

Youforce

Personeelsdossier

Koppelvlak import documenten voor
scanleveranciers

25 mei 2018

Inhoudsopgave

Documenthistorie	3
Inleiding	4
Aandachtspunten en randvoorwaarden	5
1. Documenteigenschappen	5
2. Door Raet / klant aan te leveren informatie t.b.v. scanproces.	5
3. Randvoorwaarden	7
Interfacebeschrijving output	14
4. Beschrijving van entiteiten	14
5. Bestand voor verrijking van werknemer / dienstverbanden	16
6. Bestand met documentsoorten	17
7. Kenmerken van het Outputbestand	18
8. Bestandsoverdracht en importeren naar personeelsdossier	19
Bestandsspecificatie Index.xml	20
9. Batchkenmerk	20
10. Identificatie persoon	21
11. Identificatie dienstverband	21
12. Persoonsgegevens	21
13. Documentgegevens	21
Bijlage: WC3 Mimetypes	23

Documenthistorie

Documenthistorie			
Versie	Datum	Status	Wijzigingen
0.0	08.06-2010	Concept	
0.1	14.06.2010	Concept	Wijziging in bestandslay-out medewerkersbestand
0.2	16.06.2010	Concept	Tekstuele aanpassing naar aanleiding van review binnen Interfacing.
0.3	07.07.2010	Concept	Toegestane omvang ZIP-bestand verduidelijkt.
0.4	28.07.2010	Concept	Bestand documentsoorten toegevoegd
1.0	29.07.2010	Definitief	Opmerkingen verwerkt uit review van versie 0.4
1.1	03.06.2011	Definitief	XML-tags en bestandsnaam import aangepast
1.2	05.04.2011	Definitief	Nieuwe elementen toegevoegd
1.3	10.10.2012	Definitief	Randvoorwaarden aangevuld
1.4	21.06.2013	Definitief	Documenttype .EML en .MSG toegevoegd
1.5	30.09.2013	Definitief	Afdeling verplaatst naar <DVB>
1.6	23.09.2016	Definitief	Toegestane inhoud 10c Doctitel aangepast
1.7	25.05.2018	Definitief	Optionele BSN verwijderd uit specificatie

Inleiding

Het personeelsdossier is te beschouwen als een grote verzameling van documenten die betrekking hebben op medewerkers van een klant en (indien relevant) hun dienstverband(en). In dit document wordt beschreven aan welke eisen het bestand moet voldoen dat de leverancier van gescande documenten oplevert aan de klant. Deze interface maakt het mogelijk dat het personeelsdossier automatisch gevuld kan worden met documenten die als resultaat van backlogscanning van het papieren archief aan Raet aangeboden worden.

Na het scannen van de papieren documenten levert de leverancier bestanden aan met de te archiveren documenten, aangevuld met een beschrijvende lijst met kenmerken van deze bestanden. Deze methode is geschikt om grote aantallen documenten op een eenvoudige en betrouwbare manier te verwerken, maar informatie over de status van het inlezen kan uitsluitend achteraf in een logboek teruggevonden worden.

Ten behoeve van het geautomatiseerd verwerken is een interface gedefinieerd, waarin is vastgelegd welke gegevens nodig zijn en in welk formaat de aanlevering plaats kan vinden.

De standaardset van entiteiten staat beschreven in paragraaf [Beschrijving van entiteiten](#).

De specifieke technische specificatie van het indexbestand staat beschreven in paragraaf [Bestandsspecificatie Index.xml](#)

Aandachtspunten en randvoorwaarden

1. Documenteigenschappen

De basis van Raet Personeelsdossier Online wordt bepaald door de opslag van documenten. Een document binnen het personeelsdossier bevat één of meerdere pagina's die gezamenlijk worden opgeslagen in één bestand. Aan deze documenten kunnen diverse kenmerken worden gekoppeld; waarvan de belangrijkste zijn:

- Medewerker: een medewerker of een dienstverband van de medewerker
- Documentsoort: een groepering van een of meerdere type documenten
- Documentomschrijving: een nadere specificatie van documentsoort
- Ingangsdatum: de datum waarop de inhoud van het document van toepassing is

Vooraf de documentsoort en de –omschrijving zijn binnen het dossier uitermate belangrijk voor de herkenbaarheid van het document in het dossier.

Documenten worden binnen het Personeelsdossier gekoppeld aan een medewerker of aan een dienstverband van de medewerker. Zo is een kopie ID-bewijs niet specifiek gerelateerd aan een dienstverband, maar een beoordelingsverslag wel. Dit is een keuze die de klant zelf kan bepalen, en vaak is er een relatie tussen de documentsoort en het wel of niet koppelen aan een dienstverband.

De autorisatie van Personeelsdossier biedt de mogelijkheid om een gebruiker te autoriseren op dienstverbanden, waardoor bepaalde documenten niet zichtbaar zullen zijn. Als voorbeeld heeft een medewerker twee dienstverbanden met de volgende drie documenten:

1. Kopie ID-bewijs gekoppeld aan de medewerker
2. Arbeidsovereenkomst gekoppeld aan dienstverband 1
3. Beoordelingsverslag gekoppeld aan dienstverband 2

Een gebruiker die is geautoriseerd op dienstverband 1 zal alleen het kopie ID-bewijs en de arbeidsovereenkomst kunnen raadplegen, en een gebruiker die is geautoriseerd op dienstverband 2 zal alleen het kopie ID-bewijs en het beoordelingsverslag kunnen raadplegen. Een gebruiker die voor beide dienstverbanden is geautoriseerd zal alle drie de documenten kunnen raadplegen.

Het is van belang om vooraf de documentsoort van een papieren document te bepalen, en of de betreffende documenten aan de medewerker of aan een dienstverband van de medewerker moet worden gekoppeld.

2. Door Raet / klant aan te leveren informatie t.b.v. scanproces.

Voordat de scanpartner over kan gaan tot het scannen van documenten dient deze per batch te beschikken over de volgende informatie:

Gegevens met betrekking tot de batch.

Deze gegevens worden tijdens de implementatie door de HR-consultant van Raet of de klant ter beschikking gesteld aan de scanleverancier en bevatten op klantniveau de identificerende gegevens van de batch. Deze gegevens staan beschreven in de interfacebeschrijving (velden 1-4) in paragraaf [Batchkenmerk](#)

Werknemerbestand voor de verrijking van data.

Uitgangspunt bij het scannen van het papieren dossier is dat op het werknemerdossier alleen de voor de HRM-afdeling logische werknemeridentificatie is vermeld. Deze zijn bij de upload van de batch nodig om de dossiers aan de juiste werknemers te koppelen. Echter voor het personeelsdossier dienen ook logische medewerkergegevens geïmporteerd te worden (bijv. geboortedatum, naam). Deze gegevens worden bij het document opgeslagen als de gegevens die geldig waren op het moment van aanmaak van het document. Daar kan in PDOL op gezocht worden. Bovendien leveren deze gegevens waardevolle informatie over medewerkers waarbij tijdens het inlezen een fout optreedt.

De benodigde medewerkergegevens kunnen door de klant uit het HR-systeem worden betrokken, maar worden desgewenst ook via Youforce beschikbaar gesteld als een afzonderlijk downloadbaar bestand dat aan de scanleverancier ter beschikking gesteld kan worden. In deze [paragraaf](#) vindt u een beschrijving van dit bestand.

Bestand met documentsoorten.

Aan elk documentbestand met gescande pagina's dient een documentsoort gekoppeld te worden. Om het koppelen van gescande documenten aan documentsoorten mogelijk te maken is het noodzakelijk dat de scanleverancier de beschikking krijgt over een bestand met de voor deze klant geldende documentsoorten. Omdat klanten de mogelijkheid hebben om eigen documentsoorten te bepalen, dient dit overzicht per klant aangemaakt te worden. In deze [paragraaf](#) vindt u een beschrijving van dit bestand

3. Randvoorwaarden

Om de inhoud van een document te kunnen raadplegen, bevat Personeelsdossier een documentviewer. Door in de lijst op het tabblad 'Resultaten' te dubbelklikken op een document, ziet de gebruiker de eerste pagina van dat document op het tabblad 'Weergave'. Daarnaast kan men op dit tabblad zoeken naar een (gedeelte van een) woord of zin.

Omdat deze functionaliteit bepalend is voor de gebruikersvriendelijkheid van Personeelsdossier, moeten de documenten aan bepaalde voorwaarden voldoen. Omdat Personeelsdossier in een webbrowser draait, zijn er weinig mogelijkheden om documenten te tonen. De meeste browsers zijn bijvoorbeeld niet in staat zonder plugins PDF-, Word, of Excel-documenten af te beelden. Als dat wel het geval is, gebeurt het binnen elke browser weer anders; zo kunnen menu's bijvoorbeeld worden getoond als balk of als overlay. Om deze reden wordt een document in het tabblad 'Weergave' pas afgebeeld nadat de betreffende pagina is geconverteerd naar het PNG-formaat, dat door alle moderne browsers kan worden weergegeven.

Ook de omvang van een bestand en de wijze waarop het wordt gecomprimeerd, zijn zeer bepalend voor de bruikbaarheid van Personeelsdossier. Dit bepaalt ook in grote mate de bruikbaarheid van het document buiten Personeelsdossier, bijvoorbeeld als het document door een medewerker wordt opgevraagd.

Maximum bestandsgrootte documenten

De maximum bestandsgrootte van een document is 50 megabyte. Dit maximum is van toepassing bij het inlezen van een document. Documenten groter dan 50 MB worden niet geaccepteerd.

Raet adviseert echter de bestandsomvang te beperken tot slechts enkele megabytes, met een maximum van 10 tot 15 MB. Het merendeel van de documenten bestaat uit een klein aantal pagina's (soms enkele tientallen). Gebruikt men een gemiddelde scanresolutie met een eenvoudige compressiemethode, dan zijn dergelijke bestanden tussen de 1 en 10 MB groot. Bestanden van dit formaat zijn snel te verwerken, wat het mogelijk maakt eenvoudig te bladeren door de documenten en de pagina's van de documenten.

In een uitzonderlijk geval is het mogelijk om documenten in te lezen tot een maximale omvang van 50 MB, hetgeen bij tekstdocumenten gelijk staat aan zo'n 100 tot 200 pagina's.

Documenten versus bestanden

Personeelsdossier is bestemd voor het archiveren en raadplegen van documenten in het personeelsdossier van de medewerker. De Algemene Verordening Gegevensbescherming stelt specifieke eisen aan het vastleggen en onderhouden van een personeelsdossier, waarbij de medewerkers grote zeggenschap hebben over de inhoud van het eigen dossier. Zo is het verplicht om irrelevante documenten (al dan niet op verzoek van de medewerker) uit het personeelsdossier te verwijderen en wordt voor alle documenten een minimum en/of maximum bewaartermijn vastgesteld.

Ook de toegang tot het personeelsdossier en tot bepaalde documenten is aan regels gebonden. Zo mag de leasecoördinator uit hoofde van deze functie alleen de leasecontracten raadplegen; andere documenten blijven ontoegankelijk. Veel klanten kiezen er ook voor om de medewerker toegang te geven tot het eigen personeelsdossier.

Personeelsdossier kan alleen op deze manier worden gebruikt als alle documenten afzonderlijk beschikbaar zijn. Wordt ervoor gekozen om alle documenten in één PDF-bestand op te slaan, dan kunnen de documenten niet meer afzonderlijk worden behandeld. Dit heeft de volgende belangrijke nadelen:

- **Bestanden worden te groot**
Als alle documenten in één bestand zijn opgeslagen, wordt dit dusdanig groot dat het openen of doorzoeken van een document wordt vertraagd, hetgeen ten koste gaat van de gebruikersvriendelijkheid. Ook wordt hierdoor bij medewerkers die langer in dienst zijn al snel de maximum bestandsgrootte bereikt.
- **Documenten zijn niet meer terug te vinden op basis van selectiecriteria**
Binnen Personeelsdossier kan men selecteren op kenmerken van een document. Als alle documenten in één bestand zijn opgeslagen, is dat niet meer mogelijk en moet een document worden gezocht met de zoekfunctie binnen documenten. Dit laatste is overigens alleen mogelijk als er tekstherkenning is toegepast; is dat niet het geval, dan is bladeren door de pagina's de enige optie.
- **Documenten zijn niet meer volgens de wettelijke normen (afzonderlijk) te verwijderen**
Documenten kunnen alleen worden verwijderd door het bestand met een PDF-editor te bewerken en opnieuw in te lezen in Personeelsdossier. Ook kan het mechanisme van bewaar- en geldigheidstermijnen binnen Personeelsdossier niet worden gebruikt.
- **Documenten zijn niet te distribueren (bijvoorbeeld via e-mail)**
Het is niet mogelijk een document voor een bepaald doel afzonderlijk uit Personeelsdossier te halen en te distribueren.

Raet raadt ten eerste af om alle documenten in één of enkele PDF-bestanden op te slaan.

Door elk document in een afzonderlijke PDF op te slaan kunnen documenten ook afzonderlijk worden behandeld, in en buiten Personeelsdossier.

Pagina's versus bestanden

Ook het fysiek bij elkaar houden van de verschillende pagina's van een document is van belang. Doordat men meerdere pagina's van een document in één bestand kan opslaan, is gewaarborgd dat deze informatie altijd ook logisch bij elkaar wordt gehouden en als zodanig wordt gepresenteerd.

Het wordt sterk afgeraden bij het scannen van documenten elke pagina als afzonderlijk bestand op te slaan. In dat geval bestaat er namelijk geen relatie tussen de verschillende pagina's (het documentniveau) in het personeelsdossier. Dit maakt het onmogelijk om pagina's behorend bij één bepaald document logisch te groeperen en te sorteren.

Indien desondanks toch wordt gekozen voor de opslag van individuele pagina's, is het aan te bevelen de bestandsnaam of het volgnummer van de pagina in het veld documentomschrijving op te nemen. De scanleverancier kan de bestandsnaamgeving dan het beste zodanig kiezen dat hieruit de volgorde van het scanproces te achterhalen is (via een volgnummer), wat

het mogelijk maakt binnen het personeelsdossier op een basale wijze te groeperen en sorteren.

OCR

Tekstherkenning ('Optical Character Recognition' of OCR) maakt het mogelijk de tekst van gescande documenten te gebruiken om het document te doorzoeken. Personeelsdossier biedt de mogelijkheid een selectie van documenten te doorzoeken door het eerste document te openen en vervolgens de zoekfunctie te gebruiken. Alle documenten in de selectie worden vervolgens doorzocht naar de opgegeven tekst. Deze functionaliteit is alleen mogelijk in bron-documenten (zoals een Word-, of Excel-document), PDF-documenten gegenereerd vanuit een brondocument, of PDF-documenten met gescande pagina's waarvan de tekst door middel van OCR aan het document is toegevoegd.

OCR is niet noodzakelijk als de documenten een zinnige omschrijving hebben en eenduidig gekoppeld zijn aan een documentsoort. In de meeste gevallen kan een document dan zeer snel worden gevonden aan de hand van de eigenschappen van het document. Naarmate deze indexering minder consequent is toegepast, wordt de doorzoekbaarheid van documenten steeds belangrijker om een document terug te kunnen vinden.

OCR verhoogt de gebruikersvriendelijkheid van Personeelsdossier, omdat er eenvoudig kan worden gezocht naar documenten waar een bepaald woord in voorkomt. Daarnaast is het mogelijk tekst in het document te selecteren en te kopiëren naar een andere toepassing.

Raet adviseert om bij gescande documenten tekstherkenning toe te passen om de gebruikersvriendelijkheid van Personeelsdossier te verhogen. OCR is echter geen noodzakelijke optie voor de werking van Personeelsdossier.

Resolutie

De resolutie van de gescande bestanden bepaalt voor een belangrijk deel de kwaliteit. Een te lage resolutie leidt tot slecht leesbare documenten met weinig detail, terwijl een te hoge resolutie resulteert in onnodig grote bestanden. Uitgangspunt van een gescand document is in het algemeen een pagina op A4-formaat. Als deze wordt gescand en opnieuw afgedrukt, zou er weinig tot geen kwaliteitsverschil moeten zijn tussen de 'kopie' en het origineel.

In de praktijk is een scanresolutie van 300 DPI ruim voldoende om dit te bereiken. Ook 200 DPI geeft al een zeer acceptabele kwaliteit: alleen bij inzoomen op het document is het kwaliteitsverschil met 300 DPI waarneembaar. Soms wordt er ten behoeve van het verbeteren van het herkenningpercentage op een hogere resolutie gescand, en wordt na de tekstherkenning de gescande afbeelding teruggebracht naar 200 of 300 DPI.

Raet adviseert het gebruik van 200 DPI, tenzij er een duidelijke noodzaak is om op 300 DPI te scannen. Een hogere resolutie dan 300 DPI wordt sterk afgeraden, omdat het beperkte kwaliteitsvoordeel niet in verhouding staat tot de toename van de bestandsgrootte.

Kleur

Met betrekking tot het scannen van papieren documenten zijn er drie mogelijkheden: zwart/wit, kleur, en grijs tinten.

Het gebruik van zwart/wit is alleen zinvol als er sprake is van tekstdocumenten. Het voordeel is de zeer beperkte bestandsomvang, waardoor documenten razendsnel kunnen worden getoond. Zwart/wit beperkt echter de bewijskracht van een document, omdat er nauwelijks meer details aanwezig zijn. De rechtsgeldigheid van een document is echter pas relevant als het document als bewijs wordt gebruikt in een juridisch dispuut.

Een gescand document moet daarbij worden beschouwd als een kopie van het origineel; de rechtsgeldigheid (of zuiverder geformuleerd: de toelaatbaarheid als bewijs) wordt in dat geval bepaald door de rechter. Hoe beter de kwaliteit, hoe minder discussie er ontstaat over de rechtsgeldigheid. Een kopie van een identiteitsbewijs dat in zwart/wit is gescand wordt bijvoorbeeld niet geaccepteerd als bewijs door de Inspectie SZW, omdat de persoon op de foto niet te herkennen is. In geval van een rechtszaak zal de rechter dit dus niet als bewijs accepteren.

In gevallen waar relevante grafische afbeeldingen (handtekeningen) of foto's worden gescand, is het daarom aan te raden om in kleur of in grijstinten te scannen. Het verschil in bestandsgrootte tussen kleur en grijstinten is echter minimaal, omdat kleuren en grijstinten meestal op dezelfde manier worden opgeslagen. Alleen bij gebruik van verschillende kleurdieptes ontstaat er een verschil in bestandsgrootte. Zo kan men bijvoorbeeld kiezen voor 256 grijstinten (8-bit) tegenover ruim 65 duizend kleuren (16-bit) of ruim 16 miljoen kleuren (24-bit). Er zijn echter weinig bestandsformaten die 256 grijstinten efficiënter opslaan dan de moderne technieken om 65 duizend kleuren op te slaan. Een 16-bit JPEG-bestand is daardoor kleiner dan een 8-bit GIF-bestand; JPEG kent geen 8-bit variant. In de praktijk is er nauwelijks een verschil in bestandsgrootte tussen documenten in grijstinten of in kleur, tenzij er specifieke compressiemethoden worden gebruikt (zie 0 Compressiemethoden).

Raet adviseert om te scannen in kleur of eventueel in grijstinten, zodat voldoende detail wordt geboden om het document als volwaardige kopie te kunnen gebruiken. Scannen in zwart/wit is alléén acceptabel als het tekstdocumenten betreft zonder grafische afbeeldingen (zoals handtekeningen).

Compressiemethoden

Bij het scannen van documenten worden er afbeeldingen gemaakt van de afzonderlijke pagina's. Zonder compressie worden deze pagina's zeer groot: enkele tientallen megabytes per pagina. Het is daarom belangrijk een compressiemethode te gebruiken die de bestandsgrootte beperkt. Daarnaast dienen de documenten leesbaar te zijn voor gebruikers met verschillende toepassingen op verschillende platformen.

Er zijn verschillende methoden om een document te verkleinen zonder significant kwaliteitsverlies. Soms leidt dit er echter toe dat dergelijke documenten niet meer door elke gebruiker te openen zijn, of dat het openen meer tijd in beslag neemt.

Meestal worden gescande documenten als PDF-bestand opgeslagen. Om te kunnen garanderen dat elke PDF-toepassing de documenten kan openen, wordt uit de PDF-standaard afgeleid welke compressiemethoden zijn toegestaan. Er zijn echter verschillende versies van de PDF-standaard en niet elke toepassing ondersteunt de laatste versie. Ook als dat wel het geval is, komt het vaak voor dat bepaalde documenten niet of onjuist worden weergegeven. Compressietechnieken die regelmatig tot problemen leiden zijn JPEG2000 (JP2- of JPX-

bestanden) en 'Multi Raster Content' (MRC).

JPEG-2000 is een modern compressieformaat dat formeel door de PDF-standaard wordt ondersteund, maar in de praktijk niet of slecht wordt ondersteund door PDF-toepassingen. Hoewel deze compressietechnieken tot 30% kleinere bestanden kunnen opleveren ten opzichte van een standaard JPEG-compressie (bij gelijkblijvende kwaliteit), hebben zij in de praktijk nauwelijks toegevoegde waarde voor tekstdocumenten. Bovendien is JPEG-2000 een dusdanig complexe techniek, dat het openen van een bestand vaak langer duurt.

MRC is een techniek waarbij een gescande afbeelding wordt opgesplitst in meerdere lagen; elke laag wordt weggeschreven in een voor dat doel optimale resolutie. Samen gevoegd resulteren deze lagen in een kleiner bestand dan de originele afbeelding. MRC maakt optimaal gebruik van het feit dat de PDF-standaard verschillende lagen ondersteunt. Ook deze techniek wordt echter niet door elke PDF-toepassing correct weergegeven. Daar komt bij dat het samenvoegen van de lagen meestal iets langer duurt, wat merkbaar is als het document wordt opgebouwd.

Voor Personeelsdossier is van belang dat het gebruik van complexe technieken soms resulteert in een langere wachttijd voordat een document wordt weergegeven. Doordat de meeste browsers niet in staat zijn om PDF-documenten af te beelden zonder plugins, moeten de pagina's van het document worden geconverteerd naar een PNG-bestand. Door deze complexe technieken duurt zowel het opvragen van de PDF als het converteren naar PNG langer (soms tot enkele seconden). In combinatie met een groot bestand van tientallen pagina's kan dat oplopen tot tien seconden of meer, hetgeen ten koste gaat van de gebruikersvriendelijkheid.

Daarnaast is er twijfel over de acceptatie van deze technieken voor archiveringsdoeleinden. Dit is bij (semi-)overheidsinstanties een belangrijk criterium. Ondanks de ondersteuning van deze technieken door de PDF-standaard wordt er vaak gekozen voor de oudere PDF-versie 1.4 (bijvoorbeeld PDF/A-1b), waarin alleen eenvoudige JPG-compressie wordt toegepast.

Raet adviseert om gebruik te maken van PDF met JPEG-compressie. Ondanks de iets grotere bestandsomvang waarborgt deze keuze dat de documenten universeel kunnen worden geraadpleegd, met alle PDF-toepassingen en op alle platformen. Daarnaast worden bestanden met deze compressie in Personeelsdossier veel sneller weergegeven, met minder kans op weergavefouten. MRC, JBIG2, en JPEG-2000 worden weliswaar ondersteund, maar zijn significant langzamer en kunnen in sommige gevallen tot weergavefouten leiden. Het TIFF-formaat wordt afgeraden vanwege de grote bestandsomvang en omdat veel gebruikers geen toepassing hebben om TIFF-bestanden te raadplegen..

Documentsoort en -omschrijving

De combinatie van documentsoort en documentomschrijving zou de gebruiker voldoende informatie moeten bieden om te kunnen bepalen om welk document het gaat. Als een document is gekoppeld aan de documentsoort 'ID-bewijs' en de omschrijving is 'Paspoort', hoeft men het document niet te raadplegen om te kunnen vaststellen dat het een kopie is van het paspoort van de betreffende medewerker. Is de documentsoort 'Beoordelingsverslagen' en zijn er drie documenten met respectievelijk de omschrijvingen 'Beoordeling 2010', 'Beoordeling 2011', en 'Beoordeling 2012', dan is eveneens direct duidelijk welk document benodigd is.

Vaak is het erg lastig om bij het scannen van alle papieren dossiers een zinvolle omschrijving te geven. De oplossing is dan handmatige verwerking, dan wel het ontwerpen van een geautomatiseerd systeem waarmee de omschrijving uit de gescande inhoud van het document kan worden afgeleid. Dit laatste kan worden gerealiseerd met tekstherkenning op steekwoorden, of door een tekst op het papieren document te markeren die als omschrijving kan worden gebruikt. Beide oplossingen vergen echter meestal een extra investering, waartoe veel klanten niet bereid zijn.

Raet adviseert om aandacht te besteden aan het zo correct mogelijk definiëren van de documentsoort en de documentomschrijving. De documentsoort is meestal eenvoudig vast te stellen als het papieren dossier ook goed is geïndexeerd en opgeschoond. Een zinvolle omschrijving bepalen is vaak moeilijker, omdat de inhoud van het document daarvoor moet worden gebruikt.

Is het niet mogelijk een zinvolle omschrijving te definiëren, dan kan het beste een generieke omschrijving worden gebruikt die een indicatie geeft van de inhoud. Eventueel kan als documentomschrijving de omschrijving van de documentsoort worden gebruikt (bijvoorbeeld 'Kopie ID-bewijs' in plaats van de technische documentsoortcode 'idBewijs').

Ingangsdatum

Evenals het geval is bij de documentomschrijving, is het ook vaak lastig de ingangsdatum tijdens het scanproces automatisch te bepalen. Op veel documenten staat die datum op de eerste pagina, waarmee het markeren hiervan eenvoudiger is dan het bepalen van een omschrijving. Is dat niet het geval, dan wordt er vaak gebruik gemaakt van een fictieve datum.

Het ontbreken van een ingangsdatum heeft echter tot gevolg dat de bewaartermijnen niet correct worden bijgehouden. Het is dan niet mogelijk de verwijderdatum van de gescande documenten automatisch te bepalen.

Als de feitelijke ingangsdatum niet kan worden bepaald, adviseert Raet om de scandatum te gebruiken en niet een datum die ver in het verleden ligt. Hanteert men een datum in het verleden, dan worden bij uit-dienst alle documenten onmiddellijk afgebeeld op het overzicht te verwijderen documenten. Het kan echter nodig zijn bewust een datum in het verleden te gebruiken om een HR-medewerker op dat moment alsnog de juiste ingangsdatum te laten bepalen.

De vroegst mogelijke datum is 1-1-1900.

Omvang batchbestanden

Gescande documenten worden aangeleverd door middel van batchbestanden in de vorm van een ZIP-bestand, dat een aantal documenten bevat en een indexbestand met de kenmerken van elk document. Wordt er een fout in de indexfile geconstateerd, dan wordt het hele batchbestand afgekeurd; dit voorkomt dat het bestand slechts gedeeltelijk wordt geïmporteerd en niet meer is na te gaan welke documenten wel en niet correct zijn ingelezen.

De batchbestanden moeten aan Youforce worden aangeboden via 'Zenden en ontvangen'. Hoe groter de batchbestanden, hoe langer het uploaden in Youforce duurt. Is er gedurende

het uploaden langere tijd (1 minuut) geen gegevensoverdracht mogelijk, dan wordt het uploaden afgebroken en moet het batchbestand nogmaals worden aangeboden. Gebruikt men een groot batchbestand, dan is de kans hierop groter dan bij een aantal kleinere batchbestanden. Ook als batchbestanden tijdens het importeren worden afgekeurd, is het efficiënter om met meerdere kleine batchbestanden te werken.

Batchbestanden moeten ook groot genoeg zijn om snel veel documenten te kunnen inlezen.

Raet adviseert om batchbestanden samen te stellen met een grootte tot maximaal 500 MB. Als alternatief is het mogelijk de papieren dossiers van één of enkele medewerkers in één batchbestand aan te leveren, of telkens 500 documenten tegelijk in één batchbestand op te nemen. De maximale bestandsgrootte die men via 'Zenden en ontvangen' kan aanbieden is 2 Gigabyte, maar een dergelijke grootte wordt sterk afgeraden.

Interfacebeschrijving output

4. Beschrijving van entiteiten

In deze paragraaf wordt de algemene interface voor de aanlevering van bestanden met gescande documenten beschreven aan de hand van een lijst van sleutelgegevens en aanvullende gegevens. Deze gegevens worden in een indexbestand meegeleverd bij de te importeren documenten.

1. Bronsysteem (Verplicht)

Hier wordt een code verwacht die het bronsysteem karakteriseert. Deze correspondeert met de zogenaamde "appid" die geregistreerd is in de Youforce-applicatietabel voor de betreffende applicatie. Voor aanlevering van data door de scanpartner wordt de volgende vaste waarde gehanteerd:

IMPORTSCAN Batchimport van Scanleverancier

2. Propositie (Verplicht)

Deze rubriek geeft aan uit welk bronsysteem het document afkomstig is. Hier wordt een code verwacht voor de combinatie HR-systeem / Payrollstelsel waarvan de klant gebruikmaakt. Deze stemt overeen met de corresponderende code in de Youforce Portal. Momenteel worden de volgende proposities ondersteund:

BO4GEM HR Core Beaufort – Payroll Gemal
 HRERPD HR Core Business – Payroll Business
 HRSDEM HR Core Education – Payroll Gemal

3. Klantnummer (Verplicht)

Dit nummer is gelijk aan het (beherend) debiteurnummer uit de abonnementenadministratie en identificeert de klant van Raet. Op dit niveau zijn de gebruikers van Youforce ondergebracht. Als het klantnummer niet gevalideerd is moet het gecontroleerd worden, bijvoorbeeld tegen het klantnummer van de beheerder / gebruiker die het bestand aanlevert (via Zend / Ontvang). Dit klantnummer kunt u achterhalen in Youforce onder 'Mijn profiel'

4. HR-identificatie (Verplicht)

Deze entiteit is een verbijzondering van de HR-klant en dient om verdere gegevens over werknemer en dienstverband uniek bepalend te maken binnen het klantnummer. Het is gelijk aan het opdrachtgevernnummer van de HR-klant (in het facturatiesysteem), voor zover aanwezig, of een andere identificatie van de aanleverende klant.

5. Identificatie Persoon (Verplicht)

Dit is de unieke identificatie van de werknemer in zijn relatie met de werkgever. Het gaat hierbij dus niet noodzakelijk om de identificatie van de natuurlijke persoon: niet elk HR-systeem kent dat begrip. De werknemeridentificatie is te onderscheiden in:

5a – Persoonsnummer

Technische sleutel (uniek in de gegeven context en onveranderlijk)

5b – Persoonscode

Door de klant gehanteerd kenmerk van de werknemer (mag gelijk zijn aan 5a).

6. Identificatie Dienstverband

Dit is de identificatie van het dienstverband c.q. de aanstelling van de werknemer, als deze bijvoorbeeld bij verschillende afdelingen tewerkgesteld is. Afhankelijk van de positie kan een werknemer meer dan één dienstverband hebben. Als geen dienstverband geleverd wordt, wordt het corresponderende document op het niveau van de werknemer geplaatst. Het dienstverband wordt gekarakteriseerd door:

6a – Dvbnr

Technische sleutel (uniek in de gegeven context en onveranderlijk)

6b – Dvbcode

Door de klant gehanteerd kenmerk van het dienstverband (evt. gelijk aan 6a)

6c – Afdeling

De afdeling of organisatorische eenheid waar het dienstverband betrekking op heeft. Het wordt alleen gebruikt voor nadere identificatie van het dienstverband bij de import, bij controles op uniciteit en bij foutmeldingen en logging.

7. Geboortedatum

Dit is de geboortedatum van de betreffende persoon, voor zover bekend. Dit gegeven wordt alleen gebruikt voor nadere identificatie van de werknemer bij de import, bij controles op uniciteit, en bij foutmeldingen en logging.

8. Actuele naam van de werknemer

Dit is de volledige, actuele naam van de werknemer, inclusief voorvoegsels, voorletters en eventueel de naam van de echtgenoot. Deze wordt alleen gebruikt voor nadere identificatie van de werknemer bij de import, bij foutmeldingen en logging.

9. Documentkenmerken (Verplicht)

De basis van het personeelsdossier wordt gevormd door documenten en niet de individuele pagina's van een document. Een document bevat dus een of meerdere pagina's die gezamenlijk worden opgeslagen in één bestand en geïdentificeerd wordt door de documentsoort

9a – Documentsoort (Verplicht)

Dit is de identificatie van de documentsoort. De documentsoort wordt gebruikt om het document nader te rubriceren.

9b – Documenttype (Verplicht)

Geeft het mimetype van het document aan, te karakteriseren d.m.v. de W3C-standaard. Deze moet overeenkomen met de op het MS Windows-platform gekoppelde extensie van de bestandsnaam (11d).

Voorbeelden van mimetypes zijn:

application/pdf ⇔ PDF

application/ms-word ⇔ DOC

image/jpeg ⇔ JPG, JPEG

In de [bijlage](#) is een lijst van ondersteunde mimetypes opgenomen.

9c – Documentomschrijving (Verplicht)

Deze entiteit geeft een omschrijving van het document zoals die getoond moet worden in het personeelsdossier. Deze omschrijving dient ter herkenning van een document in het personeelsdossier door een gebruiker. Het verdient de voorkeur om een naamconventie

op te stellen, waarbij delen van de omschrijving wellicht automatisch gevuld kunnen worden tijdens het scanproces.

9d – Bestandsnaam (Verplicht)

Bevat de naam van het bestand, zoals dit is opgeslagen in het bijbehorende ZIP-bestand, inclusief een met het mimetype van 11a corresponderende extensie.

9e – Ingangsdatum bestand (Verplicht)

Datum waarop het oorspronkelijke document is ontstaan. Voorbeeld: een paspoort uit 1999 zal als ingangsdatum 1999-xx-xx krijgen terwijl de feitelijke aanmaak- of importdatum van het gescande document later is. Dit is met name van belang bij het uploaden van oude documenten. Voor de massale scan zal het ondoenlijk zijn om deze datum voor elk document individueel te bepalen. In dit geval is de scandatum de meest voor de hand liggende datum.

9f – Datum einde geldigheid (Optioneel)

Bevat de datum waarop de inhoud van het document ongeldig is geworden. Dit kan van belang zijn bij document waarvan een geldig exemplaar aanwezig moet zijn in het dossier, zoals een verblijfsvergunning of een certificaat.

9g – Verwijderingsdatum (Optioneel)

Bevat een handmatig vastgestelde datum waarop het document verwijderd moet worden uit het dossier. De verwijderingsdatum wordt automatisch bepaald door het personeelsdossier, en met dit element dan deze berekende datum worden overschreven.

5. Bestand voor verrijking van werknemer / dienstverbanden

Het via Youforce geleverde werknemersbestand heeft tot doel om in het scanproces eventueel ontbrekende persoons- en/of dienstverbandgegevens aan de documenten te kunnen toevoegen. Dit bestand wordt geleverd als een CSV bestand met als separator ";". Op de eerste regel worden de kolomnamen vermeld en in de daarop volgende regels de hierbij behorende inhoud. Dit bestand wordt aangemaakt op basis van de aanwezige medewerkers in het personeelsdossier.

In dit bestand worden de volgende entiteiten opgenomen, met verwijzing naar [de beschrijving](#) in paragraaf 3.1

Klantnummer	correspondeert met veldnummer 3
Actnaam	correspondeert met veldnummer 10
Persoonscd	correspondeert met veldnummer 5b
Persoonsnr	correspondeert met veldnummer 5a
Dvbnr	correspondeert met veldnummer 6a
Dvbcode	correspondeert met veldnummer 6b
Afdeling	correspondeert met veldnummer 6c
Gebdatum	correspondeert met veldnummer 7

Per te scannen document moet bepaald worden of een document gekoppeld moet worden aan een persoon of een dienstverband. Indien u in uw organisatie nauwelijks met meervoudige dienstverbanden te maken hebt, of indien in het papieren dossier geen onderscheid wordt gemaakt naar documenten per dienstverband, zult u naar alle waarschijnlijkheid alle documenten aan een persoon willen koppelen.

Bijvoorbeeld:

- Een identiteitsbewijs (paspoort / rijbewijs) zal als document aan een persoon gekoppeld zijn.
- Indien een medewerker meerdere dienstverbanden bij een werkgever heeft en bijvoorbeeld per dienstverband een afzonderlijke beoordeling heeft, kan er voor gekozen worden deze documenten te koppelen aan de daarbij behorende dienstverband

Naar verwachting zullen tijdens het scanproces alle documenten gekoppeld worden aan de persoon.

- De sleutel voor de persoon staat vermeld in "Persoonscd"
- De sleutel van het dienstverband is de combinatie van "Persoonscd" en "dvbcode"

Indien een document gekoppeld moet worden aan een persoon dan dienen de volgende gegevens uit het bestand overgenomen te worden:

- Persnr
- Perscd
- Gebdatum

De velden 6a, 6b en 6c moeten in dit geval achterwege blijven!

Indien een document gekoppeld wordt aan een dienstverband dan dienen de volgende gegevens uit het bestand overgenomen te worden

- Persnr
- Perscd
- Dvbnr
- Dvbcode
- Afdeling
- Gebdatum

6. Bestand met documentsoorten

Het door Raet beschikbaar gestelde bestand met documentsoorten heeft tot doel om in het scanproces de ontbrekende documentsoort toe te voegen. Dit bestand wordt geleverd als een CSV bestand met als separator ";". Op de eerste regel worden de kolomnamen vermeld en in de daarop volgende regels de hierbij behorende inhoud.

In dit bestand worden de volgende entiteiten opgenomen, met eventuele verwijzing naar [de beschrijving](#) in paragraaf 3.1

Klantnummer correspondeert met veldnummer 3
 DocumentCode correspondeert met veldnummer 9a
 DocumentNaam correspondeert met veldnummer 9c
 Zichtbaar *

*

Het veld 'zichtbaar' geeft aan of een documentsoort in het personeelsdossier zichtbaar is.

1 = zichtbaar

0 = onzichtbaar

Dit veld is voor de volledigheid opgenomen in het bestand, maar is voor de scanleverancier niet van belang.

7. Kenmerken van het Outputbestand

Het outputbestand dat door de scanleverancier opgeleverd dient te voldoen aan de volgende specificaties:

Het aan te leveren bestand dient een ZIP-bestand te zijn met hierin

1. de individuele bestanden met gescande documenten
2. en 1 indexbestand met de naam 'index.xml'.

Bij het aanmaken van het ZIP-bestand dient met het volgende rekening te worden gehouden:

- De individuele documentbestanden mogen niet groter zijn dan 50 MB

Het is sterk te adviseren om één document per bestand aan te leveren, omdat het personeelsdossier document-georiënteerd is. Als er meerdere papieren documenten in één fysiek documentbestand worden opgenomen dan wordt de gebruiksvriendelijkheid van personeelsdossier belemmerd. Geadviseerd wordt om de omvang van afzonderlijke documenten te beperken tot maximaal 10 MB.

- Aan te leveren ZIP-bestand(en) mogen niet groter zijn dan 2 GB

Omwille van de hanteerbaarheid en afhandeling bij fouten is het aan te raden de omvang van de ZIP-bestanden te beperken. Een omvang rond 500 MB geeft een goede balans; groot genoeg om veel te verwerken en klein genoeg voor snelle afhandeling bij fouten.

- Er mogen geen directories in de ZIP-bestanden zijn

Als dit wel het geval is worden de documentbestanden in een subdirectory uitgepakt en kan het importproces de bestanden niet vinden.

- Alleen de compressie methode 'deflate' wordt ondersteund.
- Encryptie wordt niet ondersteund.
- Multiple volumes wordt niet ondersteund.
- Voor bestanden in de ZIP file (>2Gb of meer dan 65.535 documenten) is gebruik van de Zip64 extensies toegestaan.

In het ZIP-bestand dient het index-bestand de verplichte naam 'index.xml' te hebben. De keuze van naamgeving van de documentbestanden zelf is vrij.

Het index bestand is een XML tekstbestand dat voor ieder te bewaren document alle benodigde sleutelgegevens dient te bevatten, zie hoofdstuk 4. Het ZIP-bestand wordt aan de hand van de index verwerkt. Documenten die niet in de index vermeld staan, worden simpelweg overgeslagen.

De bestandsnaam van het ZIP-bestand dient te voldoen aan onderstaand masker.

PDOL_batchimport_scanleverancier_<jjjjmmdd>_<hhmmss>.zip

De eerste drie woorden stellen een letterlijke string voor met separator '_', net als de bestandsextensie '.zip'. De fragmenten tussen haakjes dienen vervangen te worden door actuele strings, waarbij geldt:

<jjjjmmdd> Jaar-, maand- en dagnummer van aanmaak

<hhmmss> Tijdstempel in uren, minuten en seconden van het moment van aanmaak

8. Bestandsoverdracht en importeren naar personeelsdossier

Het outputbestand wordt door de scanleverancier aangeboden aan de klant. De klant kan vervolgens het bestand via de Zend-Ontvang module van Youforce aanbieden aan het personeelsdossier. Bestand moet ingezonden worden via de keuze Zenden / Personeelsdossier.

Bestandsspecificatie Index.xml

In deze paragraaf wordt het outputformaat beschreven van het indexbestand dat door de scanleverancier aangeleverd moet worden. Deze outputbeschrijving dient opgeleverd te worden als een XML-bestand

De definitie van het XML-bestand is vastgelegd in het bij deze documentatie horende XSD-bestand "ImportKoppelvlak.xsd" en wordt hier verder toegelicht.

9. Batchkenmerk

De velden 1 tot en met 4 bevatten de identificerende gegevens van de batch. Deze velden worden de HR consultant in overleg met de klant verstrekt aan de scanleverancier.

	Veldnaam	Formaat	Verplicht	Toegestane inhoud
1	BronSysteem	Alfa (12)	Ja	Vaste inhoud: ImportScan
2	Propositie	Alfa (6)	Ja	Code voor HR+Payroll-systeem; een van de volgende waarden: BO4GEM BO4ASA HRERPD HRSGEM PRIASA
3	Klantnummer	Alfa (7)	Ja	Klantnummer . Zie Youforce / Mijn profiel
4	HR-identificer	Alfa (15)	Ja	Inhoud afhankelijk van propositie; indien propositie: BO4GEM Kenmerk van de Beaufortinstallatie, bv. de licentiesleutel. Als deze niet uniek bepalend is moeten de verzamelingen persoonsnummers uit de diverse installaties disjunct zijn. HRERPD Logische bedrijfscode van het bedrijf, veld nummer/code uit entiteit Bedrijf, gegevenselement 18 HRSGEM HRiS-klantnummer uit tabel ZD00 in directory WCU, veld CDCODE

10. Identificatie persoon

Documenten worden altijd gekoppeld worden aan een werknemer. In dit kader wordt onder een werknemer de natuurlijke persoon verstaan. Deze gegevens moeten altijd aanwezig zijn. Daarom is verplichte vulling van de velden 5, 5a en 5b noodzakelijk.

	Veldnaam	Formaat	Verplicht	Toegestane inhoud
5	Persoon	container	Ja	Omvat 5a en 5b
5a	Persoonsnr	Alfa (25)	Ja	Overnemen uit medewerkerbestand; veld: Persoonsnr
5b	Persoonscd	Alfa (25)	Ja	Overnemen uit medewerkerbestand; veld: Persoonscd

11. Identificatie dienstverband

Documenten kunnen ook gekoppeld aan een dienstverband. Deze optie zal waarschijnlijk niet in het scanproces gebruikt worden, maar behoort wel tot de mogelijkheden die het personeelsdossier biedt. Indien documenten in het scanproces gekoppeld moeten worden aan een dienstverband dienen de velden 6, 6a, 6b en 9 ingevuld te worden. Indien het document alleen aan een persoon gekoppeld moet worden dan mogen de entiteiten van dit blok niet geschreven worden. Als u toch een document wilt koppelen aan het dienstverband zonder een dienstverbandsnummer dan kunt u een streepje ("-") gebruiken.

	Veldnaam	Formaat	Verplicht	Toegestane inhoud
6	Dvb	Container	Nee	Omvat 6a en 6b
6a	Dvb-nr	Alfa (25)	Ja	Overnemen uit medewerkerbestand; veld: dvbnr
6b	Dvb-code	Alfa (25)	Ja	Overnemen uit medewerkerbestand; veld: dvbcode
6c	Afdeling	Alfa (50)	Nee	Overnemen uit medewerkerbestand; veld: Afdeling

12. Persoonsgegevens

De velden 7 tot en met 10 bevatten aanvullende gegevens van een werknemer c.q. dienstverband.

	Veldnaam	Formaat	Verplicht	Toegestane inhoud
7	Gebdatum	jjjj-mm-dd	Nee	Overnemen uit medewerkerbestand; veld: Gebdatum
8	Actnaam	Alfa (90)	Ja	Overnemen uit medewerkerbestand; veld: Actnaam

13. Documentgegevens

	Veldnaam	Formaat	Verplicht	Toegestane inhoud
9	Document	container	Ja	omvat 11a t/m 11e

9a	Docsoort	Alfa (50)	Ja	Overnemen uit documentsoorten bestand: veld DocumentCode
9b	Doctype	Alfa (75)	Ja	W3C mimetype van het document (bijlage 2)
9c	Doctitel	Alfa (50)	Ja	Beschrijvende naam over de inhoud van het document
9d	Docnaam	Alfa (50)	Ja	Bestandsnaam van het document inclusief extensie
9e	Docdatum *	jjjj-mm-dd	Ja	Aanmaakdatum c.q. datum laatste wijziging Minimale datum: 1900-01-01
9f	DatumEindeGeldigheid	jjjj-mm-dd	Nee	Datum einde geldigheid van de inhoud van het document Minimale datum: 1900-01-01
9g	Verwijderingsdatum	jjjj-mm-dd	Nee	Handmatig vastgestelde datum waarop het document verwijderd moet worden ter vervanging van de automatisch bepaalde datum Minimale datum: 1900-01-01

*

De documentdatum is bedoeld als de datum waarop de inhoud van het document van toepassing is geworden. Vaak is dit gelijk aan de datum waarop het document is ontstaan, of de datum waarop het document is geïmporteerd in het personeelsdossier. Tijdens het scanproces is het vaak technisch of financieel niet haalbaar voor elk document deze datum achteraf vast te stellen. In overleg met de klant moet dan een alternatief doel voor dit veld bepaald worden.

Als de batch handmatig aangeleverd wordt, bijvoorbeeld door een scanservicebedrijf, dan is het klantnummer mogelijk niet gevalideerd. Het klantnummer wordt dan tijdens de import gecontroleerd tegen het in Youforce vastgelegde debiteurnummer, behorende bij de gebruiker of het certificaat waarmee het ZIP-bestand wordt aangeleverd via Zend/Ontvang.

In Beaufort is het in theorie mogelijk dat door meerdere installaties bij dezelfde klant van Raet de sleutelgegevens van werknemers (Klantnummer, Persoonsnummer, Dienstverband) niet uniek zijn. Dergelijke situaties kunnen bijvoorbeeld ontstaan door fusies. In de praktijk blijkt het (nog) niet voor te komen dat verschillende werknemers bij één klant door dit verschijnsel hetzelfde persoonsnummer hebben. Waarschijnlijk zou dit overigens op andere gebieden ook tot problemen leiden. Om de koppeltabel toekomstvast te maken is het veld nummer 4 toegevoegd, zodat zonder problemen ook met meerdere Beaufort-installaties gewerkt kan worden.

Bijlage: WC3 Mimetypes

In onderstaande tabel is een lijst opgenomen van mimetypes (documenttypen) die momenteel door Personeelsdossier Online ondersteund worden.

WC3 Mimetype	Omschrijving
application/pdf	Adobe Acrobat Reader document
application/msword	MS Word bestand (tekst, *.doc)
application/vnd.openxmlformats-officedocument.wordprocessingml.document	MS Word 2007 bestand (tekst, *.docx)
application/vnd.oasis.opendocument.text	Open Document Tekst (tekst, *.odt)
application/vnd.ms-excel	MS Excel bestand (spreadsheet, *.xls)
application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	MS Excel 2007 bestand (spreadsheet, *.xlsx)
application/vnd.oasis.opendocument.spreadsheet	Open Document Spreadsheet (spreadsheet, *.ods)
application/vnd.ms-powerpoint	MS PowerPoint bestand (presentatie, *.ppt)
application/vnd.openxmlformats-officedocument.presentationml.presentation	MS Powerpoint 2007 bestand (presentatie, *.pptx)
application/vnd.oasis.opendocument.presentation	Open Document Presentation (presentatie, *.odp)
image/jpeg	Afbeelding, jpeg formaat (gecomprimeerd)
image/gif	Afbeelding of animatie, gif formaat
image/bmp	Afbeelding, bitmap formaat
image/png	Afbeelding, png formaat
image/tiff	Afbeelding, tiff formaat
image/tif	Afbeelding, tif formaat
application/vnd.ms-outlook	E-mailbestand, msg formaat
message/rfc822	E-mailbestand, eml formaat