



XML en PDF: Twee standaarden

Het Portable Document Format (PDF) en de eXtensible Markup Language (XML) zijn standaarden die genoemd worden in Regeling geordende en toegankelijke staat archiefbescheiden. Reden genoeg om kort op beide in te gaan en overeenkomsten en verschillen uit te leggen.

Concepten als structuur, inhoud en presentatie spelen bij beide standaarden een belangrijke rol. Het begrip structuur is te vertalen naar het ons bekende archivistische begrip redactionele vorm, de manier waarop de formele kenmerken van een document zijn gerangschikt. De inhoud hoeft geen uitleg. De presentatie is te vertalen naar het archivistische begrip uiterlijke vorm, de verschijningsvorm van een archiefstuk. Het begrip bestandsformaat slaat op de manier waarop gegevens zijn opgeslagen op een gegevensdrager. Dit slaat bijvoorbeeld ook op de gebruikte tekenset. De meeste bestandsformaten zijn conform een bepaalde standaard. Soms defacto door het vele gebruik, soms omdat instituten als de International Organization for Standardization (ISO) er achter staan.

PDF

PDF is een bestandsformaat, waarvan de rechten liggen bij de firma ADOBE. Kenmerk van dit formaat is dat structuur, inhoud en presentatie van een document onlosmakelijk als een geheel met elkaar zijn verbonden. Als middel voor publicatie en uitwisseling is PDF zeer geschikt. Het grote voordeel is dat het bestandsformaat wijd verbreid is. Een van de nadelen is dat het geen open standaard is. De gebruikers zijn voor de ontwikkeling van PDF afhankelijk van de eigenaar. Voor het maken van PDF-bestanden zijn commerciële producten nodig. In de wereld van de 'open source software' zijn wel zogenaamde PDF-generators beschikbaar waarmee documenten in elk willekeurig formaat op relatief eenvoudige wijze zijn 'af te drukken' naar PDF. Voor het reproduceren is de gratis 'reader', Acrobat Reader ook een product van ADOBE beschikbaar. Hoewel documenten door het gebruik van PDF gefixeerd lijken zijn ze toch relatief eenvoudig te muteren. Beveiliging is een zwak punt.

XML

XML is geen bestandsformaat maar een open standaard, vastgesteld door het W3 consortium. XML bestaat uit een aantal talen waarmee documenten kunnen worden beschreven en waarmee ze kunnen worden gepresenteerd. XML wordt gebruikt voor gestandaardiseerde gegevensuitwisseling en is ook zeer geschikt voor duurzame bewaring. Een XML document bestaat grofweg uit drie bestanden te weten een XML-schema (voorheen een Document Type Definition (DTD)), het eigenlijke document met de inhoud en een XSL-document. In XML-schema wordt de structuur van het document vastgelegd. Hierin krijgen de verschillende elementen een betekenisvolle naam in de vorm van tags. Het schema bevat de afspraken over de elementen die in een bepaald document thuishoren. Afspraken kunnen gaan over de naamgeving, het al dan niet verplicht stellen van een element, het aangeven of bepaalde elementen meer dan een keer in een document kunnen voorkomen enzovoorts. In een schema kan ook aangegeven worden of een element extra kenmerken krijgt die bijvoorbeeld te gebruiken zijn bij zoekacties.

Information is a [conceptual labyrinth](#) and at the same time a [hyperobject](#)

Een organisatie kan XML-schema's heel goed gebruiken voor het vastleggen van de vormvereisten voor bijvoorbeeld besluiten, brieven en vergunningen. Binnen ons eigen vakgebied is de Encoded Archival Description (EAD) een voorbeeld van een DTD voor toegangen op archieven. De inhoud van het document komt in een bestand waarbij de tags, de structuurelementen uit het XML-schema, de inhoudselementen omsluiten. Voor de presentatie, de verschijningsvorm is XSL (Extensible Stylesheet Language) beschikbaar. Deze taal bevat de instructies voor de opmaak. De opmaakinstructies betreffen zaken als lettertype, lettergewicht, kleur en plaatsing van de elementen. Met XSL kan de inhoud uitgevoerd worden naar het beeldscherm en naar bestanden in formaten als PDF, HTML, SGML. Uitvoer naar databases of rechtstreeks afdrukken op papier behoren ook tot de mogelijkheden. Een XSL document is zeer geschikt om de huisstijl van een organisatie in op te nemen.

Het bestandsformaat van alle soorten XML-bestanden is de ISO/IEC 10646 standaard. Deze standaard is een tekenset die niet afhankelijk is van één of meer software leveranciers. Omdat het bestandsformaat behoorlijk stabiel is, zijn garanties voor duurzaamheid aanwezig. Verder is het aardige van XML dat veel instelbaar is, mits het maar voldoet aan de algemene XML-syntax. Dit kan ook direct een behoorlijk nadeel zijn. Om XML doeltreffend te kunnen gebruiken kan voor toepassing het beste een soort traject van systeemontwikkeling doorlopen worden, waarbij onderdelen als voorbereiding, analyse voor mogelijk gebruik, implementatie en onderhoud goed geregeld moeten zijn. Bij voorkeur moet de dagelijkse gebruiker niet merken dat XML wordt toegepast. De gebruiker moet bij wijze van spreken nog steeds met de hem of haar vertrouwde tekstverwerker zijn documenten kunnen (op)maken. Opslag en/of eventueel (latere) migratie kan op de achtergrond worden uitgevoerd.

Het gebrek aan ingebouwde beveiliging is een ander nadeel, maar dat viel ook op bij PDF. De integriteit en de authenticiteit van een document zullen bij beide standaarden geregeld moeten worden via een stelsel van autorisaties, controles en eventueel encryptie.

Een eenvoudig DTD

```
<!--      DTD voor eenvoudige brieven
          datum : 10 september 2002
          Rienk Jonker
```

een eenvoudige brief kan bestaan uit een adressering,
datum en plaatsgegevens, de tekst en het einde met afsluitende informatie

```
-->
<!ELEMENT brief (adressering?, datumplaats?, detekst+, einde?)>
<!ELEMENT adressering (#PCDATA)>
<!ELEMENT datumplaats (#PCDATA)>
<!ELEMENT detekst (#PCDATA)>
<!ELEMENT einde (#PCDATA) >
```

Een XML-document

De inhoud van een heel eenvoudige brief in XML met de tags:

```
<?xml version="1.0"?>
<!DOCTYPE brief SYSTEM "brief.dtd">
```

Information is a [conceptual labyrinth](#) and at the same time a [hyperobject](#)

```
<?xml-stylesheet href="brief.css" type="text/css"?>
<brief>
  <adressering>
    <naam>Redactie Archievenblad </naam>
    <adres>Cruquiusweg 31</adres>
    <postcode>1019 AT</postcode>
    <plaats>Amsterdam</plaats>
  </adressering>
  <datumplaats>
    <datum>9 september 2002</datum>
    <plaats>Winschoten</plaats>
  </datumplaats>
  <detekst>
    <aanhef>Geachte redactie</aanhef>
    <inhoud>Hierbij een kort artikel over PDF en XML</inhoud>
  </detekst>
  <einde>
    <afsluiting>Met vriendelijke groet</afsluiting>
    <ondertekening>Rienk Jonker </ondertekening>
  </einde>
</brief>
```

Zie ook:

- Op de site van het W3 consortium [XML in 10 punten \(7, eigenlijk...\)](#) [1]
- Op de site van eDAVID - Expertisecentrum: [F. Boudrez, <XML/> en digitaal archiveren, Antwerpen, 2002](#) [2]

Tags:

- [metataal](#) [3]
- [bestandsformaat](#) [4]
- [xml](#) [5]
- [Standaard](#) [6]
- [PDF](#) [7]
- [xml schema](#) [8]
- [dtd](#) [9]

Levenscyclus:

[3.3 Fysiek beheren](#) [10]

[3.5 Documenteren](#) [11]

[4.1 Voortbestaan](#) [12]

Informatiemodel:

[Informatieobject](#) [13]

Information is a [conceptual labyrinth](#) and at the same time a [hyperobject](#)

Grondslag:

[Archiefregeling \(AR\)](#) [14]

Sharte this / Add this:

Datum eerste publicatie:

vrijdag, 18 juni 2004 - 7:45pm

 [15]



Bron-URL:<http://labyrinth.rienkjonker.nl/xmlpdf>

Links

[1] <http://www.w3.org/XML/1999/XML-in-10-points.nl> [2]

http://www.edavid.be/davidproject/teksten/DAVIDbijdragen/XML_digitaalarchiveren.pdf [3]

<http://labyrinth.rienkjonker.nl/tags/metataal> [4] <http://labyrinth.rienkjonker.nl/tags/bestandsformaat> [5]

<http://labyrinth.rienkjonker.nl/tags/xml> [6] <http://labyrinth.rienkjonker.nl/tags/standaard-0> [7]

<http://labyrinth.rienkjonker.nl/tags/pdf> [8] <http://labyrinth.rienkjonker.nl/tags/xml-schema> [9]

<http://labyrinth.rienkjonker.nl/tags/dtd> [10] <http://labyrinth.rienkjonker.nl/levenscyclus/33-fysiek-beheren> [11]

<http://labyrinth.rienkjonker.nl/levenscyclus/35-documenteren> [12]

<http://labyrinth.rienkjonker.nl/levenscyclus/41-voortbestaan> [13]

<http://labyrinth.rienkjonker.nl/contextmodel/informatieobject-0> [14]

<http://labyrinth.rienkjonker.nl/grondslagen/archiefregeling-ar> [15]

<https://www.addtoany.com/share?url=http%3A%2F%2Flabyrinth.rienkjonker.nl%2Fxmlpdf&title=XML%20en%20PDF%3A%20Twee%20standaarden>

Information is a [conceptual labyrinth](#) and at the same time a [hyperobject](#)

[\[Home\]](#) [\[Nieuws via RSS\]](#) [\[Colofon\]](#) [\[Zoeken\]](#)