

Digitale Duurzaamheid: Openheid van Informatie

Frank Schoep

17-08-2005

Samenvatting

Om te voorkomen dat digitaal opgeslagen informatie in de toekomst ontoegankelijk wordt, is het belangrijk om nu al gebruik te maken van documentformaten die openlijk beschreven staan. Voor instanties en bedrijven waar de opslag en beschikbaarheid van grote hoeveelheden tekstdocumenten en spreadsheets een cruciale rol speelt is, het noodzakelijk om deze enorme schat aan informatie vast te leggen in een gestandaardiseerd formaat zonder restricties.

1 Introductie

1.1 Korte geschiedenis

Het gebruik van een tekstverwerker voor het schrijven van documenten heeft in de laatste tientallen jaren een grote vlucht genomen, vooral in de huidige zakelijke wereld is het gebruik van computers voor het aanmaken en verwerken van informatie niet meer weg te denken.

Vanaf de eerste versies van tekstverwerkingprogramma's als *WordPerfect* en *Microsoft Word* begin jaren tachtig [1] is de software voor tekstverwerking uitgegroeid tot complete *Office-suites* zoals die momenteel bestaan in de vorm van *Microsoft Office*, *Apple iWork* en *OpenOffice.org*.

In een relatief korte periode van twintig jaar zijn er grote veranderingen doorgevoerd in de manier waarop interactie met computers plaatsvindt. Het doorbreken van grafische gebruikersinterfaces, initieel toegankelijk gemaakt door Xerox PARC en Apple [2] leidde ertoe dat *Office-suites* makkelijker in gebruik werden en op een visuele manier de opmaak van digitale documenten ondersteunen.

1.2 Doorbraak

In de jaren daarna zorgen dalende kosten voor hardware en software ervoor dat tekstverwerking niet alleen gereserveerd is voor bedrijven, maar ook betaalbaar wordt voor particulieren. In de jaren negentig zorgen de besturingsystemen *Microsoft Windows* en *Apple Mac OS* voor een massale introductie van computers in de gezinssfeer.

Het gebruik van de thuiscomputer voor het aanmaken, lezen en bewerken van documenten wordt gemeengoed, alhoewel opleidingsniveau en training inzake het gebruik en toepassen van computers nog een hindernis vormt voor potentiële computergebruikers. Het belang van informatica en het gebruik van informatiesystemen wordt onderkend [4] en een eerste stap in het nationaal onderwijs wordt gezet door het toevoegen van het vak Informatica voor scholieren in de *Basisvorming* [10] en het opnemen van ICT-gebruik in andere vakken [9].

Met het invoeren van het leersysteem *Tweede Fase* in 1998 [3] wordt in het voortgezet onderwijs uitvoerig aandacht besteed aan het ondersteunend gebruik van computers. De basisconcepten van tekstverwerking en het gebruik van werkbladen worden in de tweede helft van de opleiding prominent belicht. Op deze manier worden scholieren voorbereid op de elektronische informatieverwerking waar ze later mee te maken krijgen.

1.3 Vloedgolf aan informatie

Nederland staat niet alleen in deze stormachtige ontwikkelingen omtrent de invoering en opleiding van ICT-gebruik onder inwoners. Internationale en Europese varianten van het zogenaamde *computerrijbewijs* [8] zorgen ervoor dat inwoners kunnen aantonen bezit te hebben van basiskennis op ICT-gebied.

Het aantal informatieverwerkende mensen neemt

zoals het er nu naar uitziet in de komende decennia alleen maar toe, waardoor er steeds meer en meer documenten geproduceerd, gewijzigd en gearhiveerd worden. Dit is een probleem waar vooral overheden en bedrijven mee te maken zullen krijgen. Er komt een vloedgolf aan documenten in uiteenlopende formaten die opgeslagen moet worden, maar ook na tientallen jaren nog steeds leesbaar moet zijn.

2 Open documentformaten

2.1 Van data naar document

De huidige markt voor *Office-suites* bestaat uit een groot aantal concurrerende pakketten die stuk voor stuk hun eigen interne documentformaat hanteren. De term *documentformaat* betekent hier niet de opmaak of inhoud van een tekst of werkblad, maar de *binaire representatie* ervan die op een elektronisch medium wordt opgeslagen.

Er is een vertaalslag nodig die van de opgeslagen binaire data een grafische voorstelling maakt van het document, zodat het op het scherm of op papier getoond kan worden. Zonder een vertaalslag is het niet mogelijk om de volledige inhoud van een document te reconstrueren. Dit heeft grote gevolgen voor de toegankelijkheid van opgeslagen informatie, want wie beheert de informatie die nodig is voor de vertaling van binaire data naar ons document?

Op grond van deze vraag is een duidelijk onderscheid te maken tussen zogenaamde *proprietary*-gesloten- en *open* documentformaten. Dit onderscheid is gebaseerd op de aanwezigheid van een kostenloze, verplichtingsvrije en met name volledige specificatie van het formaat. *Open* formaten zijn op deze wijze gedocumenteerd, *gesloten* formaten beschikken niet over een open specificatie.

2.2 Wie heeft de sleutel?

De specificatie van een documentformaat is de 'sleutel' voor de vertaling van data naar document. Bij *open* documentformaten is deze sleutel openbaar en publiekelijk beschikbaar, iedereen kan en mag een programma maken dat de ruwe data leest, vertaalt en verwerkt - kortweg '*interpreteren*' genoemd.

Door de beschikbaarheid van de specificatie ontstaat de interessante eigenschap dat documenten in een *open* formaat altijd volledig en correct te interpreteren zijn, zelfs als er geen programma's zijn die het formaat ondersteunen. In dat geval is het namelijk mogelijk om een programma te (laten) onwikkelen dat het documentformaat interpreteert en *converteert* naar een ander formaat. Op deze manier zijn documenten altijd toegankelijk te maken middels een omzetting zolang van beide formaten een specificatie voorhanden is.

Als we vanuit dit oogpunt kijken naar de *gesloten* documentformaten, dan ziet de situatie er voor de eindgebruiker veel minder rooskleurig uit. De sleutel voor de vertaalslag van data naar document is hier namelijk niet toegankelijk voor de gebruiker, maar is alleen bekend bij de producent van de software die gebruikt wordt.

Dit betekent dat de toegang tot informatie opgeslagen in documenten beheerst wordt door de maker van het programma in plaats van de auteur(s) van het document. Deze twijfelachtige constructie biedt vanzelfsprekend vele nadelen aan de gebruikers van *gesloten* documentformaten.

2.3 Vendor lock-in

Het feit dat gebruikers geen conversie kunnen toepassen van het *gesloten* formaat naar een ander formaat, omdat de specificatie ontbreekt, heeft tot gevolg dat men afhankelijk wordt van een enkele softwareleverancier. Men zit ingesloten in het gebruik van een enkel product, dit effect heet '*vendor lock-in*'.

Vendor lock-in manifesteert zich in de onmogelijkheid om de eigen informatie te lezen als ondersteuning van de producent wegvalt. Daarnaast is een gebruiker verplicht om mee te gaan in een eventuele één- of tweejaarlijkse upgradecyclus van het geleverde product, waardoor men kosten moet afdragen om de *eigen* informatie te kunnen lezen en bewerken.

Een gerelateerd probleem is dat documenten alleen geïnterpreteerd kunnen worden op een door de producent ondersteund besturingssysteem en hardwareplatform. Dit betekent bijvoorbeeld dat het geautomatiseerd verwerken van documenten alleen onder het *Windows* besturingssysteem mogelijk is, waardoor een bedrijf dit systeem in de toekomst moet blijven gebruiken terwijl een overstap naar

Linux een kostenbesparing en efficiencywinst zou opleveren.

2.4 Wel voordelen, geen nadelen

Kijkend naar de genoemde voor- en nadelen van *open* en *gesloten* documentformaten valt op dat gestandaardiseerde *open* documentformaten alleen voordelen hebben, terwijl aan *gesloten* formaten nadelen kleven. De voordelen die een *gesloten* documentformaat biedt zijn alleen merkbaar voor de softwareproducent, niet voor de auteur van het document.

3 Het Valoris onderzoek

3.1 Europese Unie

Het belang van een gestandaardiseerd *open* documentformaat is duidelijk: voor opslag en archivering van documenten is het belangrijk de garantie te hebben een document voor 100% te kunnen reproduceren zoals het ten tijde van archivering bedoeld is. Deze garantie kan alleen geboden worden door het hanteren van een gestandaardiseerd en erkend *open* documentformaat.

De *Europese Unie* heeft in 2003 het onderzoeksburo *Valoris* een onderzoek laten verrichten naar een geschikt documentformaat voor elektronische archivering van documenten met als doel het vaststellen van een formaat dat *open* is en een goede basis vormt voor de toekomst. De uitslag van dit onderzoek dient als basis voor aanbevelingen inzake het selecteren van een documentformaat dat gebruikt zal worden voor internationale communicatie en archivering.

Het 78 pagina's tellende *Valoris* verslag is on-line beschikbaar [11]. Het belang van een *open* documentformaat wordt dus Europees onderkend en het verslag gaat zeer uitgebreid in op de eigenschappen en kenmerken van bestaande documentformaten, het vormt een goede basis voor een uiteindelijke keuze voor een geschikt documentformaat.

3.2 Selectiecriteria

Valoris hanteerde in het verslag criteria in een drielid categorieën die bepalend zijn voor de geschiktheid van een documentformaat. Er werden meer

dan twintig bestaande documentformaten beoordeeld op:

- Neutraliteit

Openheid Is er een publieke, *royalty-* en *licentievrije* specificatie beschikbaar?

Niet-binair Binaire formaten kunnen een vorm van platformafhankelijkheid introduceren, het formaat moet platformneutraal, dus niet-binair zijn.

Platformonafhankelijk Het documentformaat moet niet gebonden zijn aan een enkel besturingssysteem of hardwareplatform.

- Technisch vernuft

Weergave In hoeverre biedt het documentformaat controle over de visuele presentatie van het document?

Bewerkbaarheid Staat het documentformaat toe dat een document gewijzigd kan worden? Het *PDF*-formaat is bijvoorbeeld een alleen-lezen formaat.

Functionaliteit Het documentformaat moet ondersteuning bieden voor internationale tekensets zoals *Unicode*, bi-directionele tekst en het integreren van scripts of *macro's*.

Toekomst Het formaat moet uitbreidbaar zijn, zodat eventueel latere eisen als digitale handtekeningen, toegangsrechten en versiebeheer toegevoegd kunnen worden.

- Brede acceptatie

Brede acceptatie Het documentformaat moet worden ondersteund door softwareleveranciers, zodat eindgebruikers ook daadwerkelijk middelen hebben om documenten te bewerken en op te slaan.

3.3 Uitslag

Geen van de onderzochte documentformaten scoorde perfect, maar er zijn twee documentformaten die als kandidaat in aanmerking komen voor acceptatie, beide met slechts een klein aantal tekortkomingen. De twee topscoorders zijn ten eerste *OpenDocument XML*, onder andere gebruikt in het pakket *OpenOffice.org* en ten tweede *Microsoft Office*

XML dat alleen ondersteund wordt in het pakket *Microsoft Office 2003*.

3.3.1 OpenDocument XML

Het *OpenDocument XML* documentformaat was duidelijk één van de best geschikte formaten, het voldoet goed aan de eisen van een openlijke specificatie, er is in de vorm van het gratis, *open-source* pakket *OpenOffice.org* zelfs platformafhankelijke programmacode beschikbaar waarmee foutloos documenten gelezen en geschreven kunnen worden.

Het enige minpunt dat Valoris aanmerkt op het formaat is dat een aantal toekomstige wensen nog niet mogelijk zijn in de versie ten tijde van publicatie van het rapport. Dit enige kritiekpunt is inmiddels weggewerkt en *OpenDocument XML* vormt de solide basis voor het *OASIS Open Document Format* [6], hier kom ik later uitgebreid op terug.

3.3.2 Microsoft Office XML

Dit documentformaat wordt ook als kandidaat gezien, maar schiet tekort op een aantal punten:

- Documentweergave: de weergave van een document kan niet volledig worden gecontroleerd, er kunnen verschillen in weergave optreden afhankelijk van de softwareomgeving waarin men een document bekijkt.
- Platformafhankelijkheid: in dit documentformaat zijn *XML*-elementen opgenomen die afhankelijk zijn van onderdelen in het *Windows* besturingssysteem. Hierdoor is het onmogelijk om dit type documenten op een ander besturingssysteem te interpreteren.
- Het documentformaat kan niet worden gebruikt voor alle *Office* toepassingen, het is bijvoorbeeld niet mogelijk om een *Powerpoint* presentatie op te slaan in dit documentformaat.

3.4 Aanbeveling

De conclusie die uit het Valoris rapport kan worden getrokken is dat het *OpenDocument XML* formaat de beste keuze is. De hoofdredenen hiervoor kunnen samengevat worden als:

- *OpenDocument XML* biedt meer openheid: er zijn geen softwarepatent- en wetsproblemen en er is ondersteuning van het OASIS comité [5].
- Hergebruik van bestaande standaarden, waar mogelijk.
- Betere controle over de weergave en opmaak van documenten.
- *OpenDocument XML* kan door bestaande technieken als *XSLT* makkelijk geconverteerd worden naar een ander formaat.

De werkgroep van de Europese Unie gaf, na overleg met betrokken partijen, een aanbeveling die sterk neigt naar het gebruik van het *OpenDocument XML* formaat. De tekortkoming die Valoris constateerde werd opgevat als een uitdaging om het *OpenDocument XML* formaat uit te breiden zodat nieuwe mogelijkheden in de toekomst zonder problemen toegevoegd kunnen worden. Deze uitdaging werd goed onthaald en in een OASIS werkgroep werd aan het werk begonnen.

4 OASIS OpenDocument v1.0

4.1 Er is *nu* een open standaard

De planning en implementatie van de uitbreiding op het *OpenDocument XML* formaat zijn ondertussen voltooid en op 1 mei 2005 is het gereviseerde documentformaat officieel goedgekeurd en als standaard geaccepteerd door het OASIS comité onder de noemer 'OpenDocument Format for Office Applications (OpenDocument) v1.0', kortweg *ODF*.

De gezamenlijke inspanning van tientallen experts en adviseurs is nu beschikbaar als een complete en publiekelijk toegankelijke specificatie [7].

Het bestaan van een standaard betekent natuurlijk nog niet dat deze ook daadwerkelijk gebruikt zal worden, maar in het geval van het *Open Document Format* is de acceptatie nu al een feit. Twee grote *open-source* Office pakketten ondersteunen het documentformaat al in de aankomende versies: *OpenOffice.org 2.0* en *KOffice 1.4*.

4.2 Microsoft en OASIS ODF

Microsoft, met een bijna alleenheerschappij [11] in de Office markt, heeft nog geen standpunt ingenomen

men over het al dan niet ondersteunen van het OASIS documentformaat, maar probeert instanties en overheden te overtuigen van hun eigen *proprietary* formaat, dat vanuit het oogpunt van *openheid* en duurzaamheid inferieur is aan *OpenDocument XML*.

Door het gebrek aan een volledige, vrije en open specificatie van het *Microsoft Office XML* formaat en een gebrek aan OASIS documentformaat ondersteuning is er nauwelijks tot geen *interoperabiliteit* mogelijk tussen *Microsoft Office* en andere pakketten of besturingssystemen. Een ernstige waarschuwing voor *vendor lock-in* is hier dan ook op zijn plaats.

4.3 Conclusie

Dankzij het denkwerk dat de OASIS werkgroep, grotendeels bestaand uit vrijwilligers, heeft gedaan is er nu een robuuste fundering voor de toekomst van elektronische documentverwerking. Alleen met de acceptatie van een echt open documentformaat zoals *OpenDocument XML* door bedrijven, instanties en overheden zal er uiteindelijk zekerheid komen over de toekomst van informatie die opgeslagen is in elektronische documenten.

5 Dankwoord

Ik wil graag de volgende reviewers bedanken voor de beoordeling van het artikel en het aandragen van suggesties ter verbetering:

- Dalen, Maiko van
- Hubregtse, Leendert
- Pol, Hans van de
- Verwerft, Johan

Referenties

- [1] Answers.com. WordPerfect: Information. <http://www.answers.com/topic/wordperfect>.
- [2] Ars Technica. A History of the GUI. <http://arstechnica.com/articles/paedia/gui.ars>.
- [3] Ministerie van Onderwijs, Cultuur en Wetenschap. Dossier Tweede Fase. <http://www.minocw.nl/tweedefase/>.
- [4] Ministerie van Onderwijs, Cultuur en Wetenschap. Geen toekomst zonder informatica: toekomstverkenning. <http://www.minocw.nl/wetenschap/ocvsv/>.
- [5] OASIS Committee. OASIS. <http://www.oasis-open.org/>.
- [6] OASIS Committee. OASIS Open Document Format for Office Applications (OpenDocument) TC. http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office.
- [7] OASIS Committee. Open Document Format for Office Applications (OpenDocument) v1.0. <http://www.oasis-open.org/committees/download.php/12572/OpenDocument-v1.0-os.pdf>.
- [8] Het Europees Computer Rijbewijs. ECDL Nederland. <http://www.ecdl.nl/>.
- [9] Stichting Leerplanontwikkeling SLO. Ict kerndoelen basisvorming: Ict in andere vakken. <http://ict.slo.nl/service/ictanvak.doc>.
- [10] Stichting Leerplanontwikkeling SLO. Voorstel voor invulling funderende module informatica. <http://ict.slo.nl/service/fundmodu.doc>.
- [11] Valoris. Comparative assessment of Open Documents Formats Market Overview. <http://europa.eu.int/idabc/servlets/Doc?id=17982>.