som ikke lenger bor i det aktuelle borettslaget eller som driver virksomhet som tidligere har vært i kontakt med borettslaget. Et sentralt aspekt ved prosjektet er å teste ut hvorvidt en type programvare er egnet til å avgjøre hvorvidt et papirmateriale inneholder personopplysninger. Man vet derfor ikke i forkant i hvilken grad man behandler personopplysninger.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned. Ansvaret for at de registrerte får benyttet seg av sine rettigheter i forbindelse med behandlingen av opplysningene til dette konkrete forskningsformålet påhviler behandlingsansvarlig institusjon (NTNU), ikke borettslaget.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32)

Microsoft Azure er databehandler i prosjektet. NSD legger til grunn at behandlingen oppfyller kravene til bruk av databehandler, jf. art 28 og 29.

For å forsikre dere om at kravene oppfylles, må prosjektansvarlig følge interne retningslinjer/rådføre dere med behandlingsansvarlig institusjon.

KOMMENTAR TIL STUDENTENS DOBBELTROLLE
Det fremgår av meldeskjema at studenten selv har en rolle i det aktuelle borettslaget som bidrar med deler av sitt papirarkiv. Studenten har ikke tilgang til dette arkivet i kraft av sin rolle som student, men som medlem av styret i borettslaget. I forbindelse med dette prosjektet behandles opplysningene til et vitenskapelig

formåk og studentens rolle følger derav. Borettslaget må gjøres oppmerksom på dette slik at de tar skikkelig stilling til om de har anledning til å la studenten benytte opplysningene til et forskningsformål. Det er imidlertid åpenbart en fordel at opplysningene ikke gjøres kjent for andre enn de som allerede har tilgang til dem, uavhengig av rolle.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet/pågår i tråd med den behandlingen som er dokumentert.

Lykke til med prosjektet!

Kontaktperson hos NSD: Marie S. Schildmann
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Vedlegg 2 – pristabell Azure Cognitive Services

Microsoft Azure Estimate

Your Estimate

Service type	Custom name	Region
Storage Accounts		East US
Cognitive Services		West US
Storage Accounts		East US
Support		

Disclaimer

All prices shown are in US Dollar ($). This is a summary estimate, not a quote. For up to a
This estimate was created at 4/17/2019 6:11:41 PM UTC.

Description	Estimated Cost
Block Blob Storage, General Purpose V2, LRS Redundancy, Hot Access Tier, 2 GB Capacity, 100 Write operations, 100 List and Create Container Operations, 1,000 Read operations, 1 Other operations. 2 GB Data Retrieval, 2 GB Data Write	$0,04
Text Analytics: S0 size, 25,000 included transactions with 0 overages.	$74,71
Block Blob Storage, General Purpose V2, LRS Redundancy, Hot Access Tier, 2 GB Capacity, 1,000 Write operations, 1,000 List and Create Container Operations, 1,000 Read operations, 1 Other operations. 1,000 GB Data Retrieval, 1,000 GB Data Write	$0,05
Support	$0,00
Licensing Program	Microsoft Online Services Program (
Monthly Total	$74,81
Annual Total	$897,66

'ate pricing information please visit https://azure.microsoft.com/pricing/calculator/

Vedlegg 3 – utskrift av nettsiden ustrukturert.no

Velkommen til veilederen for ustrukturerte data og GDPR

Denne nettsiden er laget som en del av en bacheloroppgave ved NTNU. Formålet er å samle tips om hvordan en virksomhet kan gå fram for å få oversikt over personoppplysninger i ustrukturerte data og komme bedre i samsvar med personvernforordningen (GDPR).

Introduksjon

Når en bedrift skal sørge for at den kan oppfylle GDPR, vil det være relativt enkelt å få oversikt over data som er lagret i strukturert form, som for eksempel i kundedatabasen, HR-systemet eller et elevregister. Langt vanskeligere er det å skaffe en oversikt og sikre at man er i samsvar når det gjelder ustrukturerte data, men personvernforordningen er like gjeldende på dette området. Med ustrukturerte data menes data som ligger i e-poster, filer, papirer og så videre. Her er det ofte dårlig oversikt over hva som finnes og det er skrevet lite om problemstillingen. Dersom man gjør et google-søk er de fleste treffene konsulentselskaper som vil selge inn sine løsninger. Disse løsningene er som oftest kostbare, omfattende og skaper gjerne MER personopplysninger som lagres ved at de genererer metadata og rapporter. Derfor bestemte jeg meg for å se nærmere på hva en virksomhet selv kan gjøre av konkrete tiltak. Resultatet ble min bacheloroppgave som denne nettsiden er en del av.

Denne veilederen tar sikte på å skaffe en oversikt over tiltak en bedrift kan gjøre for å sikre samsvar med GDPR også for ustrukturerte data. Det tas utgangspunkt i at virksomhetens som skal benytte veilederen allerede har de vanlige tiltak for samsvar med personvernforordningen på plass, som for eksempel internkontroll, personvernerklæring og rutiner for behandling av strukturerte data.

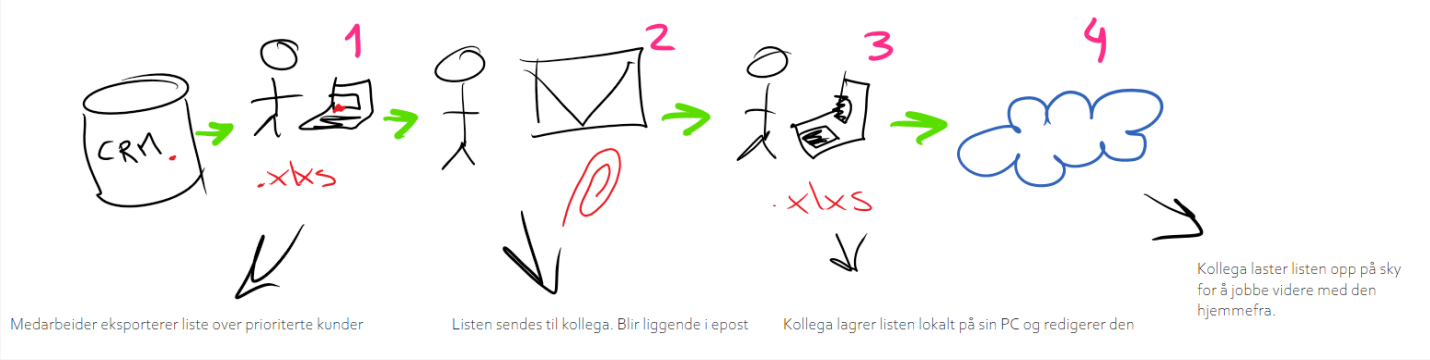
Neste

Forbehold

Denne veilederen er et studentarbeid, og kan ikke betraktes som juridiske råd. Bruk av veilederen garanterer på ingen måte at virksomheten blir i samsvar med GDPR

Steg 1: Kartlegging

Din virksomhet har antageligvis allerede gjort en kartlegging av personopplysninger i arbeidet med personvern, men er det viet tilstrekkelig oppmerksomhet til de ustrukturerte dataene? Siden GDPR gjelder ALLE personopplysninger er det fordelaktig å gå gjennom det hele på nytt, men nå med tanke på ustrukturerte data. For dette vil jeg anbefale at det arrangeres en workshop, med ulike folk i virksomheten som kan tenkes å behandle personopplysninger. Ta gjerne med noen representanter fra hver avdeling eller hvert virksomhetsområde, og sørg for at minst én fra ledelsen er med slik at man får forankret arbeidet også i toppen.
Dersom å arrangere en workshop ikke er aktuelt, kan man gjennomføre intervjuer hvor man tar med sjekklisten rundt i ulike avdelinger, og forhører seg om hvordan behandling av personopplysninger foregår.



Bildet viser en typisk flyt der en medarbeider i salgsavdelingen har eksportert en liste med viktige kunder, for å finne ut hvilke bedriften skal jobbe videre med. Her ender man opp med fem kopier av listen med alle personopplysningene liggende som lokale kopier, i sky og i e-postboksene til avsender og mottaker. Prøv å få nedtegnet slike typer prosesser som er typiske for din virksomhet.

Sjekkliste

- **Uttrekk:** Trekker dere data ut fra systemer som da går fra å være strukturert til å bli ustrukturert? Eksporterer dere for eksempel kundelister ut i fra CRM til Excel-filer som deretter sendes rundt på e-post? Eller tas lister over ansatte ut fra HRM-systemet i ulike sammenhenger?
  Prøv å få en oversikt over alle tilfeller der dette skjer, slik at dere senere kan vurdere tiltak for å gjenhente kontrollen. Skriv ned både on-demand eksporteringer og periodisk oppsatte eksporteringer.
- **E-post:** Hva befinner seg ulike e-postbokser ruundt om i bedriften? Sannsynligvis finnes det mye personopplysninger i e-poster rundt omkring. Her kan det være nyttig å skille på personlige e-postkontoer (f. eks ola.nordmann@bedrift.no), der arbeidsgiver ikke har ubetinget rett til innsyn, og upersonlige e-poster (f. eks post@bedrift.no), der arbeidsgiver har kontroll.
- **Digitale filer:** Her gjelder det samme som for e-post, prøv å skille mellom filer på delte områder og på peronlige områder. Sørg for å få med:
  - Filer med personopplysninger på hjemmeområder tilhørende ansatte.
  - Filer med personopplysninger liggende på ulike lagringsmedier som minnepinner, CD-er og så videre.
  - Backups
  - Vedlegg i eposter
  - Lokalt lagrede filer på enheter som tilhører virksomheten, som f. eks på skrivebordet eller i trash-folder.
  - Lokalt lagrede filer på andre enheter, som private PC-er, mobiltelefoner og så videre.
- **Lyd, bilder og video:** Sjekk om dere har lyd-, video- eller bildefiler liggende som kan inneholde personopplysninger. Noen eksempler på dette er opptak av telefonsamtaler, video fra overvåkningsutstyr og små filmsnutter brukt I intern eller ekstern markedsføring. Disse er i samme grad som skriftlige kilder og bilder underlagt personvernforordningen.

- **Dokumenter i papirform:** Mange dokumenter finnes kun i fysisk format og kan bli liggende i lang tid i arkiver, eller mindre organisert på diverse skrivebord og i skuffer rundt om i bedriften. De samme reglene gjelder for personopplysninger her som for opplysninger lagret i elektronisk format, og det kreves at virksomheten har oversikt. Slike dokumenter inneholder ofte personopplysninger slik som navn på kvitteringer, bilder av folk og møtereferat, eller enda mer detaljert som når søknader og relaterte dokumenter er innlevert eller skrevet ut på papir.

Tilbake Neste

Steg 2: behandlingsgrunnlag

Man har ikke lov til å holde på personopplysninger som man ikke har behandligsgrunnlag for. Det er derfor nødvendig å gå gjennom alle personopplysninger man fant i kartleggingen og avgjøre hva som er behandlingsgrunnlaget for hver enkelt av de.
Man har ikke lov til å holde på sensitive personopplysninger (med noen unntak). Som sensitive personopplysninger regnes:

- Opplysninger om rasemessig eller etnisk opprinnelse
- Opplysninger om politisk oppfatning
- Opplysninger om religion
- Opplysninger om filosofisk overbevisning
- Opplysninger om fagforeningsmedlemskap
- Genetiske opplysninger
- Biometriske opplysninger med det formål å entydig identifisere noen
- Helseopplysninger
- Opplysninger om seksuelle forhold
- Opplysninger om seksuell legning
- Opplysninger om straffedommer
- Opplysninger om lovovertredelser

Dersom man finner sensitive personopplysninger som man mener at man trenger, kan man sjekke de unntakene som finnes på datatilsynet og se om behandlingen faller inn under en av disse.
Dersom man ikke har behandlingsgrunnlag skal opplysningene slettes.

Ulike behandlingsgrunnlag

De ulike behandlinggrunnlagene man kan ha er:

- **Samtykke:** Personen det gjelder har gitt aktivt og informert samtykke til behandlingen.
- **Nødvendig for å oppfylle en avtale:** Gjelder dersom behandlingen er objektivt nødvendig for å kunne utføre en tjeneste eller levere en vare som personen har bedt om.
- **Nødvendig for å oppfylle en rettsig plikt:** Gjelder for opplysninger som virksomheten er pålagt å holde på etter lov.
- **Nødvendig for vitale interesser:** Kan kun brukes i sjeldne tilfeller der behandlingen er nødvendig for å beskytte den registrertes, eller en annen person sine vitale interesser (det vil si liv-eller død situasjoner).
- **Nødvendig for å utføre en oppgave i allmennhetens interesse eller utøve offentlig myndighet:** Dette grunnlaget omfatter i hovedsak offentlige virksomheter og private virksomheter som utfører oppdrag for det offentlige.
- **Nødvendig for å ivareta legitime interesser:** Kan brukes dersom behandlingen er nødvendig for å ivareta en berettiget interesse. Må vektes mot konsekvensene for personvernet.

På datatilsynet sine hjemmesider kan man lese mer detaljert hva som er gyldige behandlingsgrunnlag.

Protokoll

Alle virksomheter som behandler personopplysninger skal føre protokoll over disse. Protokollen skal inneholde hvilke typer personopplysninger som lagres, hvem de eventuelt deles med, hvordan de oppbevares og hvilket behandlingsgrunnlag man har.
Dersom virksomheten allerede har en slik protokoll, anbefales det at man fyller ut videre på denne med opplysningene man kom fram til i kartleggingen, og legger til behandlingsgrunnlag og tredjeparter.
Dersom virksomheten ikke har en slik protokoll kan man laste ned mal for protokoll fra datatilsynet her.

Etter å ha gått gjennom alle kilder til personopplysninger i utrukturert form og sett på behandlingsgrunnlag for hver enkelt av de, har man kanskje funnet noen som det ikke var behandlingsgrunnlag for. Disse må slettes med en gang før man går videre til neste steg: finne tiltak.

Steg 3: Finne tiltak

Etter at man har identifisert hvilke personopplysninger man har og hvorfor man har de, er det på tide å finne tiltak man kan gjøre for å være i samsvar med personvernforordningen.
Under finner du en liste over mulige tiltak virksomheten kan sette i verk for å bedre personvernet. Husk at selv om man har behandlingsgrunnlag, så er det allikevel krav om at personopplysninger skal slettes innen rimelig tid. Man er også pliktig til å ha tilstrekkelig informasjonssikkerhet.

Oppdatere personvernerklæringen

Når man har fått et fullstendig bilde av personopplysningene som lagres kan det være nødvendig at man oppdaterer personvernerklæringen. Alle registrerte har krav på å få vite hvilke opplysninger som er registrert og hva de brukes til, og dette skal være lett tilgjengelig og presenteres på en lettfattelig måte. Gå gjennom virksomhetens personvernerklæring og sjekk at man har med all behandling her.

Dersom dere ikke har en personvererklæring allerede er må det lages en og den må gjøres lett tilgjengelig for alle registrerte. Datatilsynet har laget en fin veileder som beskriver hvordan man tilfredsstiller kravene til informasjon og åpenhet. Den finner du her.

Rutiner for fysisk sikring

Dersom man i kartleggingen oppdager at data er oppbevart på måter som innebærer risiko for at personopplysninger kommer på avveie må det iverksettes tiltak for fysisk sikring. Dette kan være enkle ting som for eksempel låsing av arkivskap, låser på skuffeseksjoner og oppbevaringsmøbler på kontorer og ulike sektorer i kontorlandskapet med ulik tilgang.
En gjennomgang kan dessverre ofte resultere i at man finner ut at det er personopplysninger lagret i permer som står fritt tilgjengelig på hyller rundt om, i kontorer som står åpne eller i landskap. En annen typisk feil er at papirer med personopplysninger blir liggende uavhentet på printer. Her finnes det mange løsninger med at ansatte må taste en kode eller bruke nøkkelkort med RFID for at utskriften skal starte.
Alle som kan ha befatning med personopplysninger bør bevisstgjøres dette og det vil være fordelaktig om det utarbeides klare rutiner for hvordan papirer og utskrifter skal håndteres. Regelen om at man ikke har lov til å holde på personopplysninger legre enn det som er relevant gjelder også for papirer. Derfor kan det være smart å arrangere ryddedager på kontoret og lignende tiltak for å sørge for at man ikke oppbevarer gamle papirer. Husk også at papirene skal makuleres; å legge de til gjenvinning vil ikke være tilfredsstillende.

Eksempel på instrukser, rutiner og tiltak

- Tenk deg om før du printer. Er det nødvendig?
- Tenk deg om før du lagrer: Må dette papiret være tilgjengelig? Kan det makuleres med en gang?
- Ryddedag: Dugnad på kontoret kan arrangeres for eksempel før jul og sommerferie hvert år. Oppfordre til å makulere alt som ikke trengs. Kan gjerne kombineres med noe hyggelig som en litt ekstra god lunsj eller kakepause for å skape god stemning rundt tiltaket.
- Løsninger for kontroll av utskrifter (pull printing) For eksempel:
  - safecom PullPrint
  - HP Access Control
  - Ringdale FollowMe
- Lås på skuffer og skap der papirer med personopplysninger oppbevares

Rutinemessig sletting av tilganger og data ved avslutning av stilling

Manuelle rutiner

I sin enkleste form kan en manuell rutine gå ut på at en administrator fjerner tilgangene til en ansatt i det de slutter, samtidig som hjemmeområde, mailboks og annet relevant innhold slettes. Dersom dette skal ivaretas med manuelle rutiner fordrer det et godt samarbeid mellom IT-avdelingen, avdeling for lønn- og personal og mellomledere. Mellomledere må gi beskjed når en ansatt sier opp eller bytter stilling, både til lønns- og personalavdelingen og til IT-avdelingen, eventuelt må ansvaret for å varsle IT-avdelingen tilfalle HR-avdelingen. Det må i begge tilfeller tegnes opp helt klare rutiner, og ansvaret må plasseres tydelig for å sikre seg at avslutninger blir fulgt opp skikkelig.
Det vil ikke være i henhold til personopplysningsforordningen å oppbevare filer og e-postarkiv fra personell som ikke lengre jobber i virksomheten.

Eksempel på sjekkliste

- Slett e-post konto
- Avslutt arbeidsforhold i HRM-system og foreta sluttoppgjør
- Samle inn nøkler/nøkkelkort
- Deaktiver AD-konto
- Samle inn jobbtelefon
- Samle inn jobb-PC
- Avslutt telefonabonnement
- Slett oppføring av personen i kontaktlisten på intranett
- Sperr tilgang til fagysystemer som:
  - Saksbehandlingssystem
  - Kunderegister/CRM
  - Postjournal

- System for registrering av reiseregniger

Automatiske og halvautomatiske rutiner

Man kan kjøre sletting ved manuelle rutiner, men det er antageligvis tryggere å sette opp løsninger som automatisk utfører denne jobben for å sikre at ikke det blir glemt. Det er ulike tilbyder på markedet som lager alt fra enkle løsninger med sjekklister som sendes til ledere og IT-ansvarlig, til mer avanserte løsninger (se tiltak: IDM). Dersom virksomheten benytter Active Direcory vil det være smart å benytte AD-brukeren til autentisering på så mange av applikasjonene som brukes som mulig. Dersom virksomheten benytter Azure AD kan man knytte fagsystemer opp i Azure Application Management som gir meget god oversikt over hvilke tilganger en bruker har og gjør det enkelt å avslutte dem.

Fullautomatisert livssyklus-håndtering av brukere (IDM-system)

For større virksomheter kan det være aktuelt med et system for livssyklus-håndtering av brukere.
Et avansert IDM-system (Identity Management System) vil hver dag hente uttrekk fra HRM-systemet (Human Resource Management). Det er her alle ansatte registreres når de starter, slutter eller endrer stilling. Å knytte tilganger, grupper og brukerkontoer opp mot HRM er svært nyttig fordi det reduserer sjansen for menneskelige feil (data punches ett sted), og man knytter tilgangene opp mot det mest oppdaterte systemet. I og med at lønn og kostnader knyttes direkte til avdeling og stillingsstatus i HRM-systemet, vil dette normalt være mer oppdatert enn de fleste andre systemer en virksomhet har.
Når IDM-systemet finner en ny ansatt vil det se på hvilken stillingstype personen har, og dersom det er en type stilling som skal ha databruker vil det opprette AD-konto, e-postkonto og et hjemmeområde som blir aktivert på angitt startdato. De mest avanserte IDM—systemene har også integrasjoner mot fagsystemer som for eksempel saksbehandlingssystem, slik at tilgang opprettes også her for de stillingstypene som skal ha slik adgang. Dette kan også kobles mot systemer for fysiske tilganger gjennom API mot nøkkelkort-systemer.
De beste systemene holder også orden på om en bruker skifter stilling, og avslutter tilgangene på tidligere stilling og innvilger tilganger som er relevante for ny stilling ved skifte. Dette vil hindre at en ansatt som har jobbet i en stor organisasjon i mange år har opparbeidet seg mange tilganger til ulike systemer i løpet av årene som egentlig skulle ha vært avsluttet.
IDM-systemet vil plukke opp når det er satt slutt-dato eller at en ansatt ikke lenger finnes i uttrekket. Da legges brukeren i karantene, og AD-brukeren med tilhørende stillinger settes til inaktiv så den ikke har tilgang på systemer og e-post lenger. Dersom det har vært en feil i HRM-systemet vil dette raskt avdekkes da den ansatte ikke kommer seg på e-post eller noen systemer, og brukeren kan enkelt hentes tilbake. Etter en periode (vanligvis 30 dager) vil hjemmeområdene og e-postboksen til brukeren slettes permanent.
Disse systemene er ganske kostbare og vanligvis ikke aktuelle for små og mellomstore bedrifter. De fleste som benytter slike systemer er store bedrifter med mange ansatte og virksomheter som skoler og universitet, som har stor gjennomstrømming av brukere som skal opprettes og avsluttes. En hyggelig bieffekt av et slikt system er at det også bidrar til å holde lisenskostnader i sjakk og at man kan spare på administrasjon.
De vanligste leverandørene av IDM-systmer i Norge er:

- Innofactor
- DotNet Internals
- Identum

Automatisk sletting av filer basert på utløpsdato

Ved hjelp av Powershell og Task Scheduler kan man selv uten store budsjetter sette opp slik at filer i en mappe slettes hvis de ikke er endret etter et visst antall dager. Dette er rimelig enkelt og kan settes opp selv om man ikke har de helt store programmeringskunnskapene. Det vil fungere både på Windows Server og på klientene.
Det vil for de fleste ikke være aktuelt å sette på en slik task på alle mappene, men det kan være nyttig for å hindre at det ligger gamle ting i nedlastings-mappen for eksempel. I alle tilfeller bør man gå over søppel-folderen og sikre at det er automatisk sletting på den. I tillegg til å hindre at personopplysninger blir liggende sparer man også mye lagringplass på dette. Under vedlegg finner du en oppskrift på hvordan man setter opp dette for Windows 10.

Automatisk sletting av e-post basert på utløpsdato

E-postklienten inneholder ofte mye data man har liten oversikt over, både i e-poster og som vedlegg. Det kan derfor være nyttig å sette opp sletting av mail som er eldre enn for eksempel 30 dager. De fleste e-postklienter har muligheter for dette, i Outlook er det auto-archive funksjonen som vil brukes til dette formålet.
Det vil nok ikke være ønskelig for de fleste å slette all e-post etter en viss dato, men man kan innføre rutiner på at man arkiverer alt som skal tas vare på og auto-sletter det som blir liggende i innboksen. På denne måten tvinger man fram en mer bevisst holdning til hva man trenger å beholde og ikke. Det er også anbefalt å gå over innstillingene for sletting av området for slettede elementer, så man er sikker på at også disse blir slettet fullstendig etter en viss tid. En beskrivelse av hvordan man setter opp slik sletting i Outlook finnes under vedlegg.

Rutine for avhending av utstyr og papirer

Å rutinemessig kvitte seg med data og papirer med personopplysninger man ikke har grunn til å holde på er en naturlig konsekvens av personvernforordningen. Man må ikke glemme at denne avhendingen bør skje på en fornuftig måte. Som hovedregel bør alle papirer som kan inneholde personopplysninger makuleres. Når det gjelder lagringsmedier som USB-minnepinner, harddisker, CD-rom og så videre kan være litt mer komplisert, og norSiS foreslår følgende metoder:

- Å bruke programvare som skriver over data slik at den ikke kan hentes fram.
- Avmagnetisering. Dette vil kun fungere på tradisjonelle harddisker, disketter, magnettaper og andre magnetiske lagringsmedier.
- Fysisk destruksjon som knusing slik at det blir svært vanskelig å rekonstruere lagringsmediet.
- En kombinasjon av de tre metodene som er nevnt.

Samskriving – å dele lenker i stedet for filer

I dag er det både billig og enkelt å benytte seg av ulike samhandlingsplattformer som gjør at ansatte kan samarbeide på et dokument uten at det må sendes fram og tilbake. Det finnes både gratis produkter som kan tas i bruk av små virksomheter, og mer avanserte og kostbare løsninger for større virksomheter. Å samskrive på sentralt lokaliserte dokumenter vil føre til bedre kontroll på hvor filer med personopplysninger befinner seg, og det er ofte også mer effektivt for brukerne. Dette gir også mulighet til å trekke tilbake tilgang til filer eller områder med umiddelbar virkning.
Noen eksempler på slike systemer er:

- **Teams:** et relativt nytt produkt fra Microsoft som samler felles filområder, chattefunksjoner, videomøter og en rekke andre funksjoner i ett produkt. Teams fungerer som en hub for Office 365 og bygger på Sharepoint. Teams finnes i gratisversjon med basisfunksjoner, og som abonnementsprodukt med pris per bruker for en versjon med mer funksjonalitet.
- **Google:** en kombinasjon av googles fildelingstjeneste Google Drive og deres web-applikasjoner Sheets, Docs og slides kan brukes gratis for små volum, eller som betalingstjeneste ved større lagringsbehov.
- **Bitrix24:** full-service tjeneste med CRM, prosjektstyringsverktøy, timeregistrering, chat, videomøter og wiki-løsning i tillegg til fildeling med samskrivingsfunksjonalitet. Verktøyeteveres både som on-premise og SaaS. og er gratis for opptil 12 brukere.

Tjenestene er generelt enkle å ta i bruk. Den største vanskeligheten med dette tiltaket vil være at man må endre folk sine vaner. Når folk er vant til å sende filer i stedet for lenker på e-post kan det være en tung jobb å endre på dette. Det er derfor viktig at ledelsen tar i bruk disse verktøyene aktivt og at implementeringen blir med i arbeidet med sikkerhetskultur. Her kan det være smart å vise til tegningene man laget over prosesser i kartleggingen, slik at man skaper en forståelse for hvordan man mister kontroll på data når det blir sendt rundt som vedlegg.
Kjøper man tjenestene levert på sky vil det også være viktig å forsikre seg om at leverandøren har en databehandleravtale som er i henhold til regelverket.

Maskinlæring som verktøy

Det finnes verktøy basert på maskinlæring som kan være til hjelp i arbeidet med å håndtere ustrukturerte data. Med slike verktøy kan man finne personopplysninger i det som ellers ville ha vært ikke-søkbare filer, som skannede kvitteringer eller dokumenter med håndskrift, filer på pdf-format, bilder med tekst på og lydfiler.

Det finnes selskaper som spesialiserer seg på å tilpasse maskinlæringsteknologi til å skaffe oversikt over personopplysninger. De beste systemene greier å skanne gjennom dokumenter og identifisere og tagge alle forekomster av én type data, som for eksempel alle navn på personer eller alle IP-adresser. Dette kan kombineres med sikkerhetsmekanismer, for eksempel kan man sette kryptering på alle filer hvor man finner et kredittkort-nummer. Et eksempel på dette er Microsoft Azure Information Protection (MSIP). SailPoint er en annen aktør som spesialiserer seg i å sette tilganger til forskjellige roller av brukere basert på kategorier. I tillegg finnes det løsninger som analyserer bilder og produserer søkbar tekst. Eksempler på detter er IBMs Watson og Microsoft Cognitive services.

Det som vil være en fallgruve her er at samtidig som man får oversikt, også skaper mer data med personinformasjon. Alle taggene på dokumentene vil utgjøre enda en registrering av personinformasjonen, og må også behandles i henhold til GDPR. Her blir det viktig å ha et bevisst forhold til hvordan man håndterer slike metadata.

Rutine for å svare ut spørsmål om innsyn

Hvis virksomheten allerede har gjort en jobb mot GDPR har den antageligvis allerede en rutine på plass for hvordan det skal svares ut når noen ber om innsyn. Hos mange virksomheter, kanskje de fleste, er dette noe som ikke skjer så ofte, og det er derfor ikke nødvendig med omfattende og dyre systemer for å kunne tilfredsstille kravet til innsyn. Det bør være klart defiinert og tilgjengelig for den registrerte hvordan den skal henvende seg for å få innsyn, og på virksomhetens side må ansvaret for hvem som skal svare ut disse forespørslene delegeres. Hvis man allerede har en instruks på plass for hvilke data som skal hentes ut kan man gå gjennom funnene fra kartleggingen for å identifisere andre steder man må se for å kunne gi et fullstendig svar på slike henvendelser.
Husk å informere alle i virksomheten om at også filer, eposter og lignende med navnene til registrerte kan bli utlevert på forespørsel, slik at de tar hensyn til dette i det daglige. Man må også ta inn i rutinen en måte å unngå at man lekker andres personopplysninger når det gis innsyn. En epost vil for eksempel inneholde andre e-post adresser i mottaker- og kopifeltet så de kan ikke sendes ut sånn helt uten videre.
En anne viktig ting å huske på er at man har en god rutine for å sjekke identiteten til den som ber om innsyn så man ikke leverer ut informasjonen til hvem som helst.

Jeg har sett gjennom alle tiltakene, hva skjer nå?

Når du har sett gjennom alle disse mulige tiltakene og vurdert hvilke som er aktuelle for din virksomhet, er det på tide å gå videre til neste steg: implementering!

Tilbake Neste

Steg 4: implementering og kontinuerlig arbeid

Å innføre IDM, eller å sette opp et maskinlærings-verktøy er ganske rett fram: man går til sjefen og får et budsjett og setter deretter i gang IT-avdelingen med å implementere.
Like enkelt er det ikke med de foreslåtte tiltakene som går på å innføre rutiner i virksomheten. De fleste som har forsøkt å innføre en rutine vet at dette er mer omstendelig enn å sende ut en epost om at "nå gjør vi det på denne måten dere".

GDPR og informasjonssikkerhet er tett knyttet (informasjonssikkerhet og kontinuerlig arbeid er spesifikt nevnt i personvernforordningen), og det blir derfor naturlig å innlemme bevissthet om personvernforordningen i arbeidet med sikkerhetskultur. IMB har beregnet at så mye som 95% av uønskede hendelser i forbindelse med informasjonssikkerhet hadde menneskelig feil som årsak eller medvirkende faktor. Det er altså ingen grunn til at arbeidet med å bevisstgjøre og skape god kultur rundt sikkerhet og personvern skal få mindre plass en de tekniske tiltakene. For å lykkes med å skape en varig sikkerhetskultur holder det ikke med et kurs eller en felles e-post som sier at fra nå av må alle passe på. Man må involvere alle ledd i virksomheten i et kontinuerlig arbeid. Spesielt viktig er det at ledelsen forstår at dette er arbeid som er av høyeste viktighet for virksomheten, og at de støtter arbeidet på en synlig måte ut til alle involverte.

Hvordan jobbe med sikkerhetskultur - og ta personvern med i arbeidet?

Nasjonal sikkerhetsmyndighet er en god resurs når det kommer til dette med å innarbeide god sikkerhetskultur i en virksomhet. De påpeker også viktigheten av at ledelsen går foran som gode forbilder og at arbeidet kontinuerlig evalueres og følges opp. Hver enkelt virksomhet må se an hva som passer hos seg, hovedsaken er at alle i virksomheten forstår hvorfor sikkerhet og personvern er viktig og hvordan de kan bidra til å i vareta det. Mulige tiltak man kan vurdere å gjøre i sin virksomhet er:

- Ta med personvernet i opplæringsprogrammet av nye ansatte. Her kan det være smart å legge vekt på det grunnleggende, som for eksempel HVA personopplysninger er, og hvordan loven sier at de skal behandles. Dette vil bidra til at nyansatte får en forståelse av hvorfor det er viktig å tenke på samtidig som de får det nødvendige begrepsapparatet for å henge med på framtidige tiltak i kulturprogrammet.
- Heng på ledelsen og få de til å ta til seg alle tiltak. Det blir svært vanskelig å få de ansatte til å sende lenker til delte dokumenter dersom ledelsen fortsetter å sende filer som vedlegg.
- Informasjonsmøter/samlinger med personvern og sikkerhet som tema. Ta gjerne med siste nytt om nye typer trusler som er kommet, eller hendelser som har skjedd enten i virksomheten eller hos andre som man kan ta lærdom av.
- Små drypp av informasjon, enten ved oppslag på tavler, intranett eller som saker på avdelingsmøter som har med temaet å gjøre.
- Ha et godt fungerende system for å varsle om avvik og minn om dette jevnlig. Ros fungerer bedre enn ris, så sørg for å berømme de som oppdager og varsler avvik. Prøv å skape en kultur der det er åpenhet rundt feil.
- Konkurranser kan være med på å øke oppmerksomheten rundt temaet. Man kan for eksempel kåre mest sikkerhetsbevisste avdeling hvert år på julebordet, eller dele ut en pose twist eller lignende til de som finner hull og avvik og som melder om det i en periode.

Internkontroll

Det kan anbefales å ha periodiske gjennomganger av personvernet i en virksomhet. I likhet med sikkerhetsarbeid er arbeidet med personvern en kontinuerlig innsats. Hvor ofte det er behov for gjennomgang vil avhenge av hvilken type virksomhet det er og virksomhetens omfang. Slike gjennomganger vil bidra til å oppfylle kravet om internkontroll. Mer utfyllende informasjon om internkontroll og sikkerhetsarbeid finnes på difi.no.

Det anbefales å kalle inn til gjennomgang med nøkkelpersoner og gå gjennom:

- Protokollen
- Om noe i virksomheten har endret seg siden sist gjennomgang
- Nye avgjørelser/dommer. Lovtolkningen kan ha endret seg siden sist gjennomgang. Nyttige ressurser i den forbindelse er:
  - datatilsynets hjemmesider
  - Personvernrådet (European Data Protection Board - EDPB)
  - Personvernbloggen.no
- Nye ting man må ta hensyn til i omgivelsene, for eksempel nye former for trusler mot informasjonssikkerheten
- Avvik

I disse gjennomgangene kan en også drøfte tiltak for holdningsskapende arbeid sett i lys av hva som er kommet fram i gjennomgangen. Tiltak kan for eksempel være nyhetsbrev, foredrag, møter, oppslag på intranett eller interne konkurranser. Hva avvikene består i kan være en god pekepinn på hvor det er nyttig å rette innsatsen.

For å følge med på nyheter av interesse for dette arbeidet kan det være smart for alle som er involvert å abonnere på nyhetsbrevene til norSiS og datatilsynet. Dette er gratis, og man får god og oppdatert informasjon direkte i sin e-postboks.

Takk for oppmerksomheten

Takk for at du har bestøkt siden min. Til slutt vil jeg komme med et generelt råd i arbeidet med personvern: Utøv god kildekritikk!
I arbeidet med å lete etter informasjon om emnet har jeg kommet over veldig mange ulike fortolkninger av loven, og det er ikke alle som stemmer like godt. Man bør derfor tenke seg om før man tar til seg informasjon man finner, og vurdere kildens troverdighet. Det gjelder så klart denne siden også.

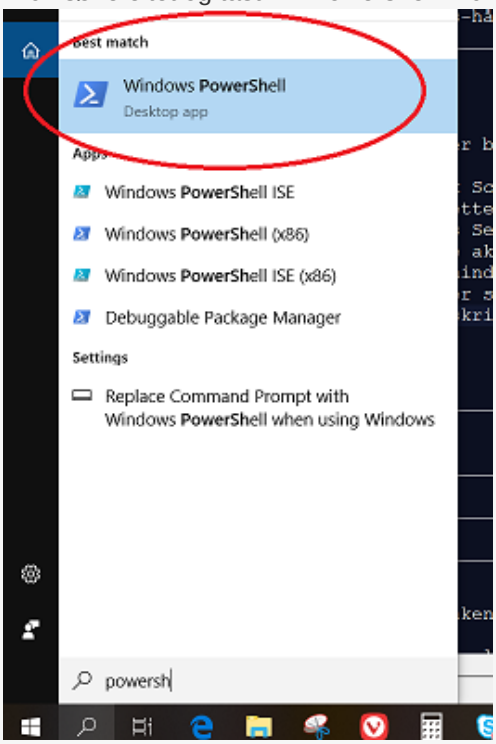
Tilbakemeldinger

Det er sikkert mange ting jeg ikke har fått med meg i denne veilederen.
I og med at jeg tror mange lurer på hvordan de skal angripe dette problemet med ustrukturerte data og GDPR lar jeg den stå litt etter at oppgaven er levert.
Dersom du har tilbakemeldinger eller idéer til ting jeg burde ta med tar jeg gjerne i mot dem på: **admin@ustrukturert.no**

Hvordan sette opp automatisk sletting av filer i en mappe i Windows 10

Steg 1: Powershell

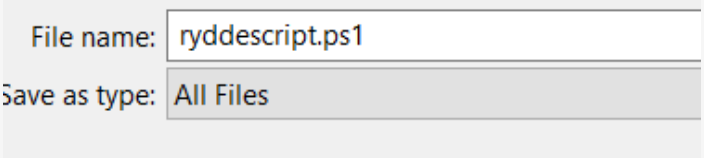
Bruk søkefeltet og tast inn Powershell. Du får opp flere alternativer, velg Windows PowerShell.



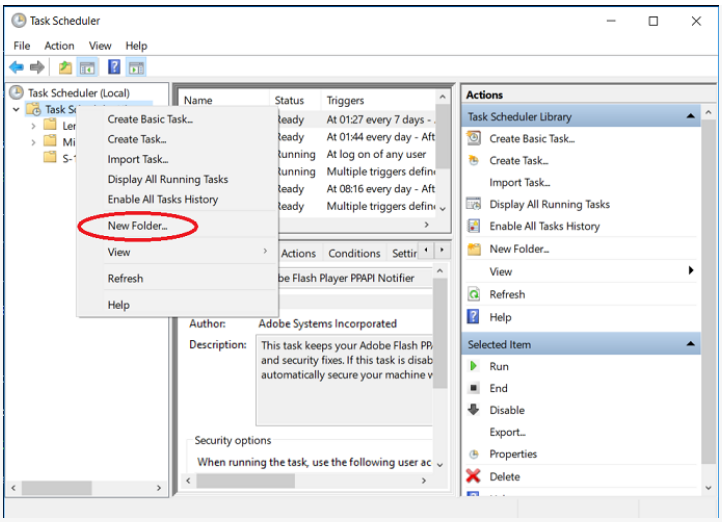
Høyreklikk og velg "run as administrator". Du får opp en dialogboks som spør om du vil tillate PowerShell å gjøre endringer på maskina, denne trykker du ja på.
Sett inn følgende poweshellkommando:
Get-ChildItem –Path "C:\sti\til\mappe" -Recurse | Where-Object {($_.LastWriteTime -lt (Get-Date).AddDays(-30))} | Remove-Item
og bytt ut C:\sti\til\mappe med sti til den mappen du vil slette fra. Dersom du ønsker en annen utløpstid enn 30 dager kan du endre på antallet i parantesen bak .AddDays. Trykk enter > du har nå sletta alle gamle filer fra mappen og fått testa scriptet ditt.

Steg 2: lage script og sette det opp til automatisk kjøring

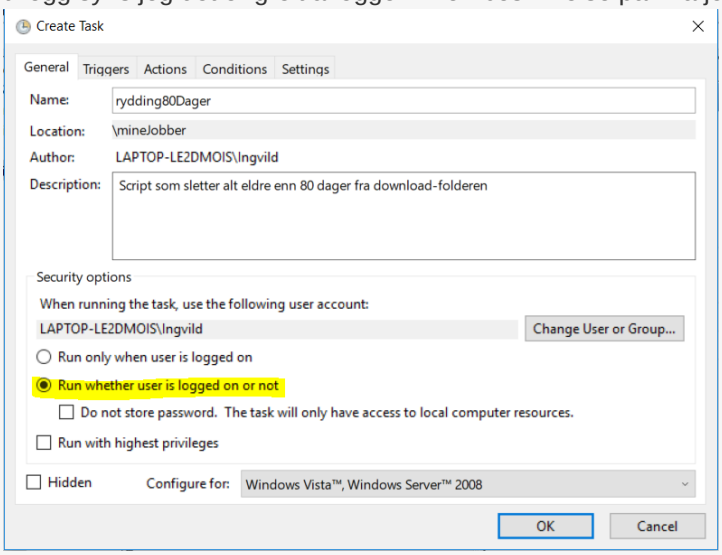
Åpne notepad, kopier poweshellkommandoen og lim den inn i en tom fil. Lagre filen som ryddescript.ps1 (viktig å få med endelsen .ps1 her!)



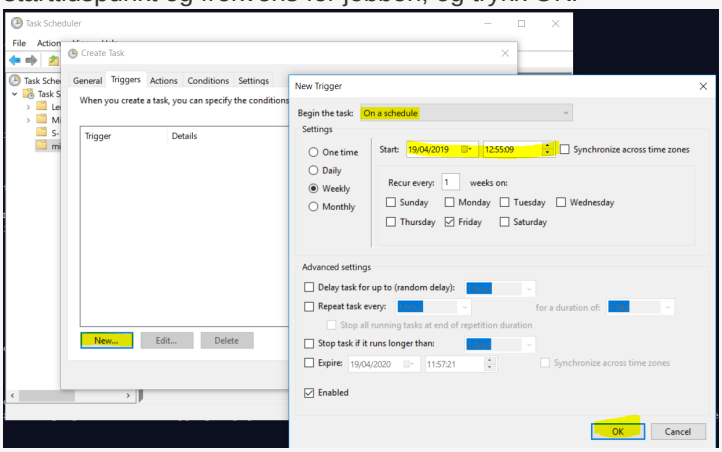
Søk opp Task Scheduler på samme måte som du gjorde med powershell og åpne den.



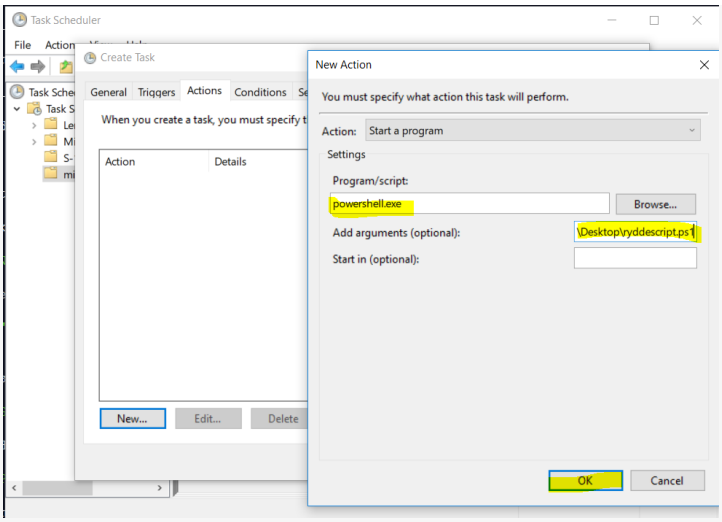
Lag en ny mappe med et navn du finner på selv, for eksempel mineJobber. Deretter høyreklikker du på mappen du nettopp lagde og velger **Create Task**. Finn på et navn til jobben og sett Run whether user is logged on or not på security options. I tillegg syns jeg det er greit å legge inn en beskrivelse på hva jobben gjør, men det er ikke påkrevd.



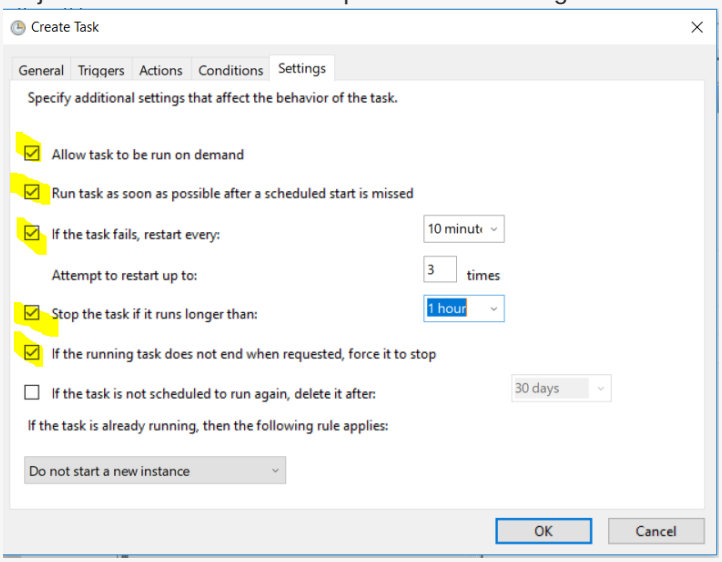
Gå deretter på Trigger-taben. Her velger du "on a schedule" for å få jobben til å kjøre på faste tidspunkt. Sett inn starttidspunkt og frekvens for jobben, og trykk OK.



Nå har du laget en hendelse som setter i gang jobben, og det er på tide å definere jobben, det vil si å legge inn så scriptet kjører på de oppsatte tidene. Gå til **Actions** og trykk på new-knappen. I menyen som kommer opp under Actions velger du "start a program". Under Program/script skriver du inn powershell.exe. I feltet for Add arguments legger du inn filbanen til scriptet ditt.



Trykk ok, og gå til settings der du velger hva som skal skje ved kjøring. Jeg syns det er best å sette max til 3 forsøk, slik at man slipper at den stjeler veldig mye ressurser hvis den henger seg opp. I og med at dette ikke er en driftskritisk jobb kan det være best å avbryte den så tidlig som mulig hvis det skjer noe som gjør at den stopper. Det bør ikke ta mange sekunder å kjøre denne så sett max tid på det minste mulige.

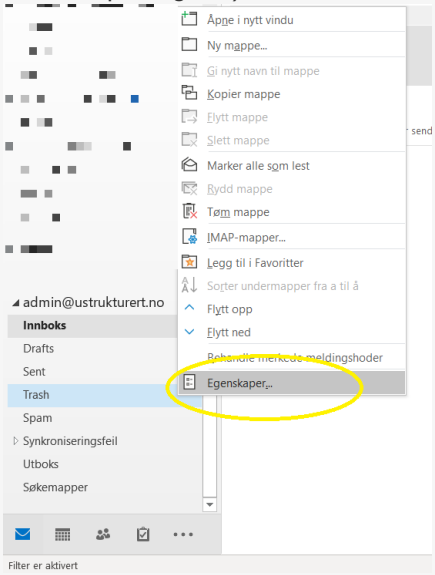


Når du trykker ok her får du beskjed om å skrive inn admin-passord. Når dette er gjort er jobben lagret og kommer til å kjøre på de tidene du har satt opp.

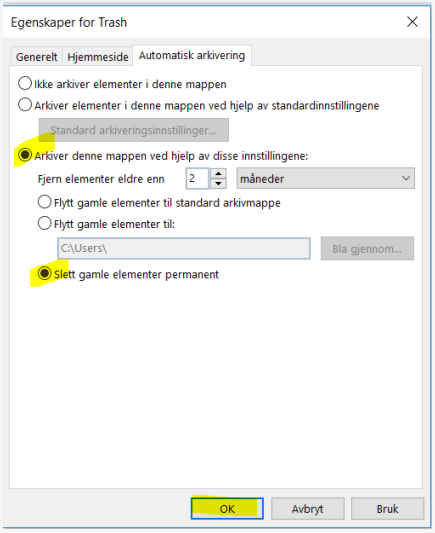
Tilbake

Bruke AutoArchive til å slette gamle e-poster i outlook

Høyreklikk på den mappen du ønsker å sette på autosletting på, og velg egenskaper (properties hvis du har outlook installert på engelsk).

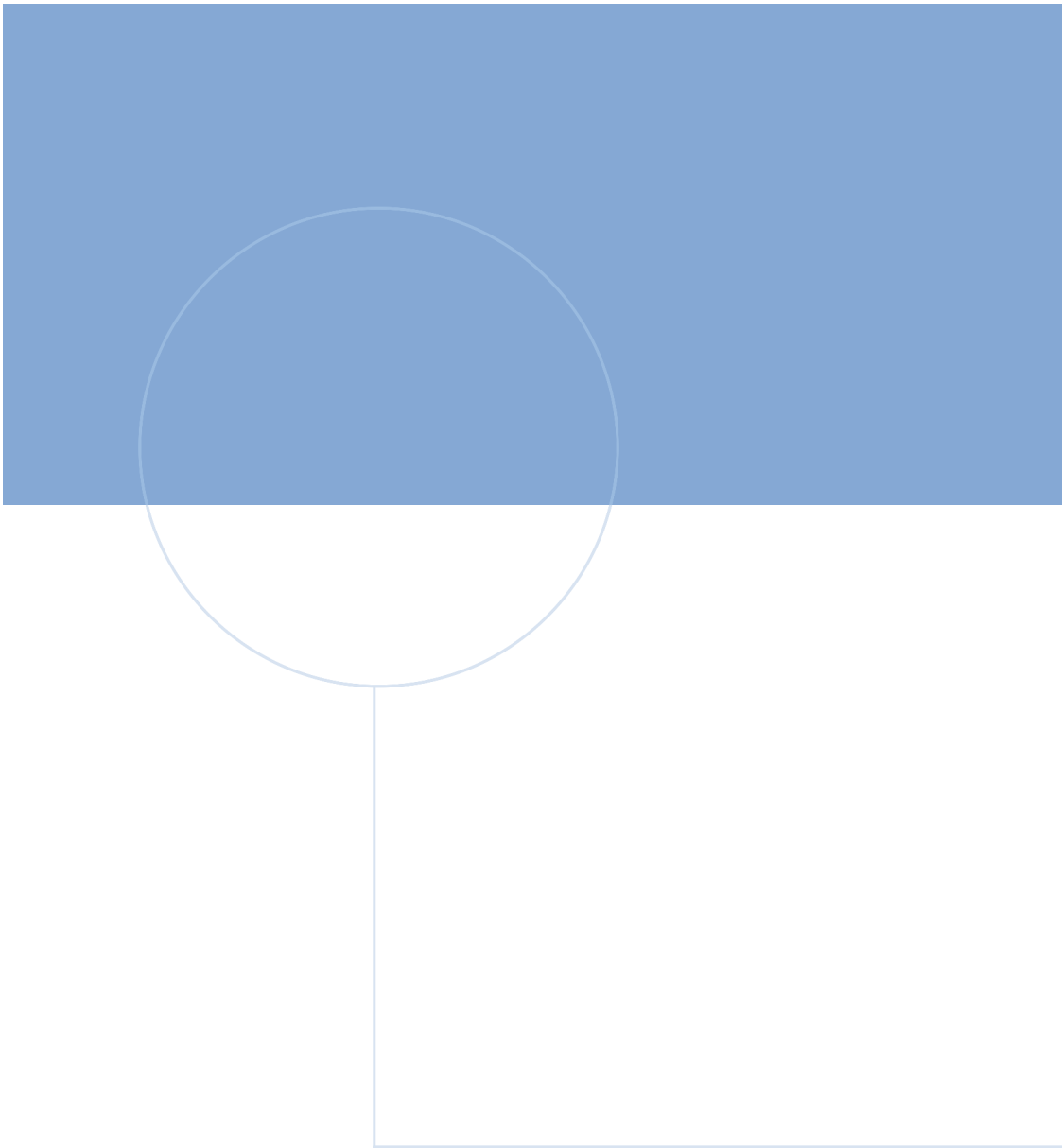


Gå til taben automatisk arkivering (auto archiving på engelsk) og velg hvor lenge du vil at ting skal kunne ligge før det slettes. Huk av for å slette gamle elementer permanent.



Trykk på OK og du er ferdig; alle e-poster i mappen vil slette seg selv etter den tiden du har valgt.

Tilbake



NTNU

Kunnskap for en bedre verden