

**Forum Standaardisatie**

Wilhelmina v Pruisenweg 104
2595 AN Den Haag
Postbus 84011
2508 AA Den Haag
www.forumstandaardisatie.nl

notitie

Aan:	Forum Standaardisatie		
Van:	Bureau Forum Standaardisatie		
Datum:	3 april 2018	Versie	1.0
Betreft:	Overzicht reacties openbare consultatie toepassingsgebieden document- en content-standaarden		
Bijlagen:	1. Reactie Geonovum 2. Reactie Koninklijke Bibliotheek		

1. Reactie van Geonovum op het expertadvies over het toepassingsgebied van OWMS

Op de aanpassing van de beschrijving van de functioneel toepassingsgebied van OWMS heb ik de volgende opmerkingen:

Voor OWMS, is onderstaande voorgesteld. Hierbij geeft de expertgroep de kanttekening dat het begrip 'informatieobject' toelichting behoeft. Deze toelichting ontbreekt, waardoor het toepassingsgebied onduidelijk is.

Daarnaast is er niets gedaan met het advies van de expertgroep om de relatie met andere standaarden voor metadata te duiden, NL profiel op ISO 19115 en 19119 staan al op de pas toe leg uit lijst, dus dat lijkt me wel een voorwaarde om daar in ieder geval helderheid over te geven.

Ik krijg de indruk dat de adviezen van de expertgroep niet zijn doorgevoerd in de aanpassing van de beschrijving van het functioneel toepassingsgebied. Het toepassingsgebied is hierdoor onduidelijk.

OWMS

De expertgroep adviseert als functioneel toepassingsgebied voor OWMS:

OWMS moet worden toegepast op het aanbieden van metadata over publieke informatieobjecten op internet.

De expertgroep maakt de kanttekening dat er diverse andere standaarden voor metadatering bestaan. Voorbeelden hiervan zijn DCAT (datasets), ISO 19115 (geografische datasets en dataset series), RDFa (tekstfragmenten) en de Richtlijn Metagegevens Overheidsinformatie (archiefbescheiden). De expertgroep adviseert om bij de omschrijving van het functioneel toepassingsgebied te wijzen op de relaties die OWMS heeft met deze en andere standaarden. Hierbij zij opgemerkt dat het niet mogelijk zal zijn om een uitputtende opsomming te geven.

De expertgroep maakt de kanttekening dat het begrip 'informatieobject' toelichting behoeft. De expertgroep adviseert daarom om bij de omschrijving van het functioneel toepassingsgebied een verklaring van dit begrip op te nemen.

2. Reactie van de Koninklijke Bibliotheek op het expertadvies over de toepassingsgebieden document- en content-standaarden

Hieronder mijn commentaar op het Expertadvies functioneel toepassingsgebieden document- en (web)content-standaarden, versie 8 december 2017.

1

Zijn er volgens u aanvullingen of wijzigingen nodig in paragraaf 1.3, 1.5 en 1.6 ('Doorlopen proces', 'Samenstelling expertgroep' en 'Toelichting standaard') gezien vanuit het doel om het Forum Standaardisatie en Overheidsbreed Beleidsoverleg Digitale Overheid te voorzien van een inhoudelijk relevante toelichting?

Nee.

6

Bent u het eens met het door de expertgroep geadviseerde functioneel toepassingsgebied voor PDF 1.7, PDF/A-1 en PDF/A-2? [paragraaf 2.5 van het expertadvies]?

Op hoofdlijnen: ja! Ik ben ook blij dat in de tekst het risico op informatieverlies bij conversies naar (en tussen) PDF expliciet wordt genoemd, plus de opmerking dat dit een reden kan zijn om geen conversie naar PDF uit te voeren. Hierbij nog wel een paar opmerkingen:

Wat is een 'document' precies?

Het expertadvies is van toepassing op "niet- of beperkt reviseerbare documenten". Wat ik hierbij mis is wat precies onder een "document" wordt verstaan. Neem bijvoorbeeld onderstaand Wikipedia lemma:

<https://en.wikipedia.org/wiki/Document>

Onder "kinds" worden hier ook spreadsheets genoemd. Ik vraag me alleen af of dit ook de beoogde scope van het advies is: het gebruik van de PDF standaarden voor spreadsheets is vanuit het oogpunt van toegankelijkheid en behoud van functionaliteit behoorlijk problematisch. Zie bijvoorbeeld onderstaand blogje dat ik daar eens over heb geschreven:

<http://openpreservation.org/blog/2016/12/09/pdfa-as-a-preferred-sustainable-format-for-spreadsheets/>

Een oplossing zou zijn om het toepassingsgebied hier nog iets specifieker te formuleren / af te bakenen.

Toegankelijk voor wie (of wat)?

Wat betreft "duurzame toegankelijkheid": een overweging die ik in het advies mis, is wie of wat (!) de toekomstige gebruiker is, en de manier waarop deze optimaal toegang heeft tot documenten. Het huidige advies lijkt sterk toegespitst op de use case van een menselijke gebruiker die een document vanuit een viewerapplicatie leest. Dit wordt inderdaad prima afgedekt met PDF/A. Maar voor toepassingen als geautomatiseerde verwerking, analyse en text-mining leveren de PDF formaten (inclusief PDF/A) veel problemen op, en vaak gaat dit juist veel makkelijker met bronformaten als docx.

Zie voor een hele berg voorbeelden van de nadelen van PDF voor dergelijke toepassingen de Twitterfeed van Peter Murray-Rust van het ContentMine programma (<http://contentmine.org/>):

<https://twitter.com/search?q=PDF%20from%3A%40petermurrayrust&src=typd>

Verder een simpel voorbeeld uit onze eigen praktijk: een tijdje terug wilden wij ten behoeve van onze web-archiveringsactiviteiten alle URLs uit een recent proefschrift extraheren. In eerste instantie gebruikten we hiervoor de PDF-versie van het proefschrift. Hierbij bleken URLs in voetnoten die over meerdere regels liepen, bij de geregeleindes te worden afgebroken. Dit probleem trad op bij meerdere linkextractietools. Linkextractie op het originele MS Word brondocument leverde dit probleem niet op.

Waar het eigenlijk vooral om gaat is dat PDF/A geen "silver bullet" is die alles op het gebied van (duurzame) toegankelijkheid afdekt; het is vaak onvermijdelijk dat je voor verschillende gebruikstypen de informatie soms in verschillende formaten nodig hebt.

Origineel bewaren vs converteren

In aanvulling hierop: de voorlaatste paragraaf ("Overigens kan de kans" ... "of het afwegingskader.") impliceert min of meer dat converteren versus het originele document intact laten een of/of keuze is. Maar je kunt natuurlijk ook allebei doen (bv origineel Word document bewaren + een hiervan afgeleide PDF/A). En in veel gevallen is dat ook de veiligste keus.

PDF/A-1 vs PDF/A-2

Over de twee PDF/A standaarden zegt het expertadvies nu:

PDF/A-1 moet worden toegepast op de uitwisseling en publicatie van niet- of beperkt reviseerbare documenten, waarbij duurzame toegankelijkheid van belang is.

PDF/A-2 moet worden toegepast op de uitwisseling en publicatie van niet- of beperkt reviseerbare documenten, waarbij duurzame toegankelijkheid van belang is en waarbij PDF/A-1 onvoldoende functionaliteit biedt.

Naar mijn mening maakt dit advies het de beoogde gebruiker van het expertadvies onnodig moeilijk. Want om welke functionaliteit gaat het hier precies, en hoe beoordeelt de gebruiker of dit een rol speelt? Verder impliceert dit een voorkeur voor PDF/A-1 boven A-2, terwijl daar volgens mij geen goede redenen voor zijn (PDF/A-1 is niet "duurzamer" dan PDF/A-2). Ik zou dit oplossen door in principe PDF/A-2 aan te bevelen (want: minste kans dat conversie niet lukt door ontbrekende functionaliteit), met een toevoeging dat PDF/A-1 ook gebruikt mag worden, maar dat deze variant op bepaalde punten minder functionaliteit biedt dan PDF/A-2, en dat de conversie hierdoor niet altijd goed mogelijk is. Voor een gebruiker lijkt dit me een stuk eenvoudiger.

8

Wilt u aanvullende informatie of overwegingen meegeven aan het Forum Standaardisatie en Overheidsbreed Beleidsoverleg Digitale Overheid die van invloed zouden kunnen zijn op het besluit om de omschrijvingen van de functioneel toepassingsgebieden van AdES Baseline Profiles, CMIS, DigiToegankelijk (EN 301 549 met WCAG 2.0), OWMS, PDF 1.7, PDF/A-1, PDF/A-2 en SKOS aan te passen op de 'pas toe of leg uit' lijst?

Zie opmerkingen onder 6.