

# PDF/A functionaliteiten en toepassingsgebieden

**Het PDF/A Competence Center organiseerde onlangs het eerste internationale PDF/A congres in Amsterdam. Organisaties en experts op het gebied van PDF-technologie kwamen hier samen om informatie en ervaringen met de 'nieuwe' ISO-standaard voor langetermijnarchivering van elektronische documenten te delen.**

DOOR ROBERTO NAGEL

Deze PDF/A standaard is inmiddels een bekend gegeven onder documentspecialisten. Maar nog een groot aantal mensen heeft geen idee hoe PDF/A werkt, wat de standaard precies inhoudt en hoe dit opgenomen kan worden in het bedrijfsproces.

## Tiff versus PDF

Traditioneel erkent men twee standaardformaten voor documentarchivering, namelijk het tiff-formaat en PDF-formaat. Een groot nadeel van het tiff-formaat is dat het document als image wordt opgeslagen, waardoor

het niet in aanmerking komt voor full-text indexerend. Daarom moeten er indexbegrippen worden geïdentificeerd en als apart bestand worden aangeleverd aan het archiefsysteem. Die indexbegrippen, eigenlijk keywords, zorgen ervoor dat het document terug te vinden is. In de praktijk wordt er vaak een maximum van zes indexbegrippen gehanteerd, omdat de indexverzameling anders onhanteerbaar wordt.

Bij een 'normaal' PDF-bestand is het apart aanleveren van indexbegrippen niet nodig, omdat de informatie die in het bestand staat beschikbaar, en dus doorzoekbaar, blijft. PDF biedt full-text search en daarnaast is het mogelijk om informatie uit het document te kopiëren en achteraf te analyseren. Bovendien biedt het PDF-formaat de mogelijkheid om met indexbegrippen te werken. Zo kunnen bijvoorbeeld gescande documenten als images in een PDF-bestand worden opgenomen en de indexbegrippen kunnen als metatags in hetzelfde bestand worden opgenomen. Bovendien kan de inhoud van de image desgewenst via OCR herkend en in leesbare vorm in het PDF-bestand worden opgenomen om alsnog een full-text search mogelijk te maken.

## Voordelen

Deze voordelen van PDF ten opzichte van tiff maken van PDF een efficiënt formaat om documenten op te slaan en te distribueren. Omdat PDF zich tot de de facto standaard voor documentuitwisseling heeft ontwikkeld, trokken verschillende partijen aan de bel bij Adobe. Organisaties die hierbij nauw betrokken waren, zijn het AIIM (Association for Information and Image Management), het NPES (National Printing Equipment Association) en archivarissen van Amerikaanse overheidsinstanties. Zij eisten commitment van de leverancier om een

bepaald bestandsformaat te blijven ondersteunen, ook bij nieuwe versies van de software. Die eis werd nijpend omdat het steeds vaker gebeurde dat nieuwe versies van de Reader oude PDF-bestanden niet ondersteunden. De ISO-werkgroep definieerde zeer snel een standaard voor elektronisch gearchieveerde documenten gebaseerd op PDF versie 1.4, te weten PDF/Archive (PDF/A). De afgelopen jaren heeft deze standaard veel aandacht gekregen en er is voldoende ervaring opgedaan zodat het ondertussen mogelijk is om over best practices te praten voor de omgang met en de toepassingsgebieden van PDF/A.

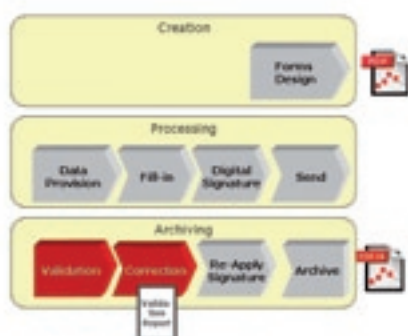
## PDF/A basisfuncties

Het bekijken van een PDF/A-document is potentieel anders dan een 'traditioneel' PDF-document. Door de functionele beperkingen van PDF/A ontstaat de vraag naar een dedicated viewer die veel sneller laadt en bovendien kan aangeven of het geopende document aan de PDF/A-specificaties voldoet.

*"Het bekijken van een PDF/A-document is potentieel anders dan een 'traditioneel' PDF-document."*

Een tweede basisfunctie van PDF/A is het aanmaken van een bestand. Daarbij kan het gaan om een conversie naar PDF/A of om het aanmaken van een geheel nieuw bestand op basis van databasegegevens. Er zijn meerdere leveranciers die software leveren om automatisch hogere volumes bestanden te converteren naar PDF/A in geval

van grote migratieprojecten. Het is dus mogelijk om eenmalig het huidige archief om te zetten naar PDF/A. Tevens is het mogelijk om de creatie van PDF/A-bestanden onderdeel te laten zijn van het outputproces. Wanneer documenten in de printstroom worden meegenomen kan er automatisch een PDF/A-bestand worden aangemaakt en naar het archief worden doorgestuurd. De conversie naar PDF/A is de laatste stap in



## Omgaan met een digitale handtekening

Bij de conversie van een elektronisch document naar PDF/A worden een aantal functies niet in het nieuwe bestand overgenomen en worden er nieuwe elementen aan toegevoegd. Zo worden JavaScript, dynamische data en formulieronderdelen verwijderd en fonts en andere resources in het bestand geplaatst. Hierdoor worden digitale handtekeningen ongeldig en zo kan het dus zijn dat een document niet meer rechtsgeldig is. Een simpele oplossing is om na de conversie naar PDF/A het document opnieuw te laten signeren. Maar dat wordt lastig als een document afkomstig is van een externe partij. Organisaties en belangenverenigingen zoeken momenteel naar een passende oplossing voor deze problematiek. Een van de oplossingen, die nu onder de naam 'TrasiDoc' wordt uitgewerkt, is om het her-signeren over te laten aan een onafhankelijke ('trusted') partij. Deze derde partij ontvangt het originele document met de handtekening, voert een validatie uit en maakt zelf de conversie naar PDF/A. Als trusted party kunnen zij dan zelf het PDF/A-document digitaal ondertekenen. Het concept biedt tevens de mogelijkheid om het beschreven proces binnen het bedrijf in te richten als trusted black-box en het gehele proces te laten certificeren door een onafhankelijke partij.

Roberto Nagel



de verwerking, behalve als men een digitale handtekening wil toevoegen. Daarbij zijn we aangekomen bij een derde belangrijke basisfunctie van PDF/A, namelijk het digitaal signeren van een document, bijvoorbeeld om het vervolgens te versturen of te archiveren.

De laatste basisfunctie is het valideren van een PDF/A-bestand om er zeker van te zijn dat een bestand aan de specificaties van PDF/A voldoet. De beheerder wordt op dit punt geconfronteerd met verschillende keuzes. Hij moet zich bijvoorbeeld afvragen welk(e) conformiteitsniveau(s) de organisatie wil hanteren. Als er sprake is van een digitale handtekening wordt ook gekeken of deze nog geldig is.

### Toepassingsgebieden

Er zijn verschillende scenario's denkbaar waarbij PDF/A kan worden toegepast. Het is bijvoorbeeld mogelijk om inkomende (fysieke) post in de postkamer te scannen met intelligente oplossingen voor scanning en herkenning om er vervolgens een PDF/A-document van te maken. Met de informatie die gevonden is aan de hand van de OCR-scanning is het wellicht ook mogelijk om bedrijfsprocessen verder aan te sturen. Groot voordeel van dit toepassingsgebied is dat de documenten direct klaar zijn voor het archiveringsproces en dat er geen afzonderlijke conversie naar PDF/A ondergaan hoeft te worden. Naast inkomende papieren documenten worden organisaties steeds vaker geconfronteerd met inkomende elektronische documenten, bijvoorbeeld opdrachtbestellingen en facturen. Het is natuurlijk mogelijk om een conversie te doen naar PDF/A, maar als het document een digitale handtekening bevat, wordt deze door de conversie ongeldig. Hier ligt echter een flinke

uitdaging. Meer hierover in het kader 'Omgaan met een digitale handtekening'. Het elektronische formulier, het derde toepassingsgebied van PDF/A, kent hetzelfde probleem. PDF/A kan tot slot worden toegepast op de migratie van het archief, waarbij een document na validatie (en validatierapport) wordt opgeslagen in het archief. Voor de conversie van een compleet archief naar PDF/A is het van belang om goede oplossingen in huis te halen die hoge volumes aankunnen en ook hoogwaardige kwaliteit leveren. Een veelvoorkomend probleem bij de conversie is nog wel de ondersteuning van bepaalde fonts. Een geavanceerde migratieoplossing gaat automatisch op zoek naar oplossingen hiervoor en zal dus vervangende fonts aanbieden.

### Volg de ontwikkelingen

PDF/A is met enthousiasme ontvangen, maar of dat enthousiasme aanhoudt is de vraag. De verwachting is namelijk dat er meerdere versies van PDF/A zullen komen. Een standaard ontwikkelt zich en blijft voortdurend in beweging. Het is zaak om deze ontwikkelingen nauwlettend te volgen en niet zomaar over te stappen naar een nieuw formaat. Laat u goed informeren wat de beste oplossing is voor uw organisatie. □

*De auteur, Roberto Nagel is Technisch Directeur bij Document Dialog, en was spreker op het congres in Amsterdam*

(ADVERTENTIE)

<b>euroscript Delt Nederland</b> Dé Documentum specialisten		
<p><b>euroscript Delt Nederland BV</b> is een 100% gespecialiseerde EMC Documentum Partner met vestigingen in de Benelux.</p> <p><b>euroscript Delt Documentum Service Maintenance &amp; Support</b> Wij bieden een breed scala aan <b>Documentum service,- onderhoud -en ondersteuningsdiensten</b>, zowel onsite als remote uitgevoerd.</p> <p><b>euroscript Delt Documentum Report Generator</b> euroscript Delt levert de <b>Delt Report Generator voor Documentum®</b>. Deze software genereert professionele management rapportages uit Documentum.</p> <p><b>euroscript Delt Documentum migratie</b> euroscript Delt verzorgt als een van de eerste dienstverleners migratietrajecten naar <b>EMC Documentum 6</b>.</p> <p><b>euroscript Delt Documentum Proven Professionals</b> Alle consultants van euroscript Delt zijn gecertificeerde "<b>Documentum Proven Professionals</b>"</p>		
		
Informatie: (040) 250 79 00, info@euroscriptdelt.nl, www.delt.nl		