



Whitepaper

Van handmatig naar automatisch anonimiseren

Introductie

Juristen en informatiespecialisten van bedrijven en overheidsinstanties behandelen elke dag enorme hoeveelheden gevoelige informatie waaronder zowel persoonlijk identificeerbare informatie (PII) als beschermde gezondheidsinformatie (PHI), intellectueel eigendom, bedrijfsgeheimen, financiële informatie en nog veel meer.

Tegelijkertijd zijn zij vaak verplicht om informatie te verstrekken aan andere juristen, de rechtbank, toezichthouders en, onder bepaalde omstandigheden, journalisten en burgers die verzoeken om hun persoonlijke dossiers of overheidsdocumenten indienen.

Anonimiseren - het verbergen van tekst - is het middel waarmee juridische teams of informatieafdelingen vertrouwelijke informatie verwijderen uit openbaar te maken documenten. Er zijn twee grote uitdagingen rond het anonimiseren: het efficiënt identificeren van de vertrouwelijke informatie die mogelijk verborgen zijn in stapels openbaar te maken informatie en het verwijderen ofwel lakken van die informatie voorafgaand aan het openbaar maken of overdragen.



Figuur 1. voorbeelden van geanonimiseerde documenten

Wat is vertrouwelijke informatie?

Vertrouwelijke informatie is informatie die moet worden beschermd omdat dit persoonlijke gegevens zijn, vertrouwelijk, geprivilegieerd of anderszins geheim is. Over het algemeen kan vertrouwelijke informatie zijn:

- persoonlijke gegevens over een persoon, die persoonlijk identificeerbare informatie (PII) kunnen zijn, zoals een naam, geboortedatum of burgerservicenummer of beschermde
- gezondheidsinformatie (PHI) zoals een medische diagnose; bevoorrechte informatie die wordt beschermd onder het verschoningsrecht van de advocaat-cliënt, als werkproduct van de advocaat of via een ander type privilege; of
- vertrouwelijke informatie die betrekking heeft op interne organisatiestrategie, intellectueel eigendom, bedrijfsgeheimen of andere beschermde informatie.

Vertrouwelijke informatie komt vaak voor in de context van documenten die openbaar worden (of zelfs moeten worden) gemaakt. Die openbaarmaking kan plaatsvinden in de context van de Wet Open Overheid, eDiscovery zaken, juridische geschillen of elders in de loop van een (mogelijke) rechtszaak.

Wat is anonimiseren of lakken?

Anonimiseren is het proces waarbij gevoelige informatie volledig wordt verwijderd uit informatie die een organisatie verlaat en wordt overgedragen aan een andere partij. Wanneer een ontvanger recht heeft op het ontvangen van gegevens die ook informatie bevatten die hij niet mag zien, moeten die gegevens worden bewerkt om de gevoelige informatie in die gegevens te beschermen. Redacties (geanonimiseerde delen van ene document) verschijnen over het algemeen als zware zwarte vlakken over afzonderlijke woorden of cijfers of, in het geval van uitgebreidere redacties, balken die tekstregels verbergen.

Anonimiseren kan een tijdrovende, verzwarende en foutgevoelige taak zijn. Het is het typische naald-in-een-hooiberg dataprobleem: juridische teams moeten pagina's met niet-gevoelige informatie ontleden om de kleine stukjes gevoelige informatie te detecteren - namen, bedragen, identificatienummers en meer - die erin verborgen kunnen zijn.

Maar het vinden van gevoelige informatie is slechts de eerste uitdaging bij het anonimiseren: juridische teams moeten die informatie ook grondig opschonen, zodat deze op geen enkele manier kan worden ontdekt.

Je verhoogt het risico dat per ongeluk gevoelige informatie openbaar maakt wanneer er niet voldoende aandacht besteed wordt aan beide aspecten. Het gebruik van inefficiënte handmatige processen of verouderde technologieën om gevoelige informatie te lokaliseren, kan ertoe leiden dat stukjes informatie volledig gemist worden, zodat een redactie niet wordt toepast.

Evenzo kan het gebruik van onjuiste en ineffectieve methoden ertoe leiden dat informatie geproduceerd wordt die geïdentificeerd is en waarvan is gedacht dat deze geanonimiseerd was, maar die nog steeds kan worden achterhaald.

Wetgeving

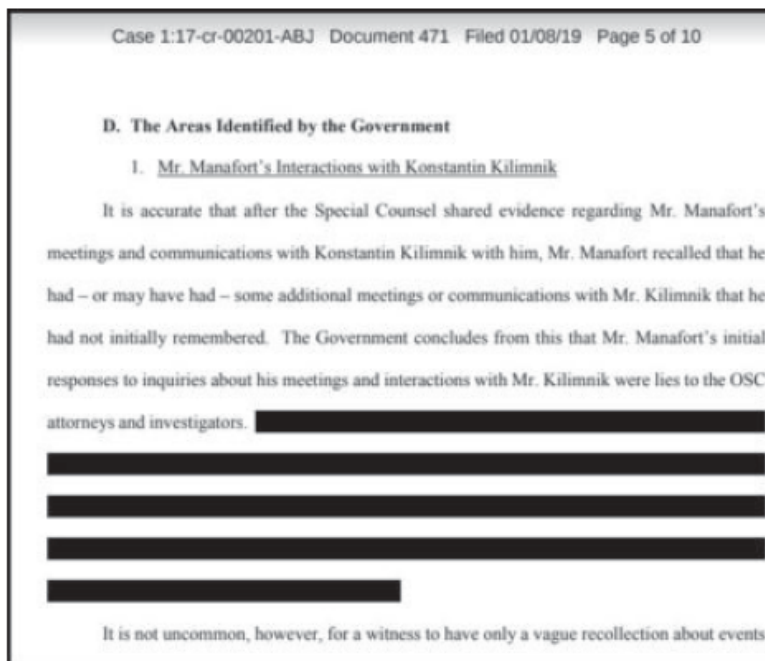
Natuurlijk anonimiseren wij informatie om vertrouwelijke gegevens te beschermen maar er zijn ook wettelijk vastgestelde regels voor het omgaan met informatie en met name bij het publiceren of overdragen van informatie aan andere partijen zijn er diverse regels.

Denk aan de Algemene Verordening Gegevensbescherming (AVG), de Wet Open Overheid (Woo), voorheen de Wet Openbaarheid van Bestuur (Wob), maar ook de Wet Financieel toezicht (Wft) en General Data Protection Regulation (GDPR). Ieder land heeft zijn eigen wetten en regels die van toepassing kunnen zijn voor het anonimiseren van informatie.

Kan het misgaan?

In de Verenigde Staten kan er via de Freedom of Information Act (FOIA) verzoek worden gedaan publieke informatie vrij te geven. Ook hier gelden bepaalde uitzonderingsgronden, waardoor informatie moet worden afgeschermd. Zonder gebruik van de juiste tools is anonimiseren en weglakken een van de meest tijdrovende klussen in een FOIA-proces. En wanneer er bij het anonimiseren fouten worden gemaakt, heeft dit vaak grote consequenties.

Tijdens de openbaarmaking van het Mueller-rapport werd dit temeer duidelijk. Speciaal aanklager Robert Mueller leverde in april 2019 het zogenoemde Mueller rapport af. Dit was een rapport dat onderzoek deed naar de vermeende Russische inmenging in de Amerikaanse presidentsverkiezingen van 2016. In het rapport werd de campagnevoorzitter van Donald Trump, Paul Manafort, beschuldigd van het delen van kiezersdata met Konstantin Kilimnik, volgens de FBI een Russisch geheim agent, die een lijntje had met het Kremlin.



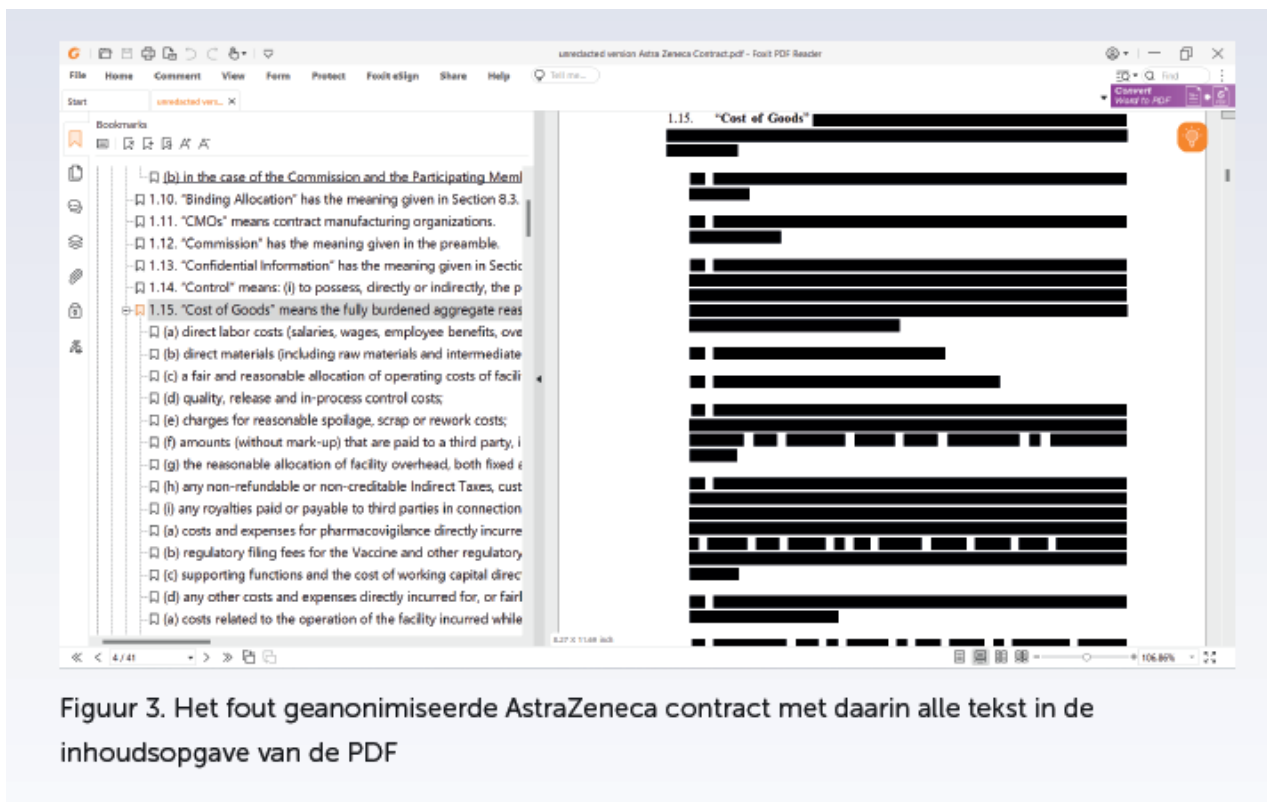
Figuur 2. PDF document dat geanonimiseerd leek, maar waar je tekst eenvoudig kon achterhalen

'De advocaten van Manafort kwamen met een officieel antwoord, waarin de beschuldigingen werden ontkend. Een deel van het antwoord was afgelakt voor de buitenwereld. Afgelakt, totdat een verslaggever van de krant the Guardian het in handen kreeg en simpelweg de inhoud van het PDF-document kopieerde en plakte in een nieuw document. Omdat de het gelakte gedeelte alleen op cosmetisch op de PDF was gezet, was de tekst erachter nog steeds relatief eenvoudig leesbaar te maken. Uit de voorheen weggelakte delen bleek wel degelijk dat Manafort meerdere ontmoetingen had gehad met Konstantin Kilimnik, wat veel stof heeft doen opwaaien.

In het door de advocaten van Manafort gepubliceerde pdf-bestand was een zwarte balk over de af te lakken tekst heen geplakt. Wie voor deze oplossing koos, stond er niet bij stil dat deze vorm van beveiliging simpel te omzeilen is: de originele tekst is namelijk niet weg. De tekst achter het zwarte balkje is gewoon te kopiëren en te plakken, en dus volledig leesbaar te maken ondanks de poging tot aflakken.

In januari 2021 was het in Europa weer eens raak: journalisten vonden in een gelakt document toch nog informatie die verborgen had moeten blijven. Dit keer was het de Europese Commissie, die naar aanleiding van de commotie over het leveren van de COVID-19 virus vaccins het bewuste contract met producent AstraZeneca op Internet plaatste.

In het bewuste contract was veel informatie netjes gelakt, echter was men vergeten ook de bookmarks-sectie van het document na te kijken. Deze bookmarks (of bladwijzers) in een PDF worden automatisch gegenereerd aan de hand van de gebruikte kopteksten: ook wanneer de tekst in de koptekst zelf is gelakt, is dit niet het geval in de bookmarks.



Figuur 3. Het fout geanonimiseerde AstraZeneca contract met daarin alle tekst in de inhoudsopgave van de PDF

Helaas gebeurt het vaker dat gebruik van “lak-tools”, zoals Adobe Acrobat, voor problemen zorgen. Het kan voorkomen dat er informatie ongewild wordt gepubliceerd, ook in Nederland. Het heeft wel eens geleid tot het uitlekken van staatsgeheimen, met alle gevolgen van dien. Dit komt omdat in veel organisaties nog standaard software wordt gebruikt voor het aflakken en niet alle gebruikers de juiste stappen doorlopen.

Als gevolg worden organisaties steeds voorzichtiger, juist om lakfouten te voorkomen: er zijn organisaties die na het lakken in Adobe alle documenten printen en op papier distribueren, of worden dezelfde documenten na het printen nogmaals gescand naar PDF en alsnog digitaal vrijgegeven. Extra en onnodig werk als er op de juiste manier gewerkt wordt.

Besparingen

Kunnen we bepalen wat de kosten zijn van anonimiseren? Helaas zijn er weinig publieke en specifieke cijfers beschikbaar. Dat is natuurlijk ook lastig omdat er veel factoren meespelen, zoals de inzet van medewerkers (Beleid, Juridische zaken, etc.), de kosten van advocaten en de kosten van software licenties. Met name de inzet van medewerkers is lastig te meten, maar als je hoort dat iemand met het afhandelen van een Woo-verzoek (lees: anonimiseren) soms maanden bezig is, dan kan dat aardig oplopen qua arbeidskosten.

Gelukkig zijn er een aantal artikelen verschenen die meer inzicht geven in de kosten van het anonimiseren en het afhandelen van Wob/ Woo verzoeken.

Centrale Commissie Dierproeven: €1,2 miljoen in 4 jaar

Het bij elkaar vinden van 30.000 documenten en het aanstellen van een extra jurist was voor de Centrale Commissie Dierproeven (CCD) nodig om te voldoen aan de Wob-verzoeken die bij het overheidsorgaan binnenkwamen. Het grootste aantal Wob-verzoeken werd ingediend door twee fanatieke dierenliefhebbers. In totaal zijn er in drie jaar tijd zo'n 131 Wob-verzoeken door hen ingediend.

Het kostte de CCD van 2015 tot en met 2018 gemiddeld €315.000 per jaar om te voldoen aan dit aantal Wob-verzoeken. Aan dwangsommen en proceskosten heeft de CCD al bijna €30.000 moeten overmaken aan het tweekant. Volgens de CCD-secretaris was hier sprake van 'misbruik van recht'.

Bron: <https://www.telegraaf.nl/nieuws/1381860859/onvrede-over-wobterreur-dierenactivisten>

Nederweert: advocaat alleen al kost €115.000

De gemeente Nederweert heeft onlangs te maken gehad met een 'Wobmisbruiker'. Dat heeft de Raad van State in haar uitspraak op 6 november 2019 geconcludeerd. De gemeente werd overladen met 96 Wobverzoeken, 35 Wbp-verzoeken en 11 verzoeken op grond van de AVG. De gemeente Nederweert heeft alleen al vanwege deze verzoeken €115.000 moeten uitgeven aan advocaatkosten.

Bron: <https://www.1limburg.nl/ontslagen-ambtenaar-trolt-nederweertvoor-ruim-een-ton>

Lisse: Sociale Dienst huurt een externe kracht ingehuurd voor Wobverzoeken

Een ander voorbeeld is de Sociale Dienst in Lisse. Een burger diende in alleen al in 2018 60 Wob-verzoeken, klachten en bezwaren in. De sociale dienst in Lisse heeft een externe kracht ingehuurd om alle Wob-verzoeken van de burger te behandelen.

Bron: <https://www.omroepwest.nl/nieuws/3714782/Superklager-voor-derechter-gesleept-door-sociale-dienst>

Er zijn ongetwijfeld meer cijfers beschikbaar bij diverse overheidsorganisaties, en in gesprekken hoor je wel eens schattingen die variëren van duizenden euro's voor kleine Wob/Woo-verzoeken tot meer dan honderdduizend euro voor één enkel groot en complex Wob/Wooverzoek.

Waar moet je op letten als je software gebruikt om te anonimiseren?

Weglakken, aflakken, witlakken, zwartlakken, voorlakken, lakken, anonimiseren, redacten, black-linen, redigeren allemaal termen die gebruikt worden om informatie zoals persoonlijke informatie, telefoonnummers, en andere informatie weg te halen uit documenten.

Er zijn veel verschillende softwareoplossingen en tools beschikbaar die meerdere of enkele documenten kunnen lakken met de hand of (deels) automatisch.

Wetgeving en verantwoording

Bij het anonimiseren van informatie is het belangrijk dat duidelijk is wat de reden is voor het anonimiseren. Het moet voor de ontvanger duidelijk en transparant zijn wat de reden is waarom bepaalde delen informatie geweigerd is. Door middel van weigeringsgronden of uitzonderingsgronden wordt normaal gesproken op het geanonimiseerde gedeelte de reden weergegeven.

Wat is wel of niet openbaar?

E-mailadressen worden bij een Woo besluit (bijna) altijd weggehaald, om de persoonlijke informatie weg te halen. Echter, het kan zo zijn dat de domeinnaam van het adres wél openbare informatie is of kan zijn. Alle e-mailadressen met de hand lakken bij een groot Woo verzoek is eigenlijk onbegonnen werk. Met behulp van automatisering gaat dat veel makkelijker, door met het juiste programma alle e-mailadressen te vinden in de documenten en deze weg te lakken.

Enige nadeel: in veel gevallen wordt dan het volledige e-mailadres wordt weggehaald. In een dergelijk geval moet er vervolgens met de hand een correctie plaatsvinden. Dat is natuurlijk veel extra werk, daarom is het belangrijk dat email adressen in één keer kunnen worden geanonimiseerd waarbij de domeinnaam zichtbaar blijft.



To: [redacted] - BD/DRB/RG [redacted] @minjenv.nl; [redacted] - BD/DRB [redacted] @minjenv.nl
Cc: [redacted] - BD/DRB/RG [redacted] @minjenv.nl; [redacted] - BD/DRB/RG [redacted] @minjenv.nl; [redacted] - BD/DRB/RG [redacted] @minjenv.nl; [redacted] - BD/DRB/RG [redacted] @minjenv.nl
From: [redacted] - BD/DRB/TR
Sent: Fri 3/12/2021 11:44:15 AM
Subject: RE: inschrijffvoorwaarden RvR
Received: Fri 3/12/2021 11:44:16 AM

Hoi [redacted],
Dank voor je terugkoppeling. Ik heb met [redacted] gesproken dat ik zal aanschuiven bij een gesprek met de Raad en

Figuur 4. Een email header met gelakte email adressen to het @-teken

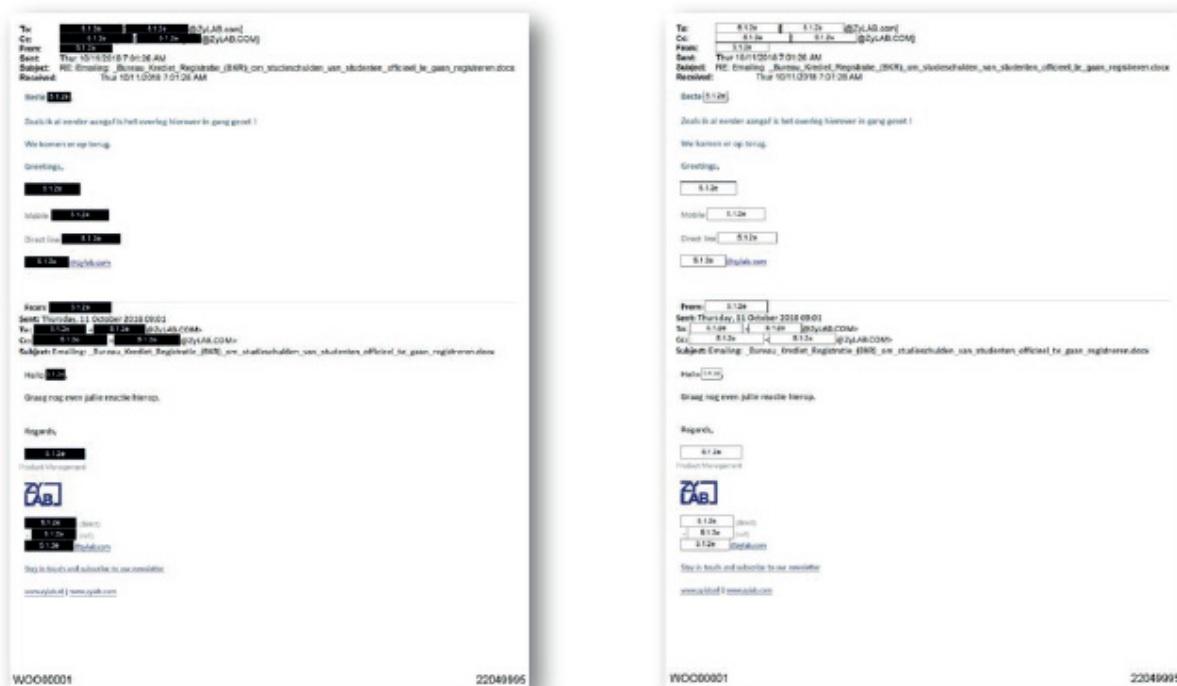
Pagina's weigeren of anonimiseren

Niet alleen namen en persoonlijke gegevens mogen worden geweigerd bij het overdragen van informatie, soms moeten volledige documenten, pagina's of pagina reeksen worden geweigerd. Het kan zijn dat binnen een document van 100 pagina's slechts 10 pagina's relevant zijn en er 90 worden geweigerd. Het is dan noodzakelijk om dit duidelijk en transparant te maken, maar ook de afhandeling simpel en eenvoudig.

Transparantie en Reputatie

Hoe breng je overheidsdocumenten naar buiten waarvan een deel als 'niet openbaar' onder de reikwijdte van een Woo-verzoek blijft vallen? De beeldvorming over een transparante overheid wordt er niet beter op als in documenten tientallen pagina's zijn zwartgelakt.

Figuur 5. Zwart of wit lakken? Wat oogt vriendelijker?



Figuur 5. Zwart of wit lakken? Wat oogt vriendelijker?

De Kinderopvangtoeslagenaffaire is letterlijk één grote zwarte bladzijde te noemen. Pagina na pagina was zwartgelakt in de openbaar gemaakte overheidsdocumenten over de affaire. Het leidde onder meer tot een verzoek van de Tweede Kamer om de lakinstructies openbaar te maken. En het inspireerde Arjan Lubach tot een onthullend filmpje over Wob-procedures waarbij in documenten soms wel 90 pagina's waren weggelakt.

Door anders te gaan lakken, namelijk met gekaderd-wit, kan de beeldvorming veranderen. Diverse organisaties hebben nu afgesproken om deze lak-methode standaard te gaan gebruiken, met name omdat het veel vriendelijker oogt en besparingen oplevert bij het printen.

Ondersteuning van de workflow van het anonimiseren

Gaat het alleen om het anonimiseren van documenten? Nee, anonimiseren van informatie is een heel proces. Het begint met het identificeren van documenten die geanonimiseerd moeten worden, deze documenten moet daarna ingelezen worden, ongeacht het formaat. Vervolgens moeten de documenten automatisch geanalyseerd worden en de gebruiker helpen te ontdekken wat er geanonimiseerd moet worden. Het systeem moet documenten automatisch (voor)lakken. De gebruiker moet simpel en eenvoudig zelf kunnen lakken en/of aanpassingen kunnen maken. Meerdere personen moeten toegang hebben tot de documenten en elkaar kunnen aanvullen en controleren. Het moet eenvoudig zijn om het werk te verdelen.

Resultaat

Uiteindelijk moeten alle documenten veilig en doorzoekbaar worden gepubliceerd. Dus alle gelakte vlakken moeten ook echt weg zijn, maar alle overige informatie moet wel doorzoekbaar zijn voor zoek engines op het Internet. Voordat het resultaat wordt vrijgegeven of gepubliceerd moet je een kwaliteitscontrole kunnen doen die verdedigbaar en geregistreerd is.

Verzamelen en verwerken van documenten om te anonimiseren

Alleen het verzamelen van documenten voor anonimiseren is vaak niet genoeg om documenten goed te kunnen beoordelen. Veelal moeten de verschillende documenten worden geanalyseerd en op de juiste manier worden verwerkt om deze bruikbaar te maken voor het anonimiseren. Met name als je gebruikt wilt maken van meer automatisering, dan is het een voordeel als alle documenten zowel individueel als in relatie tot elkaar bijvoorbeeld volledig doorzoekbaar zijn onafhankelijk van het type document.

Dit helpt bij het selecteren van de documenten die onder het verzoek vallen, je kan op de inhoud van het documenten zoeken en daarnaast ook bij de analyse van deze documenten door persoonsnamen en andere gegevens te herkennen.

Vaak zijn documenten die uit het “archief”, zaakstelsel of document managementsysteem direct bruikbaar om te verwerken. Dit zijn dan nette Microsoft Word documenten of PDFdocumenten die zelf zijn gemaakt door bijvoorbeeld een e-mail of Microsoft Word document om te zetten naar PDF. Maar het komt met regelmaat voor dat documenten niet altijd geschikt zijn om direct gebruikt te worden. Deze documenten hebben een extra verwerking nodig om ze geschikt en leesbaar te maken en deze te kunnen anonimiseren.

PDF

PDF is een alom geaccepteerd formaat en wordt veel gebruikt. PDF-documenten kan je zelf makkelijk maken door bijvoorbeeld een Microsoft Word document op te slaan als PDF. Je hebt dan direct een mooie doorzoekbare PDF die iedereen kan lezen. Maar het gebeurt ook vaak genoeg dat PDF documenten van een derde partij komen en dat het op het eerste gezicht niet duidelijk is of dit document wel doorzoekbaar is. Iedereen kent wel nu wel de mogelijkheid om een document te scannen op kantoor en deze vervolgens in de mail te ontvangen. De meeste apparaten die hiervoor gebruikt worden, maken een keurige doorzoekbare PDF. Maar oudere apparaten doen dat vaak niet.

Je krijgt dan wel een PDF, maar deze is eigenlijk alleen terug te vinden op de bestandsnaam. De inhoud van deze documenten is niet doorzoekbaar. Door de huidige thuiswerksituatie als gevolg van COVID-19 gebeurt het nu vaker dat documenten even met de telefoon worden “gescand”. Vaak even een foto maken van het document of via een “gratis” documenten scanner app even scannen en doorsturen. Veel van deze documenten zijn niet doorzoekbaar en indien deze worden opgenomen in het “archief” zijn deze moeilijker terug te vinden.

In de praktijk is het wel eens voorgekomen dat documenten zoals contracten met een derde partij, een leverancier van een overheidsinstantie, werden opgenomen in het “archief”. Tijdens een onderzoek naar inkoop en leveranciers werden contracten opgezocht in het “archief” waarbij de desbetreffende documenten van die ene leverancier niet werden gevonden doordat deze alleen plaatjes van het document bevatten en geen doorzoekbare tekst. Op een later moment is dit wel achterhaald door gebruik te maken van de juiste tooling.

Tip: Met de gratis App Microsoft Office Lens maak je makkelijk scans en doorzoekbare PDF

Embedded (Ingesloten) Objecten en Bijlagen

Tja, wat is de juiste Nederlandse naam hiervoor, een Embedded Object is een bestand dat is ingesloten in een ander bestand. Een voorbeeld is een plaatje in een Microsoft Word document. Dat is daardoor een object in een document geworden. Dit is nog vrij gangbaar en komt vaak voor. Maar het kan ook gebeuren dat er een PDF document wordt ingesloten in een Microsoft Word document. In de praktijk komen wij wel eens bijzondere documenten tegen zoals bijvoorbeeld een Excel sheet van 50MB.

Dat is natuurlijk een hele grote Excel sheet, maar in plaats van heel veel cijfers, bevatte deze Excel sheet 40 PDF documenten als bijlagen (ingesloten objecten). De eigenaar van het document gebruikte de Excel om zijn PDF documenten bij elkaar te houden. Erg lastig om dit type documenten direct te verwerken voor een Wob of Woo verzoek, dus ook hiervoor zijn extra bewerkingen nodig.

Andere voorbeelden hiervan zijn:

- Microsoft Word documenten met Microsoft Excel tabellen;
- PDF documenten met bijlages
- E-mails met plaatjes in de tekst en niet als bijlage
- E-mail berichtenreeksen of conversaties
-

Meer en meer zijn het niet alleen de documenten uit het archief die moeten worden geanonimiseerd, maar met name e-mail maakt een steeds groter deel uit van alle documenten die geanonimiseerd publiek gemaakt moeten worden. Soms tot wel 70% van de informatie komt oorspronkelijk uit e-mail. E-mail heeft natuurlijk bijlages in allerlei variaties. Een e-mail kan bijlages hebben zoals Microsoft Word en PDF, maar bijvoorbeeld ook een RAR of ZIP bestand met daarin allerlei documenten verpakt.

De relatie tussen de e-mail en bijlages moet natuurlijk bewaard blijven tijdens de beoordeling en alle verschillende type documenten moeten bekeken kunnen worden. Een geautomatiseerde verwerking waarbij hieraan wordt voldaan kan veel tijd en “gedoe” besparen.

Daarnaast zijn er ook nog de zogenaamde berichtenreeksen of conversaties van opeenvolgende reacties op eerder e-mails. Het inzichtelijk maken hiervan maakt het mogelijk om zo snel te zien welke e-mails volledig zijn en welke niet. Visualisatie van deze berichtenreeksen kan voor de beoordelaars veel tijd schelen en kan dubbel werk bij het anonimiseren voorkomen.

ZyLAB ONE kan helpen bij het verzamelen en verwerken van documenten om deze geschikt te maken om te anonimiseren en beschikt over veel functionaliteit, zoals:

- Eenvoudig uploaden van enkele tot meerdere documenten
- Verwerken van meer dan 700 verschillende type bestanden
- Analyseren van e-mail bestanden
- Dedupliceren van documenten
- Herkenning van soortgelijke documenten
- Omzetten van plaatjes naar doorzoekbare documenten door middel van OCR (Optical Character Recognition)
- Omzetten van audio en video naar tekst (Speech to Tekst)
- Documenten direct uit bronsystemen halen zoals Microsoft 365 en Google Workspace
- Herkennen van verborgen informatie/documenten (embedded objects)

Anonimiseren met ZyLAB ONE

Zonder gebruik van de juiste tools, is anonimiseren één van de meest tijdrovende klussen bij het publiceren of exporteren van vertrouwelijke informatie. ZyLAB ONE is een totaaloplossing dat helpt bij o.a. het snel en efficiënt anonimiseren van informatie.

Hieronder beschrijven we de verschillende methoden hoe je snel met ZyLAB ONE kan anonimiseren.

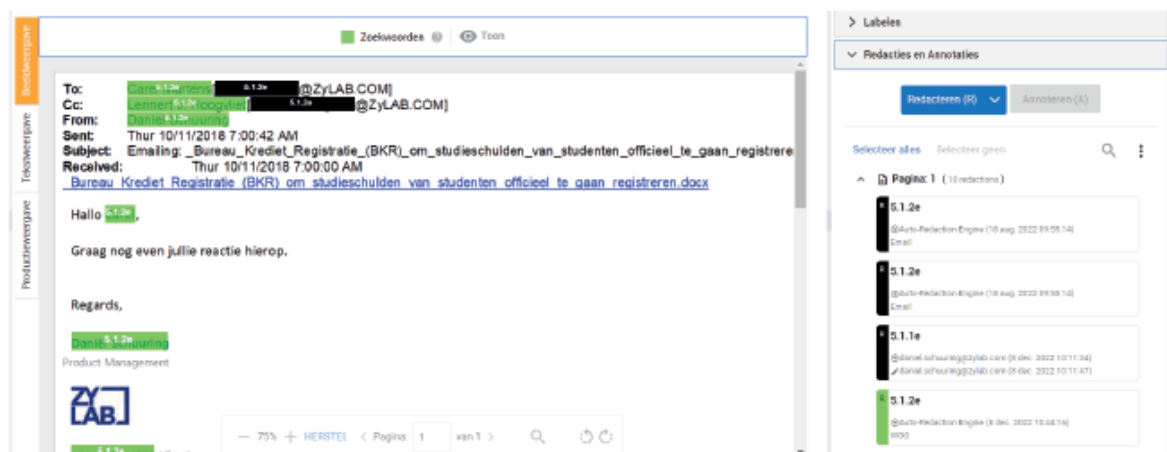
- Handmatig anonimiseren
- Semi-automatisch anonimiseren
- Automatisch anonimiseren
- Anonimiseren of weigeren van documenten of pagina's

Handmatig anonimiseren

Met de hand anonimiseren of wel het “tekenen” van zwarte vlakken op een document is de traditionele manier van anonimiseren en een zeer tijdrovend proces. Document voor document moet worden nagelezen en de beoordelaar moet goed opletten en nauwgezet lezen om de gegevens te vinden die moeten worden gelakt.

ZyLAB ONE maakt dit proces van handmatig anonimiseren sneller en efficiënter met de verschillende mogelijkheden zoals:

- Lakken met standaard instellingen zoals de weigeringsgrond
- Snel kunnen selecteren van één of meerdere uitzonderingsgronden
- Snel toepassen van al eerdere opmaak en uitzonderingsgronden van eerdere redacties
- Meerdere redacties kunnen selecteren en tegelijkertijd aanpassen
- Naast de weigeringsgrond ook de kleur van de redactie aanpassen om zo snel bepaalde type redacties te kunnen herkennen
- Anonimiseren in eerste instantie transparant te kunnen doen zodat snel te zien is wat er geanonimiseerd is. Ook handig voor tweede beoordelaar. Bij publicatie wordt de redactie permanent gemaakt en de teksten definitief weggehaald.
- Het aanmaken van indicatoren waar mogelijk geanonimiseerd moet worden, door bijvoorbeeld overal de woorden “de heer” of “mevrouw” te markeren.
- Het lakken van volledige documenten, pagina's of pagina reeksen die later vervangen kunnen worden door enkele pagina met de uitzonderingsgronden.



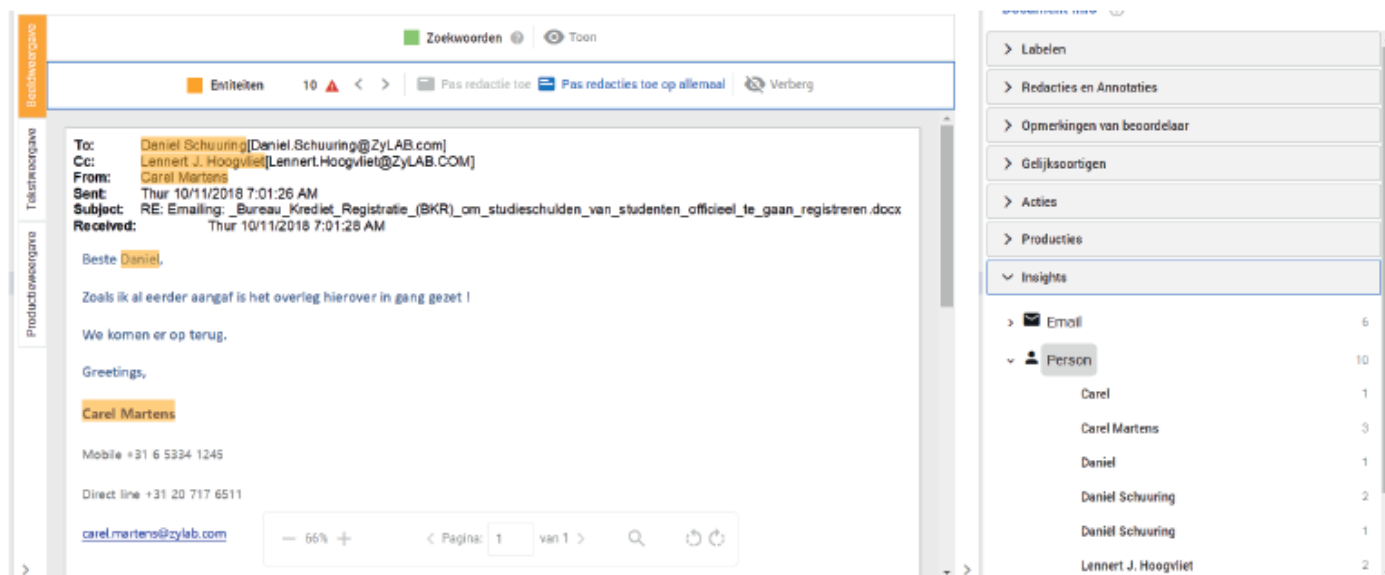
Figuur 7. Document dat gelakt is met verschillende kleuren, email adressen in het zwart en namen in het transparant-groen. Het document is nog niet gepubliceerd.

Semi-automatisch anonimiseren

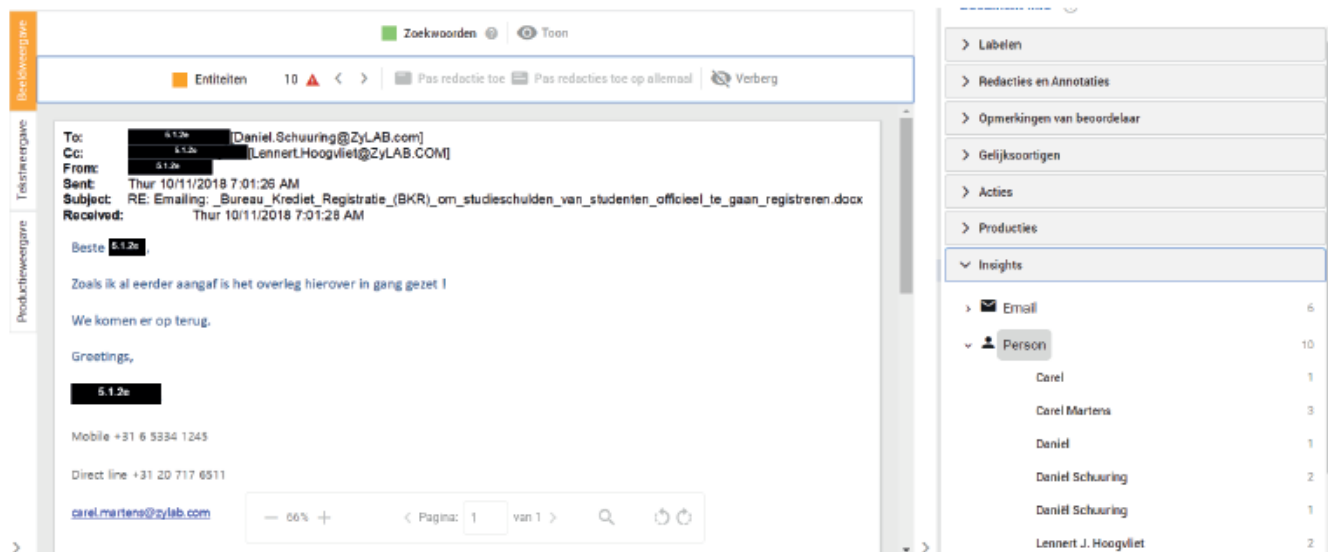
Volledig handmatig lakken is een tijdrovende bezigheid. ZyLAB ONE kan helpen bij het identificeren van teksten die geanonimiseerd moeten worden door deze tekst te markeren. Vervolgens kan deze markering simpel worden omgezet in een redactie. ZyLAB ONE geeft de gebruiker dus hints welke informatie privacygevoelig is.

Door alle documenten en teksten te analyseren door gebruik te maken van specifieke technieken kan de ZyLAB ONE o.a. persoonsnamen, organisaties, burgerservicenummers, telefoonnummers en functie titels detecteren en markeren. Dit kan heel veel werk schelen en eventueel ook fouten voorkomen. Het blijft echter belangrijk om de documenten te controleren omdat de gebruikte technologie soms niet alle gegevens kan herkennen. Maar het neemt wel een groot gedeelte van het werk uit handen.

Om de te anonimiseren gegevens automatisch te herkennen en hints te kunnen geven maakt ZyLAB ONE gebruik van verschillende technieken en regels.



Figuur 8. Automatische herkenning van te anonimiseren gegevens. Dit kan in één keer geanonimiseerd worden.



Figuur 9. Met één druk op de knop zijn de entiteiten geanonimiseerd

Zoekvragen

Door middel van eenvoudige of meer complexe zoekvragen kunnen gegevens waarvan verwacht wordt dat die in de documenten voorkomen worden herkend en gemarkeerd. Denk hierbij aan namen van medewerkers in verschillende schrijfwijzen of functie titels die al bekend zijn. In ZyLAB ONE kunnen ook zeer lange zoekvragen worden gebruikt en dus heel veel namen in een keer worden gevonden en gemarkeerd.

Patronen

Patronen zijn eigenlijk vaste schrijfwijzen voor specifieke gegevens. Een voorbeeld is een IBAN-rekening. Deze rekeningnummers hebben een vast patroon dat eenvoudig door de computer herkend kan worden door gebruik te maken van een regel. We noemen dit een reguliere expressie ook wel RegEx genoemd. Andere voorbeelden zijn telefoonnummers, creditcardnummers, postcodes en andere gegevens die een vast patroon hebben.

Kunstmatige intelligentie

Naast het zoeken naar gegevens en het opstellen van patronen, kunnen we ook ZyLAB ONE zelf laten herkennen waar informatie staat die we moeten anonimiseren. Daarvoor maakt ZyLAB ONE gebruik van kunstmatige intelligentie. Door ZyLAB ONE te leren wat bijvoorbeeld persoonsgegevens zijn, kan ZyLAB ONE zelfstandig persoonsgegevens zoals namen herkennen in documenten. Dat leren is al gedaan door het Data Science team van ZyLAB en er zijn zogenaamde AI-modellen ontwikkeld voor verschillende talen waaronder het Nederlands. Deze modellen worden regelmatig aangepast en extra getraind om zo de herkenning te verbeteren.

Deze genoemde technieken versterken elkaar ook dus het is belangrijk om deze technieken te combineren om zo de beste resultaten te verkrijgen.

In ZyLAB ONE worden de gevonden gegevens om te anonimiseren gemarkeerd zodat deze duidelijk in de documenten te zien zijn en deze kunnen vervolgens direct worden omgezet naar een redactie.

Belangrijke kenmerken van ZyLAB ONE voor het herkennen van (herleidbare) persoonsgegevens zijn:

- Uitgebreide standaard bibliotheek voor het herkennen van persoonsgegevens zoals, persoonsnamen, functies, telefoonnummers, BSN, creditcard gegevens, bankrekeningnummers, huwelijkse staat, geboortedatum, geslacht, adressen, postcodes, plaatsnamen, etc.
- Naast persoonsgegevens ook de mogelijkheid voor het herkennen van organisaties, landen, datum, geldbedragen en meer.
- Voldoet aan de Interdepartementale standaard voor de Woo; lakken tot het @-teken, publiceren van het Woo-besluit met witte zones met een zwart kader en opdruk van de gehanteerde uitzonderingsgrond.
- De mogelijkheid om zelf namenlijsten toe te voegen.
- Zelf configureren van extra regels aan de hand van zoekvragen of patronen.
- Kunstmatige intelligentie maakt gebruik van de laatste stand van de technologie (deep learning) en de gebruikte modellen worden regelmatig verbeterd.

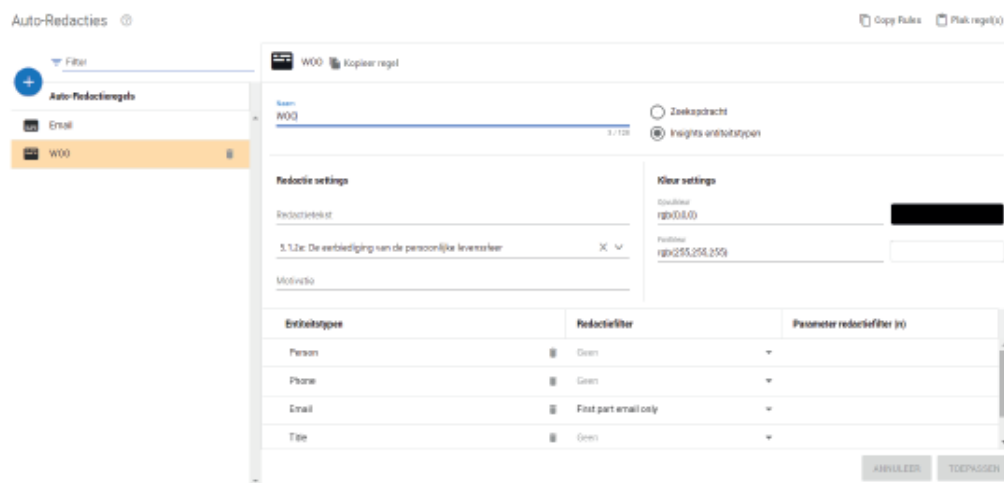
Automatisch anonimiseren

Een volgende stap is het volledig automatisch anonimiseren van specifieke gegevens. Dit betekent dat alle documenten automatisch worden geanonimiseerd op basis van de herkenning van persoonsgegevens door ZyLAB ONE. Het systeem herkent dus automatisch persoonsgegevens en zal deze direct anonimiseren.

Wij noemen dit ook wel voorlakken van de documenten. Uiteraard zullen de documenten nog visueel gecontroleerd moeten worden in sommige gevallen zoals bij Woo-besluiten. In ZyLAB ONE kunnen regels opgesteld worden om bepaalde informatie in alle documenten direct te anonimiseren. Dat kan op 2 manieren, via zoekvragen of door het omzetten van herkende persoonsgegevens naar redacties.

De regels hebben verschillende mogelijkheden:

- Regels maken met één of meerdere zoekvragen
- Regels maken op basis van herkende entiteiten
- Een regel kan een zelfgekozen opmaak (kleur, transparantie) hebben voor de redactie
- Eén of meerdere uitzonderingsgronden kunnen worden toegevoegd aan een regel
- Regels kunnen worden gekopieerd tussen verschillende dossiers of archieven zodat een eenmaal gemaakte regel snel kan worden gebruikt in een ander dossier of archief.



Figuur 10. Auto-redactie regel om automatisch alle persoonsnamen, telefoonnummers, functietitels en e-mail adressen te anonimiseren

Er zijn dus verschillende opties om te anonimiseren en deze toe te passen. Ons advies is:

1. Maak gebruik van de automatische herkenning van persoonsgegevens of andere informatie.
2. Maak automatische redactieregels aan om zo alle documenten voor te lakken
3. Maak zoveel mogelijk gebruik van informatie die al beschikbaar is om de kwaliteit van het voorlakken te verbeteren, zoals b.v. een lijst met namen van medewerkers of betrokken

Anonimiseren of weigeren van documenten of pagina's

Het is natuurlijk niet echt anonimiseren, maar het weigeren van een document of een deel van een document wordt ook wel gezien als onderdeel van het lakken of anonimiseren. ZyLAB ONE heeft een speciale functie voor het weigeren of niet verstrekken van documenten, pagina reeksen of pagina's. Met deze functie kan in 1x een document geweigerd worden, maar ook een enkele pagina of meerdere pagina's. Dit doen we door deze pagina's te anonimiseren of te lakken. Echter op het moment dat deze documenten gepubliceerd gaan worden, zal ZyLAB ONE deze op een speciale manier publiceren.

Om een paar voorbeelden te geven:

- Een document van 100 pagina's wordt volledig geweigerd. Tijdens het publiceren zal dit document vervangen worden door een document van 1 pagina met daarop de weigeringsgrond.
- In een document van 100 pagina's wordt pagina 15 tot en met 25 geweigerd, na het publiceren zijn pagina 15 tot en met 25 vervangen door een enkele pagina met daarop de informatie dat deze pagina's zijn geweigerd met daarop de reden(en).

Op deze manier wordt het knippen in documenten en het definitief verwijderen van informatie voorkomen en ook het tonen van allemaal zwart gelakte pagina's.



Figuur 11. Weigeren van pagina's voor een Woo-besluit waarbij geweigerde pagina's vervangen worden door een enkele pagina met de uitzonderingsgronden

Publiceren

Als alle documenten zijn verwerkt, de juiste tekstdelen zijn geanonimiseerd, kan de informatie worden gepubliceerd. In ZyLAB ONE noemen wij dat het produceren van de documenten.

Bij het maken van de productie worden de handelingen die zijn verricht in het document, zoals de geanonimiseerde teksten en de wettelijke gronden daarvoor, definitief gemaakt. Deze acties worden in de documenten 'gebrand', zodat deze openbaar kunnen worden gemaakt, zonder dat er onbedoeld informatie uitlekt.

Tijdens de productie kan ook een inventarisatielijst worden geproduceerd. Het is ook mogelijk om dit apart te doen van de productie, om tussentijds een overzicht te kunnen maken. In sommige situaties zoals bij een Woo-besluit is er vaak nog een laatste accordering nodig of iemand die de documenten nog wil inzien voordat ze definitief worden gepubliceerd of openbaar worden gemaakt. Met de speciale kwaliteitscontrole optie kan een persoon heel snel en eenvoudig door alle geproduceerde pdf-documenten bladeren. Dit kan op afstand en ieder document kan individueel goedgekeurd worden waarvan ook een registratie plaatsvindt. Zo is de laatste controle snel en eenvoudig uit te voeren.

Is alles akkoord bevonden, dan kan de productie worden gedownload uit ZyLAB ONE en kunnen de documenten worden vrijgegeven.

Meer weten over ZyLAB ONE

Meer weten over ZyLAB ONE en hoe ZyLAB kan helpen bij het verwerken en anonimiseren van documenten, neem dan contact met ons op via onze website, www.zylab.nl of stuur een email naar info@zylab.com.