

OVERZICHT VAN OFFICE OPEN XML



ECMA TC45
TOM NGO (NEXTPAGE), REDACTEUR

1 INLEIDING

Office Open XML (Open XML) is een ter goedkeuring voorgelegde open standaard voor tekstverwerkingsdocumenten, presentaties en spreadsheets, die in meerdere toepassingen en op meerdere platformen vrij kan worden gebruikt. De standaardisatie van Open XML is onder meer van belang voor organisaties die toepassingen willen implementeren waarin deze indeling kan worden gebruikt, commerciële en overheidsinstellingen die gebruikmaken van dergelijke software, en leraren en auteurs die hun kennis over deze indeling overdragen. Uiteindelijk zullen echter alle gebruikers voordeel hebben bij een XML-standaard voor hun documenten, bijvoorbeeld op het gebied van stabiliteit, opslag en uitwisselbaarheid en de voortschrijdende ontwikkeling.

Het standaardisatiewerk voor Open XML werd uitgevoerd door het Technical Committee 45 (TC45) van Ecma International, waarin Apple, Barclays Capital, BP, de British Library, Essilor, Intel, Microsoft, NextPage, Novell, Statoil, Toshiba en de Amerikaanse Library of Congress zijn vertegenwoordigd (1).

Deze whitepaper biedt een overzicht van Open XML en behandelt de volgende thema's:

- Inzicht in de doelstellingen van Open XML en de structuur van de specificatie.
- Kennis van de eigenschappen: hoe gaat Open XML om met achterwaartse compatibiliteit, opslag, uitbreiding, aangepaste schema's, subsetting, meerdere platformen, internationalisatie en toegankelijkheid?
- Informatie over het volgen van de high-level structuur van een Open XML-bestand en snel navigeren naar een onderdeel van de specificatie waarover u nadere informatie wilt.

2 DE AANLEIDINGEN VOOR STANDAARDISATIE

Open XML is vanaf het prille begin ontworpen om de bestaande tekstverwerkingsdocumenten, presentaties en spreadsheets, gecodeerd in door Microsoft Corporation gedefinieerde binaire bestandsindelingen, correct weer te geven. Tijdens het standaardisatieproces werden de vereiste voorzieningen voor weergave van de bestaande bestandsindelingen opgenomen in XML en vervolgens uitgebreid, werd gedetailleerde informatie verstrekt en interoperabiliteit mogelijk gemaakt. Op dit moment hebben meer dan 400 miljoen gebruikers naar schatting al meer dan 40 miljard documenten in binaire bestandsindelingen geproduceerd, terwijl hier elk jaar nog eens miljarden documenten bijkomen.

De oorspronkelijke binaire indelingen van deze bestanden werden ontworpen in een periode dat schijfruimte schaars was en de benodigde tijd voor het parseren een negatief effect had op de gebruikerservaring. De bestandsindelingen waren gebaseerd op directe serialisatie van de gegevensstructuren in het geheugen die door Microsoft® Office®-toepassingen werden gebruikt. De huidige infrastructuur van hardware, netwerken en standaarden, en met name XML, laat echter een nieuw ontwerp toe, dat meerdere producenten op meerdere platformen de ruimte geeft en verdere ontwikkeling mogelijk maakt.

Gelijke tred houdend met de technologische vorderingen heeft ook het marktaanbod zich verbreed en zijn er nieuwe toepassingen ontwikkeld, die in het oorspronkelijke, eenvoudige ontwerp van programma's voor documentbewerking niet waren voorzien. Deze nieuwe toepassingen kunnen bijvoorbeeld:

- Automatisch documenten genereren aan de hand van bedrijfsgegevens.
- Bedrijfsgegevens extraheren uit documenten en de gegevens overbrengen naar bedrijfstoepassingen.
- Afgebakende taken uitvoeren die op een kleine subset van een document betrekking hebben, maar waarmee het document toch bewerkbaar blijft.
- Documenten toegankelijk maken voor gebruikers met fysieke beperkingen, zoals blinden en slechtzienden.
- Op verschillende typen hardware worden uitgevoerd, inclusief mobiele apparaten.

De wellicht grootste uitdaging heeft betrekking op de langetermijnopslag. Wij zijn eraan gewend geraakt steeds meer informatie te maken. We hebben deze informatie echter steeds gecodeerd via digitale weergaven die zo nauw met de gebruikte programma's zijn verweven, dat de informatie na een jaar of twintig vrijwel niet meer leesbaar is zonder aanzienlijk gegevensverlies. Het veiligstellen van de financiële en intellectuele investeringen in deze documenten, zowel in de bestaande als de nieuwe, is dan ook uitgegroeid tot een belangrijke prioriteit.

Doordat zich nu de volgende vier omstandigheden voordoen - het gebruik van binaire bestandsindelingen in zeer brede kring, technologische vooruitgang, marktmechanismen die uiteenlopende toepassingen vereisen en het groeiende probleem van de langetermijnopslag - is de noodzaak ontstaan een open XML-indeling te ontwikkelen die migratie van de miljarden documenten met een minimum aan gegevensverlies mogelijk maakt. De standaardisatie van deze open XML-indeling en het onderhouden van de standaard op de lange termijn maakt een omgeving mogelijk waarin elke organisatie kan vertrouwen op de blijvende stabiliteit van de specificatie, in de wetenschap dat het standaardisatieproces zorgt voor evenwichtige verdere ontwikkelingen.

Er bestaan reeds diverse documentstandaarden en -specificaties, zoals HTML, XHTML, PDF en subsets daarvan, ODF, DocBook, DITA en RTF. Net als bij de talloze standaarden voor bitmap-afbeeldingen, zoals TIFF/IT, TIFF/EP, JPEG 2000 en PNG, kwam elk van deze standaarden en specificaties voort vanuit andere doelstellingen. Open XML voorziet in de behoefte aan een standaard voor de voorzieningen in de bestaande documentindelingen. Voor zover ons bekend, biedt alleen XML ondersteuning voor alle voorzieningen in binaire bestanden.

Open XML bevat definities voor de indelingen van tekstverwerkings-, presentatie- en spreadsheetdocumenten. Elk van deze documenttypen wordt gespecificeerd via een primaire opmaaktaal, de primary mark-up language: WordprocessingML, PresentationML of SpreadsheetML. Door middel van ingesloten methoden kan een document dat tot een van deze drie typen behoort, ook gegevens bevatten die in de andere opmaaktalen of in een van de vele ondersteunende opmaaktalen zijn gespecificeerd.

De specificatie omvat zowel normatieve gegevens die via Open XML worden gedefinieerd als informatieve gegevens die het begrip van de lezer ten goede komen en niet zijn voorgeschreven. Op grond van de uiteenlopende gebruikersbehoeften werd de structurering opgesplitst in verschillende “parts”.

Part 1 - Basiseigenschappen 165 pagina's	<ul style="list-style-type: none"> — Definieert de woordenlijst, notatieconventies en afkortingen. — Vat de drie primaire opmaaktalen en ondersteunende opmaaktalen samen. — Creëert voorwaarden voor conformiteit en biedt richtlijnen voor interoperabiliteit. — Beschrijft de beperkingen in de Open Packaging Conventions die op elk van de documenttypen betrekking hebben.
Part 2 - Open Packaging Conventions 125 pagina's	<ul style="list-style-type: none"> — Definieert de Open Packaging Conventions (OPC). Elk Open XML-bestand bevat een verzameling bytestromen, “parts” genaamd, die worden gecombineerd in een container, die “package” wordt genoemd. De indeling van een package wordt gedefinieerd door de OPC. — Beschrijft een aanbevolen fysieke implementatie van de OPC die de bestandsindeling ZIP gebruikt. — Declareert de XML-schema's voor de OPC als XML Schema Definitions (XSD) (2), in een appendix die uitsluitend in elektronische vorm wordt uitgegeven. De appendix bevat ook niet-normatieve weergaven van de schema's, mogelijk gemaakt door RELAX NG (ISO/IEC 19757-2) (3).
Part 3 - Primer 466 pagina's	<ul style="list-style-type: none"> — Introduceert de voorzieningen van elk van de opmaaktalen en biedt context en illustratieve elementen via voorbeelden en grafische voorstellingen. Dit part dient ter informatie en is niet-normatief. — Beschrijft de voorziening voor het opslaan van aangepaste XML-gegevens binnen een package, waardoor de integratie van bedrijfsgegevens wordt ondersteund.
Part 4 – Referentie voor mark-up language 5756 pagina's	<ul style="list-style-type: none"> — Definieert elk element en kenmerk, de hiërarchie van parent/child-relaties van elementen en zo nodig de aanvullende semantiek. Dit part is bedoeld voor gebruik ter referentie wanneer volledige informatie over een element of kenmerk is vereist.

- Definieert de voorziening voor het opslaan van aangepaste XML-gegevens.
 - Declareert de XML-schema's voor de opmaaktalen als XSD (2), in een appendix die uitsluitend in elektronische vorm wordt uitgegeven. Deze appendix biedt ook niet-normatieve weergaven van de schema's door gebruik te maken van RELAX NG (ISO/IEC 19757-2) (3).
- Part 5 - Mark-up
compatibiliteit
en uitbreidbaarheid
34 pagina's
- Beschrijft de voorzieningen voor uitbreiding van Open XML-documenten.
 - Specificeert elementen en kenmerken, waardoor toepassingen met verschillende uitbreidingen interoperabel worden.
 - Duidt de regels voor uitbreidbaarheid aan met behulp van NVDL (ISO/IEC 19757-4) (4).

De elektronische versies bevatten een groot aantal intern actieve koppelingen, die het lezen en navigeren door deze documenten vergemakkelijken. Met name part 4 bevat overal koppelingen naar parent/child-elementen.

4 DE EIGENSCHAPPEN VAN DE STANDAARD

Dit hoofdstuk biedt een inleiding op nader onderzoek van Open XML aan de hand van een aantal high-level eigenschappen. Elk van de volgende gedeelten beschrijft een van deze eigenschappen en verwijst naar specifieke voorzieningen in Open XML.

- In het gedeelte “Interoperabiliteit” wordt beschreven hoe Open XML onafhankelijk werkt van merkgelabelde bestandsindelingen, functies en de runtime-omgeving, zodat ontwikkelaars over een grote keuzevrijheid beschikken.
- In het gedeelte “Internationalisatie” wordt een klein aantal karakteristieken genoemd van de wijze waarop Open XML ondersteuning biedt voor alle belangrijke taalgroepen.
- In de gedeelten “Laagdrempelig voor ontwikkelaars”, “Compactheid” en “Modulariteit” worden de specifieke manieren beschreven waarop Open XML de praktische hindernissen voor gebruik door uiteenlopende groepen omzeilt of wegneemt: leercurve, minimale functieset, prestaties.
- In het gedeelte “Betrouwbare migratie” wordt beschreven hoe Open XML beantwoordt aan de overkoepelende doelstelling dat de informatie, inclusief alle intenties van de oorspronkelijke auteur, bewaard moet blijven in bestaande en nieuwe documenten.
- In het gedeelte “Integratie met bedrijfsgegevens” wordt beschreven hoe Open XML zakelijke informatie opneemt in aangepaste schema's om integratie en hergebruik van gegevens tussen productiviteitstoepassingen en informatiesystemen mogelijk te maken.
- In het gedeelte “Ruimte voor innovatie” wordt beschreven hoe Open XML inspelt op toekomstige ontwikkelingen door zowel verdere uitbreidingsmethoden te specificeren als interoperabiliteit tussen toepassingen met verschillende functiesets mogelijk te maken.

De overige hoofdstukken in dit document, inclusief het gedeelte dat u nu leest, bieden een thematische leidraad voor Open XML. Verwijzingen naar de specificatie worden weergegeven als §Part: sectie. subsectie - zo verwijst bijvoorbeeld §1:2.5 naar part 1, sectie 2.5 van de specificatie. Voor verwijzingen naar andere titels in dit artikel wordt de naam gebruikt.

4.1 INTEROPERABILITEIT

Ontwikkelaars hebben de mogelijkheid toepassingen te bouwen die op meerdere platformen Open XML kunnen gebruiken of creëren.

Als eerste moet worden opgemerkt dat de interoperabiliteit van Open XML tot stand is gekomen via een groot aantal bijdragen, aanpassingen en herzieningen van de specificatie door leden van de Ecma TC45-commissie (1) met uiteenlopende achtergronden en bedrijfsbelangen. Vertegenwoordigd waren:

- Producenten (Apple, Intel, Microsoft, NextPage, Novell en Toshiba) van verschillende besturingssystemen (Linux, MacOS en Windows) met verschillende gebruiksdoelen voor Open XML.
- Ondernemingen (BP, Barclays Capital, Essilor en Statoil) met hoge investeringen in bestaande content, inclusief bedrijfskritische transactiesystemen.
- De British Library en de United States Library of Congress, die beide een direct belang hebben bij voorzieningen voor langetermijnopslag.

Gedurende de voorbereidingsperiode werden door de commissieleden honderden problemen op het gebied van beleid, transparantie, semantiek en mogelijke omgevingsafhankelijkheid aangekaart en opgelost. Ter illustratie een aantal van de problemen en daarmee samenhangende acties:

- Voorzieningen ter ondersteuning van de platformafhankelijkheid van methoden die in de oorspronkelijke binaire bestandsindelingen merkgelinkt waren.
- Voorwaarden voor conformiteit.
- Inhoud van de schema's.
- Alternatieve weergaven van de schema's en uitbreidingsmethoden door gebruik te maken van RELAX NG (ISO/IEC 19757-2) en NVDL (ISO/IEC 19757-4) (4).
- Ontwikkeling van tools voor automatische analyse en visualisatie van de schema's.
- Internationalisatie.
- Volledigheid, juistheid en transparantie van alle beschrijvingen in de specificatie; in veel gevallen doordat werd geprobeerd delen van de specificatie te implementeren.

De verdere tekst in dit gedeelte beschrijft een aantal specifieke deelgebieden waarin Open XML uitgaat van de oorspronkelijke binaire bestandsindelingen om de interoperabiliteit te waarborgen.

Een van de belangrijkste voorwaarden voor interoperabiliteit is onafhankelijkheid van een bepaald type broncontent.

- Open XML kent geen beperkingen voor afbeeldings-, audio- en videotypen. Voor bijvoorbeeld afbeeldingen zijn alle typen toegestaan: GIF, PNG, TIFF, PICT, JPEG of een willekeurig ander afbeeldingstype (§1:14.2.12).
- Alle typen ingesloten besturingselementen zijn toegestaan, zoals Java of ActiveX (§1:15.2.8).
- De lettertypespecificaties in WordprocessingML mogen informatie over de afmetingen en Panose-nummers bevatten, zodat gemakkelijker een vervangend lettertype kan worden gevonden als het oorspronkelijke type niet meer beschikbaar is (§3:2.10.5).

Daarnaast werkt Open XML onafhankelijk van de runtime-omgeving van de applicatie waarin een document is gemaakt.

- Een klassiek voorbeeld hiervan doet zich voor wanneer externe besturingselementen of toepassingen een afbeelding genereren die slechts een gedeelte van het beeldscherm beslaat. Ter voorkoming van situaties waarin het besturingselement of de applicatie niet beschikbaar is of niet kan draaien in een bepaalde runtime-omgeving, kan in het documentbestand een weergave van de afbeelding worden opgeslagen. Deze methode was ook beschikbaar in de oudere binaire bestandsindelingen.
- Open XML introduceert echter een meer algemene methode met de naam Alternate Content Block (§3:2.18.4), die kan worden gebruikt in diverse situaties waarin een ontvangende applicatie mogelijk niet in staat is de content van een producerende applicatie correct te interpreteren. Doorgaans zal deze aanpak worden toegepast in het kader van uitbreidbaarheid. Een nadere beschrijving van deze methode vindt u in het gedeelte “Ruimte voor innovatie”.
- Daarnaast werkt Open XML onafhankelijk van parameters die zinvol zijn in de omgeving waarin een document wordt gemaakt, maar mogelijk geen betekenis hebben in de omgeving van ontvangende toepassingen. Zo bevat bijvoorbeeld de parameter CT_SYSCOLOR een index voor een kleurentabel in de omgeving waarin een document wordt gemaakt. Ter ondersteuning van de portabiliteit met andere omgevingen kan de producerende applicatie in PresentationML de systeemkleuren die tijdens het maken van document actief waren, opslaan in het cachegeheugen.

Ter afsluiting van dit gedeelte het belangrijkste aspect: Office Open XML is afgestemd op open W3C-standaarden, zoals XML (5) en XML Namespaces (6). Dit gegeven maakt een basisniveau van interoperabiliteit mogelijk met alle platformen en besturingssystemen die van deze open standaarden gebruikmaken.

4.2 INTERNATIONALISATIE

Open XML ondersteunt de voorzieningen voor internationalisatie die vereist zijn voor talen als Arabisch, Chinees (drie varianten), Hebreeuws, Hindi, Japans, Koreaans, Russisch en Turks.

Vanzelfsprekend biedt Open XML ook ondersteuning voor Unicode, dat eveneens op XML is gebaseerd. Bovendien beschikt Open XML over een uitgebreide functieset voor internationalisatie, die in de loop van vele jaren steeds verder is verfijnd. Ter illustratie het volgende overzicht:

Tekstrichting - Open XML ondersteunt talen die van links naar rechts (LTR) of van rechts naar links (RTL) worden geschreven. Ook voor bidirectionele (“BiDi”) talen, zoals het Arabisch, Farsi, Urdu, Hebreeuws en Jiddisch, is ondersteuning aanwezig. Deze talen worden weliswaar van rechts naar links geschreven, maar kunnen ook tekst bevatten die van links naar rechts wordt geschreven. In WordprocessingML kan de tekstrichting zowel worden bepaald op alinea-niveau (§4:2.3.1.6) als op het niveau van een run (de tekststroom) binnen een alinea (§4:2.3.2.28). Voor DrawingML-tekst kan de tekstrichting worden bepaald op hoofdstekstniveau (§4:5.1.5.1.1), alinea-niveau (§4:5.1.5.2.2) en binnen genummerde opsommingstekens (§4:5.1.5.4).

Tekstdoorloop - In WordprocessingML kan de richting van de tekstdoorloop zowel worden bepaald op sectie- of tabelniveau (§4:2.3.1.41) als op alinea-niveau (§4:2.3.2.28). Op sectie- en tabelniveau kan de tekstdoorloop worden bepaald in verticale en horizontale richting. Hierdoor biedt Open XML ondersteuning voor elke mogelijke tekstmaak, bijvoorbeeld ook voor het Mongools, dat wordt geschreven in verticale regels die van boven naar beneden lopen en van links naar rechts zijn gerangschikt. Deze opmaken hebben betrekking op lijsten, tabellen en andere presentatie-elementen. DrawingML kan ook gebruikmaken van Kumimoji-instellingen op alinea- en tekststroomniveau, waarmee tekst horizontaal en getallen verticaal worden weergegeven (§4:5.1.5.2.3, §4:5.1.5.3.9). In WordprocessingML (§4:2.3.1.16) en PresentationML (§4:4.3.1.15) kan de tekenstroom ook worden gespecificeerd aan de hand van Kinsoku-instellingen, zodat kan worden ingesteld met welke teken een tekstregel mag beginnen of eindigen.

Getalnotatie - Voor het opmaken van velden in WordprocessingML (§4:2.16.4.3), nummeren van alinea's en lijsten in WordprocessingML (§4:2.9) en nummeren in DrawingML (§4:5.1.5.4, §4:5.1.12.61) zijn enkele tientallen getalnotaties beschikbaar, zoals Hiragana, Arabisch, Abjad, Thais, voluit geschreven hoofdtelwoorden (bijvoorbeeld "honderdrieëntwintig"), Chinees, Hebreeuws, Hindi, Japans, Koreaans (Chosung en Ganada), Romeins en Vietnamees. Deze voorzieningen bieden eveneens ondersteuning voor de weergave van decimaalteken (bijvoorbeeld "1.00" t.o.v. "1,00") en lijstscheidingstekens. Ook in SpreadsheetML is de internationalisatie van getalnotaties krachtig doorgevoerd en worden alle voorzieningen zowel in de celopmaken (§4:3.8.30) als in verwijzingen naar externe gegevens (§4.3.13.12) ondersteund.

Datumnotatie - In WordprocessingML (§4:2.18.7) en SpreadsheetML (§4:3.18.5) zijn de volgende notaties voor kalenderdatums beschikbaar: Gregoriaans (drie varianten), Hebreeuws, Hijri, Japans (keizerrijk), Koreaans (Tangun-tijdperk), Saka, Taiwanees en Thais.

Formules - Voor het specificeren van formules in SpreadsheetML zijn diverse, aan internationalisatie gerelateerde conversiefuncties beschikbaar, zoals BAHTTEXT (§4:3.17.7.22), JIS (§4:3.17.7.185) en ASC (§4:3.17.7.11).

Taalidentificatie - In WordprocessingML (§4:2.3.2.18) en DrawingML (§4:5.1.5.3) kan elke alinea en run (tekststroom) worden gemarkeerd met een taal-id, waardoor toepassingen automatisch de juiste taalcontroleprogramma's en andere taalspecifieke functies kunnen selecteren. Naast de taal-id voor elke taal kunnen in Open XML ook tekensets, lettertypefamilies en Panose-nummers worden benoemd, zodat toepassingen gemakkelijker een vervangende tekenset kunnen selecteren wanneer lokale ondersteuning niet mogelijk is.

4.3 LAAGDREMPELIG VOOR ONTWIKKELAARS

Ervaren ontwikkelaars hoeven niet meer dan een paar uur te besteden aan het lezen van de specificatie om eenvoudige Open XML-conforme toepassingen te kunnen schrijven.

Hoewel de specificatie een grote functieset beschrijft, hoeft een Open XML-conforme applicatie niet alle functies in de specificatie te ondersteunen. In de conformiteitsverklaring (§1:2) wordt alleen de eis gesteld dat een conforme ontvangende applicatie "geen conforme documenten van het [verwachte] documenttype weigert" en een conforme producerende applicatie "in staat moet zijn conforme documenten te produceren" (§1:2.5). Daarnaast worden richtlijnen voor interoperabiliteit gegeven, waarin de betekenis van de semantiek voor elementen wordt omschreven (§1:2.6).

Een conforme applicatie kan over uiterst specifieke functionaliteit beschikken. De applicatie kan bijvoorbeeld een batchverwerker zijn die geen andere functie heeft dan het actualiseren van copyrightverklaringen in een verzameling tekstverwerkingsdocumenten, maar ook een tekst-naar-spraak lezer die is afgestemd op diavoorstellingen en teksten naar audio kan omzetten terwijl de gebruiker de dia's een voor een bekijkt. De structuur van de bestandsindeling maakt het mogelijk dat dergelijke programma's al met minimale kennis over Open XML kunnen worden geschreven. Specifieke kenmerken:

- De bestandsindeling is conform gevestigde standaarden, met name XML en ZIP, waarvoor volledig doorontwikkelde tools bestaan.
- De bestandsindeling maakt gebruik van de Open Packaging Conventions, die XML en ZIP combineren met standaardmethoden om de relaties binnen een bestand uit te drukken. Hierdoor kan in veel gevallen door de bestandsinhoud worden genavigeerd zonder dat inzicht in de tag-semantiek van de primaire of ondersteunende opmaaktalen in Open XML is vereist.
- Diep geneste elementen in de XML-boomstructuur zijn toegankelijk en kunnen worden aangepast zonder de totale structuur aan te tasten.

Compacte gegevens overall in de bestandsindeling, die soms niet aanwezig waren in de binaire indelingen, bieden door middel van waarden in het cachegeheugen ondersteuning voor toepassingen met minimale functionaliteit. Voorbeelden:

- Wanneer geen gebruik wordt gemaakt van een paginator kan bijvoorbeeld een leesapplicatie voor blinden toch paginanavigatie mogelijk maken op basis van de laatst berekende pagina-einden (§4:2.3.3.13).
- Wanneer geen formules worden gebruikt en geïntegreerd met een externe gegevensbron, kan een spreadsheetprogramma toch werken op basis van in het cachegeheugen geplaatste berekeningen (§3:3.2.9) en externe gegevens (§4:3.14 en §4:3.10.1.76).

Een document met minimale conformiteit is uiterst eenvoudig. Zie voor meer informatie het gedeelte “Minimaal WordprocessingML-document”.

4.4 COMPACTHEID

De Open XML-bestandsindeling ondersteunt het maken van high-performance toepassingen. Dit gedeelte beschrijft een aantal ontwerpkenmerken waardoor compacte bestanden mogelijk worden die de verwerking en parsering versnellen. In het volgende gedeelte wordt beschreven hoe een applicatie dankzij modulaire bestandsstructuren een groot aantal taken kan uitvoeren door slechts een kleine subset van een document te parsen of aan te passen.

Open XML-bestanden worden doorgaans opgeslagen in een ZIP-archief en ingepakt en gecomprimeerd, conform de aanbevolen werkwijze in de Open Packaging Conventions. Het wellicht verrassende gevolg hiervan is dat Open XML-bestanden gemiddeld 25% kleiner zijn, en soms zelfs 75% kleiner, dan de binaire versies. In de binaire bestandsindeling zou deze whitepaper bijvoorbeeld 85 procent groter zijn!

Een andere eenvoudige bron voor compactheid, vooral geschikt wanneer de weergave niet wordt gecomprimeerd, is de lengte van de id's in de XML-code. Vaak gebruikte tag-namen moeten kort zijn en het is raadzaam ook korte naamruimtevoorvoegsels te gebruiken. Zo is “w” bijvoorbeeld het gebruikelijke voorvoegsel voor de naamruimte in WordprocessingML.

Verdere compactheid wordt mogelijk door herhalingen in de totale bestandsindeling te voorkomen.

Eén voorbeeldklasse voorkomt redundante opslag van grote objecten:

- In SpreadsheetML worden herhaalde tekenreeksen opgeslagen in een teken tabel in het werkboek en wordt gebruikgemaakt van indexverwijzingen (§3:3.3).
- In SpreadsheetML wordt een formule die op meerdere cellen van toepassing is, als één basisformule opgeslagen in de cel linksboven, terwijl de overige cellen binnen het formulebereik via een groepeerindex naar de basisformule verwijzen (§3:3.2.9.2).
- In DrawingML worden shape-namen (§4:5.1.12.56), tekstconfiguraties (§4:5.1.12.76) en andere vooraf ingestelde waarden (meerdere in §3:5.8, §3:5.9 en §4:5.1.12) niet in detail beschreven, maar via een naam of nummer. De betekenissen van de namen en getallen bevinden zich hierbij niet in het bestand, maar in de specificatie. Deze manier van verwijzen is het resultaat van een expliciet compromis tijdens het standaardisatieproces. De gekozen aanpak is echter compact en maakt bewerkingen op het juiste abstractieniveau mogelijk: door slechts één attribuut te wijzigen, kan een rechthoek bijvoorbeeld worden veranderd in een ovaal (§4:5.1.11.18).

In een andere voorbeeldklasse wordt hiërarchie gebruikt om overerving van semantiek mogelijk te maken.

Een gunstig neveneffect hiervan is dat de prestaties verbeteren doordat bestanden kleiner worden.

- In WordprocessingML zijn stijlen hiërarchisch (§3:2.8.9).
- In DrawingML zijn shapes hiërarchisch gegroepeerd (§4:5.1.2.1.20).
- In PresentationML worden diamodelen, dia-indelingen en dia's verbonden door een standaardhiërarchie (§3:4.2).

Ook andere aspecten van Open XML zijn ontworpen om een efficiënt gebruik mogelijk te maken.

Zo worden bijvoorbeeld in de celtabellen van SpreadsheetML uitsluitend niet-lege cellen opgeslagen, terwijl samengevoegde cellen worden beschreven als een eenheid. Deze techniek maakt voor spreadsheets met grote spreiding een significante afname van de bestandsgrootte mogelijk.

4.5 MODULARITEIT

Toepassingen kunnen een groot aantal taken uitvoeren door slechts een kleine subset van een document te parsen of aan te passen.

De Open XML-indeling bevat drie functies die samenwerken om deze modulariteit mogelijk te maken.

- Een document is niet één consistent geheel, maar opgebouwd uit meerdere onderdelen.
- De relaties tussen de diverse onderdelen worden opgeslagen in parts.
- Het archiveringsindeling ZIP, dat doorgaans wordt gebruikt voor Open XML-documenten, biedt ondersteuning voor willekeurige toegang tot elk van de parts.

Voorbeelden:

- Een applicatie kan probleemloos de weergavepositie van dia's verplaatsen, samen met de bijbehorende bronnen zoals afbeeldingen en lay-outs, zonder dat hiervoor de content van de dia's hoeft te worden geparseerd (§3:13.3.8). In dit voorbeeld worden de dia's en bijbehorende bronnen gevonden via gegevens die "expliciete relaties" worden genoemd. Expliciete relaties worden gedefinieerd door de Open Packaging Conventions en kunnen zonder enige kennis van de tag-semantiek in PresentationML worden geparseerd (§1:9.2, §2:8.3).
- Een applicatie kan alle opmerkingen uit een WordprocessingML-document verwijderen, zonder dat hiervoor content hoeft te worden geparseerd (§1:11.3.2). In dit voorbeeld worden de opmerkingen gevonden via gegevens die "impliciete relaties" worden genoemd. Impliciete relaties zijn specifiek voor Open XML en vereisen dan ook enige kennis van de desbetreffende mark-up language (§1:9.2).

4.6 BETROUWBARE MIGRATIE

Open XML is ontworpen om ondersteuning te bieden voor alle functies in de binaire bestandsindelingen van Microsoft Office 97 t/m 2003.

Doordat het realiseren van deze doelstelling gepaard ging met tal van levensgrote problemen, is het des te opmerkelijker dat Open XML toch in staat blijkt deze ongeëvenaarde prestatie te leveren. Sommige indelingen, zoals PDF, zijn ontworpen om eindgebruikers een exacte visuele kopie van een voltooid document te bieden. Open XML is echter juist bedoeld om toekomstige bewerkingen en handelingen toe te staan op het abstractieniveau waarover de maker van het oorspronkelijke document kon beschikken. Het herleiden van een vectorafbeelding naar een bitmap zou bijvoorbeeld niet passen binnen deze opzet, evenmin als het samenvouwen van een stijlhiërarchie tot onafhankelijke stijlen. Ook is het mogelijk dat een document calculatiesemantiek bevat die de oorspronkelijke maker wilde behouden, zoals formulelogica die via tussentijdse berekeningen tot stand komt, inclusief de foutcodes of animatieregels die dynamisch gedrag teweegbrengen.

De volgende verwijzingen naar de specificatie zijn illustratief voor de mogelijkheden van Open XML om subtiele aspecten van binaire bestandsindelingen weer te geven:

- De SpreadsheetML-beschrijving bevat een uitgebreide formulespecificatie (§4:3.17.7).
- De WordprocessingML-specificatie bevat informatie over de gevolgde regels voor het samenstellen van de eigenschappen van alinea's, tekens, nummering en tabellen met directe opmaak (§3:2.8, in het bijzonder §3:2.8.10).
- De PresentationML-specificatie bevat informatie over de animatiefuncties (§3:4.4).

Open XML maakt meerdere conforme werkwijzen mogelijk, zonder dat overeenkomst met elk afwijkend detail noodzakelijk is. Dit is vooral van belang wanneer numerieke berekeningen een rol spelen, zoals bij lay-outs, het genereren van effecten en evalueren van formules. Als meer consistentie zou worden verlangd dan in de praktijk nodig is, zou het ontwikkelaars onnodig moeilijk gemaakt worden om conformiteit te bereiken. De volgende verklaringen onderstrepen een aantal besluiten die de commissie over dit onderwerp heeft genomen:

- Open XML definieert effecten zoals oppervlakweergaven (§5.1.12.50), maar verplicht de ontwikkelaar niet om deze effecten op pixelniveau te laten overeenkomen.
- Open XML definieert parameters zoals paginamarges (§4:2.6.11), lettertype (§4:2.8) en invullen (§4:2.3.1.13). Ontwikkelaars kunnen in Open XML verschillende flow-algoritmen gebruiken, op voorwaarde dat deze parameters worden nageleefd.
- De SpreadsheetML-formulespecificatie (§4:3.17.7) probeert niet om verschillen in drijvende-komma-berekeningen op te heffen, omdat de conformerende toepassingen hiervoor veelal langzame emulatie zouden moeten gebruiken in plaats van de oorspronkelijke hardware. In plaats daarvan wordt het minimumaantal bits voor de exactheid van numerieke berekeningen gespecificeerd (§4:3.17.5).
- De SpreadsheetML-formulespecificatie laat ook bepaalde voorwaardelijke beslissingen afhangen van het gebruik, zodat toekomstige innovaties mogelijk worden. Zo zijn er bijvoorbeeld geen beperkingen voor het aantal keren dat een berekening zoals NORMINV (§4:3.17.7.227) mag worden herhaald. (De NORMINV-functie voert door middel van een iteratieve zoekactie de normale distributie uit in omgekeerde volgorde.)

Een aantal oudere functies, zoals VML (§3:6), is hoofdzakelijk opgenomen om achterwaartse compatibiliteit mogelijk te maken. Voor het schrijven van nieuwe documenten wordt het gebruik aanbevolen van de moderne standaarden die al in Open XML zijn opgenomen, zoals DrawingML (§3:5).

4.7 INTEGRATIE MET BEDRIJFSGEGEVENS

Doordat in Open XML-documenten aangepaste schema's kunnen worden gebruikt, stelt Open XML organisaties in staat productiviteitstoepassingen te integreren met de informatiesystemen voor het beheren van bedrijfsprocessen. Zo kan een organisatie bijvoorbeeld streven naar hergebruik en geautomatiseerde verwerking van bedrijfsgegevens, die anders vrijwel onvindbaar ingebed zouden zijn in documenten waar ze niet door de bedrijfstoepassingen kunnen worden gelezen of opgeslagen.

Enkele beschikbare toepassingen zijn:

- *Zoeken* - Een eindgebruiker kan bijvoorbeeld in een verzameling spreadsheets zoeken naar bedrijven met winstmarges van meer dan 20 procent.
- *Metagegevens labelen* - Een bedrijf kan bijvoorbeeld goedgekeurde voorstellen labelen.
- *Documenten assembleren* - Een marketingteam kan bijvoorbeeld het aanmaken van voorstellen stroomlijnen door het verzamelen van de benodigde gegevens te automatiseren.
- *Gegevens hergebruiken* - Een salesmanager kan bijvoorbeeld een rapport aanmaken met alle verkoopcontracten in een bepaalde periode, inclusief een overzicht van de klanten, transactiewaarden en aangepaste voorwaarden.
- *Line-of-business toepassingen* - Professionals in een gespecialiseerde verticale organisatie kunnen hun werkzaamheden verrichten in een vertrouwde werkomgeving en de eindresultaten automatisch laten opnemen in het bedrijfssysteem.

Als bedrijven deze mogelijkheden willen benutten, moeten de gegevensstructuur en het gegevenstype voor de documentklasse worden gedefinieerd, terwijl de informatie op elke willekeurige plaats in de tekststroom van de desbetreffende documenten moet kunnen worden weergegeven. Als eenvoudig voorbeeld nemen we een curriculum vitae. Hiervoor wordt een gegevensstructuur gedefinieerd met de velden Naam, Telefoonnummer, Adres, Carrièreplanning en Diploma's. Vervolgens moet ervoor worden gezorgd dat de velden op de juiste posities verschijnen, op plaatsen waar mensen de veldinhoud normaal gesproken in documenten opnemen. Vanzelfsprekend zouden in andere bedrijfsomgevingen, zoals een financieringsmaatschappij of een ziekenhuis, waarschijnlijk andere keuzen voor de structuur en gegevensvelden worden gemaakt.

Open XML volgt voor dit proces echter een gestandaardiseerde aanpak.

Als eerste wordt de structuur van de bedrijfsgegevens aangeduid met behulp van een aangepast XML-schema. Dankzij dit schema kunnen organisaties ook gelabelde gegevens weergeven die vanuit zakelijk perspectief belangrijk zijn. Een organisatie kan zowel eigen schema's maken als standaardschema's voor de branche gebruiken, zoals XBRL voor financiële rapportages (7) of HL7 voor gezondheidszorginformatie (8). Schema's worden gebruikt in de publieke sector, binnen ondernemingen en als standaarden voor de branche, en voor tal van doeleinden, van geboortebewijzen tot verzekeringsinformatie, toegepast. Elk aangepast schema kan worden gebruikt, op voorwaarde dat het XSD-formaat wordt gehanteerd.

Vervolgens worden de aangepaste gegevens in een willekeurig Open XML-document ingesloten in een Custom XML-part (§3.7.3), waarna de gegevens kunnen worden beschreven in een part voor Custom XML Data Properties (§4:7.5). Omdat Open XML deze aangepaste gegevens afzonderlijk bewaart, kunnen schone gegevens worden geïntegreerd terwijl eindgebruikers de gegevens in een brede context, zoals in documenten, formulieren, dia's en spreadsheets, kunnen bekijken en bewerken. Hierdoor wordt interoperabiliteit op een elementairder en in semantisch opzicht nauwkeuriger niveau mogelijk.

4.8 RUIMTE VOOR INNOVATIE

Open XML is ontwikkeld om ontwikkelaars te stimuleren nieuwe toepassingen te maken, die nog niet werden overwogen toen de binaire Office-indelingen werden vastgelegd en zelfs tijdens het definiëren van Open XML nog niet speelden.

Als eerste worden in dit gedeelte de uitbreidingsmethoden besproken die samenwerken om interoperabiliteit tussen applicaties met verschillende functiesets mogelijk te maken. Als voorbeeld nemen we een *up-level* applicatie (deze bevat een nieuwe functie die niet in Open XML is gedocumenteerd) en een *down-level* applicatie (deze is niet compatibel met de nieuwe functie).

De drie belangrijkste doelen van uitbreidbaarheid zijn:

- *Visuele exactheid* - Het vermogen van de down-level applicatie om weer te geven wat de up-level applicatie zou weergeven. Hiervoor moet een bestand vanzelfsprekend meerdere weergaven van dezelfde gegevens opslaan.
- *Bewerkbaarheid* - De mogelijkheid om een of meer van de weergaven te bewerken.
- *Privacy* - De mogelijkheid om te waarborgen dat oude versies van een bepaalde weergave daadwerkelijk worden gewist nadat een andere weergave is bewerkt, zodat door de gebruiker verwijderde of aangepaste informatie niet onbedoeld kan blijven bestaan. Een applicatie kan dit mogelijk maken door weergaven te verwijderen of te synchroniseren.

Ontwikkelaars die de Open XML-functieset willen uitbreiden, beschikken over twee belangrijke mogelijkheden:

- *Alternatieve contentblokken* - In een alternatief contentblok (§3:2.18.4 en §5:9.2) kunnen meerdere weergaven van dezelfde content worden opgeslagen, waarbij elke weergave in het eigen blok wordt geplaatst. Vervolgens kan een down-level applicatie het blok inlezen waarmee het compatibel is. Tijdens het bewerken zal de down-level applicatie zo veel mogelijk blokken opslaan.
- *Uitbreidingslijsten* - In een uitbreidingslijst (§3:2.6) kan arbitraire, aangepaste XML worden opgeslagen, zonder visuele weergave.

Voor ontwikkelaars is voldoende ruimte beschikbaar om buiten deze beide uitbreidingsmethoden innovaties aan te brengen.

- *Alternatieve interactieparadigma's*: De specificaties in Open XML reiken verder dan de syntaxis van documenten, maar niet tot en met het applicatiegedrag. Zoals wordt beschreven in de conformiteitsverklaring, ligt het accent vooral op semantiek (§1:2.2, §1.2.3). Hierdoor kan een conformerende applicatie op een groot aantal manieren met de eindgebruiker communiceren, of juist alle communicatie uit de weg gaan, op voorwaarde dat de gespecificeerde semantiek wordt nageleefd.
- *Novel-computeromgevingen*: De conformiteitsverklaring laat zowel toepassingen toe die over een lage capaciteit beschikken, zodat deze kunnen draaien op kleine apparaten, als toepassingen die alleen een subset van Open XML gebruiken (§1:2.6). Met behulp van de methode Additional Characteristics kan een producerende applicatie informatie over de capaciteitsgrenzen verstrekken (§3:8.1).

Zoals in het vorige gedeelte al werd gesteld, hoeven voor een aantal van de belangrijkste innovatiemogelijkheden geen documenten te worden gegenereerd voor directe gebruikersinteractie. In plaats daarvan wordt de verwerking direct tussen computers uitgevoerd met behulp van XML-berichtindelingen, zoals XML Web Services (9). Hoewel dergelijke toepassingen naast het verwerken van de gegevens in Open XML-documenten geen voor de gebruiker zichtbare activiteiten ontplooiën, zijn toch zowel de louter syntactische documentconformiteit (§1:2.4) en applicatieconformiteit (§1:2.5) als de richtlijnen voor interoperabiliteit (§1:2.6), waarin de semantiek is vervat, van toepassing.

Het is niet mogelijk een overzicht te bieden van alle mogelijke gebruikssituaties waarin aangepaste XML wordt verwerkt. Niettemin mag worden verwacht dat bijvoorbeeld op XML gebaseerde services zullen worden ontwikkeld die Open XML-documenten verwerken voor het automatisch extraheren en invoegen van aangepaste gegevens. Ook aangepaste beveiligingsservices zoals XML Digital Signature (10) en XML Encryption (11), en zelfs arbitraire XSLT-transformaties (12) die vanuit en naar andere XML-indelingen converteren, liggen in de lijn der verwachting. Open XML staat al deze vormen van verwerking toe en legt geen beperkingen op.

5 STRUCTUUR VAN EEN OFFICE OPEN XML-DOCUMENT

Een van de belangrijkste oogmerken van deze whitepaper is de lezer inzicht te verschaffen in de high-level structuur van Open XML-bestanden. Om dit mogelijk te maken, wordt een gemiddelde hoeveelheid informatie verstrekt over de Open Packaging Conventions (OPC), terwijl aan de afzonderlijke opmaaktalen minder aandacht wordt besteed.

5.1 OPEN PACKAGING CONVENTIONS

De Open Packaging Conventions (OPC) bieden een manier om meerdere contenttypen (bijvoorbeeld XML, afbeeldingen en metagegevens) op te slaan in een container, zoals een ZIP-archief, zodat een document volledig kan worden weergegeven. De OPC beschrijven een logisch model voor het weergeven van containment en relaties.

In de aanbevolen werkwijze voor de OPC wordt gebruikgemaakt van het ZIP-archiveringsformaat. Met behulp van een ZIP-viewer kan de structuur van alle Open XML-bestanden worden bekeken. Het is zinvol om tijdens het lezen van deze beschrijving de inhoud van een klein Open XML-bestand met een dergelijke viewer te bekijken. Onder Windows hoeft u hiervoor alleen maar de extensie “zip” toe te voegen aan de bestandsnaam en te dubbelklikken.

Logisch gezien is elk Open XML-document een *OPC-package* (§5:8). Een package bestaat uit een niet-hiërarchische verzameling *parts* (§5:8.1). Ieder part heeft een niet-hoofdlettergevoelige *part name* die is opgebouwd uit een door slashes gescheiden reeks segmentnamen, zoals “/pres/slides/slide 1 .xml” (§5:8.1.1). Ook heeft ieder part een *content type* (§5:8.1.2). Fysiek gezien is het ZIP-archief één package, vormt ieder ZIP-onderdeel in het archief één part, en staan de padnamen in het ZIP-archief in direct verband met de part names.

In de ZIP-implementatie kunnen gebruikers via “/[Content_Types].xml” het contenttype van elk part in het package (§2:9.2.6) bepalen. De syntaxis en definitie van mediatypen is conform sectie 3.7 van RFC 2616 (13).

Packages en parts kunnen *expliciete relaties* (§1:9.2) bevatten met andere parts in het package, maar ook met externe bronnen. Aan elke expliciete relatie wordt een relatie-id toegekend, waardoor de content van een part naar de relatie kan verwijzen. Ook wordt een type toegekend, waardoor een applicatie kan beslissen hoe de relatie wordt verwerkt. Relatietypen worden benoemd met behulp van URI's, die niet-coördinerende partijen in staat stellen veilig en zonder conflicten nieuwe typen te maken.

De set expliciete relaties voor een bron-package of part wordt opgeslagen in een *relationships part*. Het relationships part voor het totale package heeft de naam “/_rels/.rels”, terwijl aan het relationships part voor het part “/a/b/c.xml” de naam “/a/b/_rels/c.xml.rels” wordt gegeven. De relationships parts (en in de ZIP-implementatie het part content type) zijn de enige speciaal benoemde parts in een package. Om een package te kunnen openen, moet de applicatie het package relationships part parsen en de relaties van het desbetreffende type volgen.

Alle andere parts in een Open XML-document bevatten Open XML, aangepaste XML of arbitraire content, zoals multimediaobjecten. Doordat een part aangepaste XML kan bevatten, zijn bijzonder krachtige methoden voor het insluiten van bedrijfsgegevens en metagegevens mogelijk.

5.2 WORDPROCESSINGML

Een WordprocessingML-document is opgebouwd uit een verzameling “stories” (§3:2.1). Elke story maakt deel uit van een van de volgende entiteiten: het hoofddocument (§3:2.2), de woordenlijst (§3:2.13), een subdocument (§3:2.18.2), een koptekst (§3:2.11.1), een voettekst (§3:2.11.2), een opmerking (§3:2.14.5), een frame, een tekstvak (§3:2.18.1), een voetnoot (§3:2.12.1) of een eindnoot (§3:2.12.2).

De enige verplichte story is het hoofddocument. Dit is het doel van het volgende type package relationship:

<http://schemas.openxmlformats.org/officeDocument/2006/relationships/officeDocument>

Een pad van de basis naar de top in de XML-boomstructuur bevat normaal gesproken de volgende XML-elementen (§3:2.2):

- `document` — het basiselement van het hoofddocument (§3:2.3).
- `body` — body (§3:2.7.1). De hoofdtekst kan meerdere alinea’s bevatten. Dit element kan ook section properties bevatten, gespecificeerd in een `sectPr`-element.
- `p` — paragraph (§3:2.4.1). De alinea kan een of meer runs bevatten. Dit element kan ook paragraph properties bevatten, gespecificeerd in een `pPr`-element. Het `pPr`-element kan op zijn beurt weer standaard run properties bevatten, gespecificeerd in een `rPr`-element (§3:2.4.4).
- `r` — run (§3:2.4.2). De tekststroom kan meerdere typen tekststroomcontent bevatten, hoofdzakelijk tekstbereik. Dit element kan ook run properties bevatten (`rPr`). In Open XML is de run een essentieel begrip. Een run is een aaneengesloten stuk tekst met identieke eigenschappen, dat geen aanvullende tekstopmaak bevat. Wanneer een zin bijvoorbeeld de woorden “dit zijn **drie** tekststromen” bevat, wordt deze zin beschreven door ten minste drie runs: “dit zijn”, “**drie**” en “tekststromen”. In dit opzicht wijkt Open XML sterk af van indelingen waarin het willekeurig nesten van eigenschappen is toegestaan, zoals HTML.
- `t` — text range (§3:2.4.3.1). Het tekstbereik bevat een willekeurige hoeveelheid tekst zonder opmaak, regeleinden, tabellen, grafieken of andere onderdelen die geen tekst zijn. De opmaak van de tekst wordt overerfd vanuit de run properties en de paragraph properties. Dit element maakt vaak gebruik van het attribuut `xml:space="preserve"`.

In dit gedeelte hebben we kennis gemaakt met de directe opmaak van tekst door het specificeren van paragraph properties en run properties. Directe opmaak komt aan het einde van een toepassingsvolgorde die daarnaast de volgende elementen omvat: teken, alinea, nummering, tabelstijlen en standaardinstellingen voor het document. Deze stijlen zijn ingedeeld in overervingshiërarchieën (§3:2.8.9).

Zie het gedeelte 5.6 “Minimaal WordprocessingML-document” voor een volledig overzicht van een WordprocessingML-document.

5.3 PRESENTATIONML

Een PresentationML-document wordt beschreven door een presentation part. Het presentation part is het doel van het volgende type package relationship:

<http://schemas.openxmlformats.org/officeDocument/2006/relationships/officeDocument>

Het presentation part verwijst naar de volgende basiselementen (§3:4.2), weergegeven van de top naar de basis in de standaardhiërarchie:

- Dia masters, notitiemodellen en hand-outmodellen (§3:4.2.2) - overerven eigenschappen uit het presentation part.
- Dia lay-outs (§3:4.2.5) - overerven eigenschappen van dia master.
- Dia's (§3:4.2.3) en notitiepagina's (§3:4.2.4) - overerven eigenschappen van respectievelijk dia lay-outs en notitiemodellen.

Iedere master, lay-out en dia wordt opgeslagen in het eigen part. De naam van elk part wordt gespecificeerd in het relationship part van het presentation part. Met uitzondering van het presentation part, hebben alle zes parts in hoofdzaak dezelfde structuur. Een pad van de basis naar de top in de XML-boomstructuur bevat normaal gesproken de volgende XML-elementen (§3:2.2):

- `sld`, `sldLayout`, `sldMaster`, `notes`, `notesMaster` of `handoutMaster` — het basiselement.
- `csld` — slide (§4:4.4.1.15). Dit element kan DrawingML-elementen bevatten (zie de twee opsommingen hierna) en andere structuurelementen (zie hierna).
- `spTree` — shape tree (§4:4.4.1.42). De shape-boom kan zowel group shape properties in een `grpSpPr`-element (§4:4.4.1.20) bevatten als niet-visuele group shape properties in een `nvGrpSpPr`-element (§4:4.4.1.28). Zowel dit knooppunt als de afgeleide elementen zijn DrawingML-elementen. Een ander DrawingML-element dat een centrale rol speelt in PresentationML:
- `sp` — shape (§4:4.4.1.40). De shape kan zowel shape properties in een `grpSpPr` element (§4:4.4.1.20) bevatten als niet-visuele shape properties in een `nvGrpSpPr` element (§4:4.4.1.28).

Een `csld` kan naast de shape-content uit DrawingML ook andere structuurelementen bevatten, afhankelijk van het basiselement waarin het zich bevindt. Zie het overzicht in de volgende tabel:

	Dia	Dia lay-out	Dia master	Hand-outmodel	Notitiemodel	Notitie pagina
Gemeenschappelijke gegevens	X	X	X	X	X	X
Overgang	X	X	X			
Timing	X	X	X			
Kop- en voetteksten		X	X	X	X	
Overeenkomende naam		X				
Type lay-out		X				
Behouden		X	X			
Overzicht lay-out			X			
Tekststijl			X			

Wanneer lagere objecten in de standaardhiërarchie (dia master, dia lay-out en dia) de eigenschappen specificeren, worden de door hogere objecten in de hiërarchie toegekende eigenschappen genegeerd. Als bijvoorbeeld geen overgang is gespecificeerd voor een dia, wordt de eigenschap gezocht in de dia lay-out. Als de eigenschap ook daar niet is gespecificeerd, wordt de eigenschap gezocht in de dia master.

5.4 SPREADSHEETML

Een SpreadsheetML-document wordt in het hoogste niveau beschreven door een workbook part. Het workbook part is het doel van het volgende type package relationship:

<http://schemas.openxmlformats.org/officeDocument/2006/relationships/officeDocument>

In het workbook part wordt informatie opgeslagen over het werkboek en de werkboekstructuur. Voorbeelden hiervan zijn de bestandsversie, de applicatie waarin het werkboek is gemaakt en een te wijzigen wachtwoord. Logisch gezien bevat het werkboek een of meer bladen (§3:3.2). Fysiek gezien wordt elk blad opgeslagen in het eigen part, terwijl de verwijzingen op de gebruikelijke wijze vanuit het workbook verlopen. Ieder blad kan een werkblad, grafiekblad of dialoogblad zijn. Omdat het werkblad het meest wordt gebruikt, wordt alleen dit type hier besproken. In een werkbladobject bevat een pad van de basis naar de top in de XML-boomstructuur normaal gesproken de volgende XML-elementen:

- `worksheet` — het basiselement in een werkblad (§3:3.2).
- `sheet Data` — de celtabel, beschrijft elke niet-lege cel in het werkblad (§3:3.2.4).
- `row` — één rij cellen in de celtabel (§3:2.8).
- `c` — één cel (§3:3.2.9). Het attribuut “r” duidt aan de hand van coördinaten (A1, B2, enz.) de cellocatie aan. Daarnaast kan een cel zijn voorzien van een stijl-id (attribuut “s”) en een gegevenstype (attribuut “t”).
- `v` en `f` — de waarde (§3:3.2.9.1) en de optionele formule (§3:3.2.9.2) van de cel. Als een cel een formule bevat, is de waarde het resultaat van de laatste berekening.

Zowel de tekenreeksen als de formules worden opgeslagen in gedeelde tabellen (§3:3.3 en §3:3.2.9.2.1), zodat redundante opslag of het op onjuiste wijze laden en opslaan van gegevens wordt voorkomen.

5.5 ONDERSTEUNENDE OPMAAKTALLEN

De content van een Open XML-document kan ook met behulp van diverse ondersteunende opmaaktalen worden beschreven:

- DrawingML (§3:5) — geschikt om shapes en andere grafische objecten in een document te beschrijven.
- VML (§3:6) — deze indeling voor vectorafbeeldingen is in Open XML opgenomen om achterwaartse compatibiliteit mogelijk te maken, maar zal in de toekomst worden vervangen door DrawingML.
- Gedeelde opmaaktalen: Math (§3:7.1), Metadata (§3:7.2), Custom XML (§3:7.3) en Bibliography (§3:7.4).

5.6 MINIMAAL WORDPROCESSINGML-DOCUMENT

Dit gedeelte beschrijft een minimaal WordprocessingML-document dat uit drie parts bestaat.

Het part content type “/[Content_Types].xml” beschrijft de contenttypen van de overige twee vereisteparts:

```
<Types xmlns="http://schemas.openxmlformats.org/package/2006/content-types">
  <Default Extension="rels"
    ContentType="application/vnd.openxmlformats-package.relationships+xml"/>
  <Default Extension="xml"
    ContentType="application/vnd.openxmlformats-officedocument.wordprocessingml.document.main+xml"/>
</Types>
```

Het part package relationship “/_rels/.rels” beschrijft de relatie tussen het package en het part hoofddocument:

```
<Relationships xmlns="http://schemas.openxmlformats.org/package/2006/relationships">
  <Relationship Id="rId1"
    Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/officeDocument"
    Target="document.xml"/>
</Relationships>
```

Het part document , hier “/document.xml”, bevat de content van het document:

```
<w:document xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main">
  <w:body>
    <w:p>
      <w:r>
        <w:t>Hello, world.</w:t>
      </w:r>
    </w:p>
  </w:body>
</w:document >
```

De specificatie voorziet in minimale documenten en aanvullende informatie voor WordprocessingML (§1:11.2), PresentationML (§1:13.2) en SpreadsheetML (§1:12.2).

6 SAMENVATTING

Open XML is het resultaat van een gemeenschappelijke inspanning van vertegenwoordigers uit een groot aantal bedrijfssectoren en publieke instellingen met uiteenlopende achtergronden en bedrijfsbelangen. Open XML komt niet alleen tegemoet aan de volledige functieset die voor bestaande documenten werd gebruikt, maar ook aan de internationalisatiebehoefte van alle belangrijke taalgroepen in elk deel van de wereld. Het standaardisatiewerk door Ecma TC45 (1) en de bijdragen in de vorm van openbaar commentaar hebben voor Open XML een grote mate van interoperabiliteit en platformafhankelijkheid mogelijk gemaakt. Daarnaast is de documentatie zeer compleet (dankzij uitgebreid referentiemateriaal) en voor iedereen toegankelijk (dankzij de niet-normatieve beschrijvingen). Bovendien bevat Open XML voldoende informatie voor producten die op gebruikers met fysieke beperkingen zijn afgestemd, zodat ook hier het verwerken van documenten op de juiste wijze kan plaatsvinden. Open XML-implementaties kunnen zowel bijzonder klein zijn en alleen gerichte functionaliteit bieden als op de complete functieset worden afgestemd. Dankzij de uitbreidingsmethoden die in de indeling zijn ingebouwd, is er voldoende ruimte voor innovaties.

De standaardisatie van de indelingsspecificatie en het onderhouden van de standaard op de lange termijn zorgen ervoor dat meerdere partijen kunnen vertrouwen op blijvende stabiliteit, in de wetenschap dat het proces voor open standaarden evenwichtige verdere ontwikkelingen waarborgt. Er bestaat grote behoefte aan een open standaard voor documentindelingen die het mogelijk maakt niet alleen de miljarden, in oude binaire indelingen gemaakte documenten te behouden, maar ook de miljarden documenten die daar jaarlijks bijkomen. De technologische vooruitgang op het gebied van hardware, netwerken en een op standaarden gebaseerde software-infrastructuur maakt dit mogelijk. De explosief groeiende spreiding van de marktvaag en de grote investeringen in bedrijfskritische bedrijfssystemen maken dit noodzakelijk.

1. **Ecma international:** overzicht van Office Open XML. *Ecma International*.
[Website] <http://www.ecma-international.org/memento/TC45.htm>
2. **W3C:** XML-schema. *World Wide Web Consortium*.
[Website] <http://www.w3.org/XML/Schema>.
3. **ISO:** ISO/IEC 19757-2:2003. *Internationale organisatie voor standaardisatie*.
[Website]
<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=37605&ICS1=35&ICS2=240&ICS3=30>
4. **ISO:** ISO/IEC 19757-4:2006. *Internationale organisatie voor standaardisatie*.
[Website]
<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=38615&ICS1=35&ICS2=240&ICS3=30>
5. **W3C:** Extensible Markup Language (XML) 1.0 (vierde editie). *World Wide Web Consortium*.
[Website 2006] <http://www.w3.org/TR/2006/REC-xml-20060816/>
6. **W3C:** Naamruimten in XML 1.0 (tweede editie). *World Wide Web Consortium*.
[Website 2006] <http://www.w3.org/TR/2006/REC-xml-names-20060816/>
7. **XBRL International:** XBRL-specificaties. *Extensible Business Reporting Language*.
[Website] <http://www.xbrl.org/Specifications/>
8. **Health Level Seven:** goedgekeurde HL7/ANSI-standaarden. *Health Level Seven*.
[Website] http://www.hl7.org/about/directories.cfm?framepage=/documentcenter/public/faq/ansi_approved.htm
9. **W3C:** Architectuur van W3C-webservices. *World Wide Web Consortium*.
[Website 2002] <http://www.w3.org/2002/ws/>
10. **W3C:** W3C XML-handtekening. *World Wide Web Consortium*.
[Website] <http://www.w3.org/Signature/>
11. **W3C:** W3C XML-encryptie. *World Wide Web Consortium*.
[Website 2001] <http://www.w3.org/Encryption/2001/>
12. **W3C:** XSL en XSLT. *World Wide Web Consortium*.
[Website] <http://www.w3.org/Style/XSL/>
13. **W3C:** Hypertext Transfer Protocol - HTTP/1.1. *World Wide Web Consortium*.
[Website] <http://www.w3.org/Protocols/rfc2616/rfc2616.html>