



Wim Hooghwinkel

Wat is DITA nu eigenlijk

DITA staat voor Darwin Information Typing Architecture en is een XML-toepassing, ontwikkeld voor het maken, beheren en publiceren van modulair georganiseerde informatie.

De DITA-specificatie wordt beheerd door het OASIS DITA Technical Committee en is nog steeds in ontwikkeling. De huidige standaard is versie 1.2 sinds eind 2008.

In het OASIS subcomitee zijn de diverse grote spelers uit de informatieverwerkende industrie vertegenwoordigd, dat zijn zowel documentatie-afdelingen van multinationals of uit de IT-industrie als softwareontwikkelaars en 'global information providers' (vertaal- en lokalisatiebureaus, leveranciers van CMS-systemen).

Dat het een open standaard is en dat deze gesteund wordt door de industrie is belangrijk voor bedrijven en organisaties die moeten investeren in technologie voor productinformatie. Omdat het een standaard is wordt continuïteit gewaarborgd en omdat het open is kan niet gebeuren dat één producent zich een monopolie toe-eigent. Dit stelt bedrijven in staat hun investeringen naar de toekomst te verantwoorden en flexibel te blijven in keuze voor leveranciers.

De ontwikkeling van DITA past in een lange geschiedenis van ideeën over het structureren van informatie en scheiding van vorm en inhoud door middel van markeringen (tagging). In de jaren 60 van de vorige eeuw werden de eerste stappen voor informatie-typing (information typing) gezet door Robert Horn met Information Mapping¹. In dezelfde tijd werd bij IBM onder leiding van Charles Goldfarb de mark-up taal GML ontwikkeld, wat later zou uitgroeien tot de ISO standaard SGML. Jarenlang was dit de standaard voor gestructureerde documentatie en is nog steeds in gebruik in de vorm van de SGML-applicatie DocBook. In 1981 werd voor het eerst, weer door een team van IBM, het idee naar buiten gebracht

¹ In the mid-'60's [Robert Horn](#) (winner of an ACM SIGDOC Lifetime Achievement Award for Documentation) developed [Information Mapping](#) techniques and founded the company by that name. Common "Information Types" were identified in dozens of standard document types like user manuals, policy and procedure manuals, annual reports, etc. Identifying standard information types is at the heart of DITA (Darwin Information Typing Architecture).

dat documentatie taak-georiënteerd zou moeten zijn om zo optimaal aan te sluiten bij de behoeften van de eindgebruiker van een (software) product. Een toepassing van deze ideeën was de computer-documentatie van de eerste Apple macintosh in 1984. De WYSIWYG weergave ontketende een revolutie in computergebruik en leidde tot de opkomst van Desk Top Publishing-pakketten als Pagemaker, FrameMaker en Interleaf, waarbij Interleaf en later FrameMaker ook het werken met gestructureerde documenten (SGML) ondersteunde. Diverse theorieën over informatietyping werden gelanceerd. De terugkerende informatietypen waren Learning, Using en Reference, later in DITA vertaald naar Concept, Task, Reference.

Minimalism werd geïntroduceerd door John M. Carroll in 1990². Minimalism is de uiterste vorm van taak-georiënteerd documentatie maken en gaat uit van kleine onderdelen informatie die leesbaar zijn in willekeurige volgorde. De nadruk komt te liggen op het 'lezen om te doen' in plaats van 'lezen om te weten'. Dit principe werd nog eens opnieuw gedefinieerd in 1995 tijdens een workshop van de Society of Technical Communication (STC). In Nederland wordt deze stroming vertegenwoordigd door Hans van der Meij³ die zich nog steeds bezig houdt met onderzoek naar minimalism en verdere uitdieping daarvan.

Rond 1995 volgde de introductie van XML door een werkgroep van het W3C. XML was bedoeld als een vereenvoudigde en meer flexibele vorm van SGML met als doel het markeren van teksten voor publicatie via het World Wide Web. De SGML-applicatie HTML voldeed niet meer aan de toenemende eisen. XML bleek al snel voor meer doeleinden geschikt en groeide uit tot een algemeen gebruikte standaard voor gegevensuitwisseling. DITA werd geboren in 1996 toen,

² The Nurnberg Funnel, 1990; bijgewerkte uitgave in 1995: Minimalism Beyond the Nurnberg Funnel.

³ <http://users.edte.utwente.nl/meij/>



alweer, een werkgroep van IBM zich boog over het maken van een helpstelsysteem voor software en daarbij gebruik maakte van een combinatie van XML, minimalism en datatypering. Een van de doelstellingen was een systeem, of architectuur, te bedenken dat kon voorzien in een oplossing voor nieuwe problemen die zich voordeden bij IBM. Dit had te maken met het toenemende aantal producten, dat in steeds kortere tijd ontwikkeld werd en vervolgens regelmatig vernieuwd. *Time to Market* speelt een belangrijke rol: ontwikkelingen gaan zo snel dat zo gauw een product klaar is het direct in de markt gezet moet worden (= verkocht), inclusief bijbehorende documentatie in alle talen. Het ontwikkelteam (met onder andere Don Day en Michael Priestley) besloot om te gaan zoeken naar een werkwijze die parallel kon lopen aan de software- en hardware ontwikkelingen bij het bedrijf. Er werd gekozen voor een pragmatische en taakgerichte modulaire opzet met als basisdomein de drie informatietypen concept, task en reference. De werkgroep introduceerde het idee van uitbreidbare informatietypering door middel van specialisatie van de basistopics. Later werd hier ook nog domeinspecialisatie aan toegevoegd.

Na het invoeren van de eerste versie van DITA binnen IBM in 2001 volgde een korte strijd met de auteursafdelingen, die gewend waren om zelfstandig producten (achteraf) te beschrijven en dat in boekvorm: van begin tot eind. De nieuwe werkwijze echter was gericht op het schrijven van losse componenten, parallel aan de productontwikkeling, zonder op voorhand te hoeven weten hoe of in welke vorm deze uiteindelijk zouden worden samengevoegd. Een nieuwe rol werd geïntroduceerd: de *informatie-architect*. De informatie-architect is degene die een 'plattegrond' maakt van alle benodigde componenten (informatie), reeds bestaande componenten aanmerkt voor hergebruikt, eventueel nieuwe definieert en de auteurs aanstuurt om de nog ontbrekende informatie-objecten te gaan maken. Toch bleek DITA een succes en om verdere ontwikkeling mogelijk te maken deed IBM de standaard in 2004 aan OASIS en zette de Open Toolkit in het Open Domain (SourceForge). Veel bedrijven zagen de (vooral zakelijke) voordelen en besloten de standaard in gebruik te nemen en te ondersteunen.

Wat maakt DITA nu zo bijzonder

Bij DITA komen alle karakteristieken die van belang zijn bij de productie van informatie bij elkaar - modulariteit, gestructureerd schrijven, informatietypering, scheiding van inhoud en presentatie, single-sourcing, minimalism, topics, taakgericht, hergebruik, conditionele verwerking, vertaalvriendelijk, meerdere uitvoerkanalen, component publishing, gebruiksvriendelijk, consistent, objectgeorieerd, overerving, specialisatie, eenvoudige XML.

Documentatie-afdelingen en informatieleveranciers zien zich gesteld voor een aantal uitdagingen die te maken hebben met globalisering en snelle ontwikkelingen van producten. Productdocumentatie moet multi-functioneel zijn, gericht op diverse doelgroepen en geschikt voor een brede product-range. Dezelfde informatie moet beschikbaar zijn in diverse formaten, geschikt voor meerdere kanalen (online, print, mobile) en in steeds meer talen. Tegelijkertijd moeten de productiekosten omlaag, dat betekent lagere vertaalkosten en minder DTP-kosten. De informatie moet bovendien herbruikbaar zijn, binnen de eigen productgroep maar ook daarbuiten en als het even kan geschikt zijn voor hergebruik door andere afdelingen van de organisatie met andere doelen. Tot slot verlangen eindgebruikers stipte en gerichte informatie, goed doorzoekbaar, op het juiste

moment beschikbaar en altijd bijgewerkt met de meest recente gegevens.

XML-applicatie

XML maakt het mogelijk om inhoud te scheiden van presentatie, onderscheid te maken tussen structuur en inhoud en om de verwerking los te koppelen van het maakproces. Auteurs kunnen zich concentreren op de inhoud, hoeven zich niet druk te maken om de presentatie. Dezelfde inhoud kan, geautomatiseerd, worden uitgevoerd naar diverse applicaties voor verschillende vormen van publicatie. Eventuele nieuwe vormen van uitvoer kunnen makkelijk worden toegevoegd zonder dat dat invloed heeft op de inhoud of het auteursproces: dezelfde data wordt gebruikt. XML leent



zich door zijn hiërarchische opbouw uitstekend voor beheer in een database. Wijzigingen kunnen zo eenvoudig worden doorgevoerd en deze zijn dan ook direct beschikbaar.

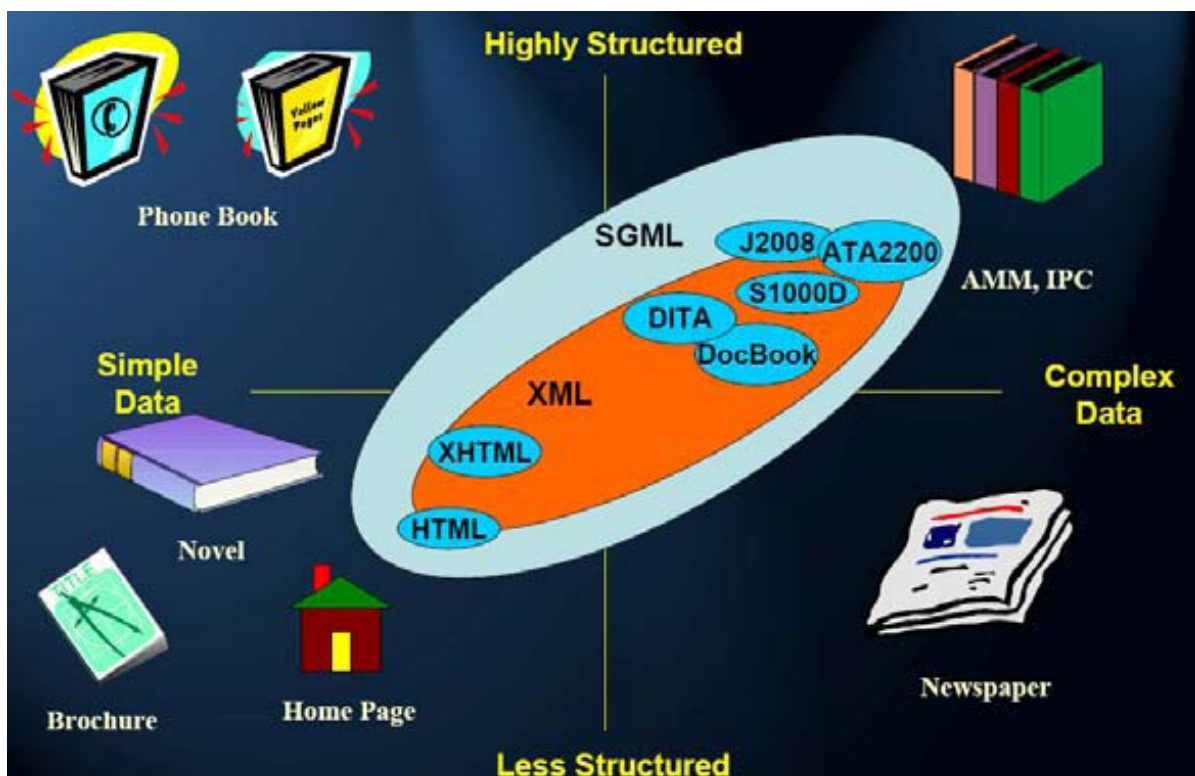
Wat is XML

XML is een mark-up taal, voortgekomen uit SGML. De mark-up, of tagging, is bedoeld als semantische markering, dat wil zeggen dat het wordt gebruikt om de inhoud van een document te beschrijven, niet de opmaak. Hierin onderscheidt XML zich van HTML, wat een SGML-toepassing is, gericht op het beschrijven van weergave in browsers, dus de opmaak.

Door de mark-up ontstaat een structuur van elementen. In XML hebben alle elementen verplicht een begin- en sluitmarkering (open and close tag), wat bijvoorbeeld in SGML en HTML

niet verplicht is. Elementen kunnen een of meer attributen hebben die meer informatie geven over de inhoud van dat element.

Een DTD of een Schema is een document dat beschrijft welke elementen, eventueel in welke volgorde of hiërarchie, beschikbaar zijn om te gebruiken. Voor diverse typen documentatie kunnen verschillende DTD's of Schema's worden gemaakt. Een XML-standaard is een vastgestelde set van elementen, vastgelegd in een standaard set DTD's of Schema's. Wanneer je informatie maakt aan de hand van zo'n standaard weet je dat de te gebruiken tools daarmee overweg kunnen. Enkele bekende standaarden zijn S1000D (vliegtuigindustrie), SCORM (e-learning), ATA2100, DocBook (technische documentatie) en DITA, maar in feite is HTML ook een standaard. Iedere standaard heeft zijn eigen mate van complexiteit en eigen toepassingsgebied.



Afbeelding: Bill Wheat, Delta Airlines © Suite Solutions



Scheiding inhoud, opmaak en structuur:

Formatting

Heading 1 + 20 pt, Centered

Arial, 14pt, bold

Heading 2

Normal + Arial, 14 pt, Bold, Left: 0", Hanging: 1.25"

Heading 2

Sample Task Card

EqType: A-123
Skill Code: 06
Area: 02
Date: Sept. 11, 2007
Keyword: VHF Transceiver

Description
Remove and replace the Nbr. 1 VHF Radio Transceiver

WARNING: Power must be removed before performing maintenance to this system.

Mech	Insp	
		1. Open circuit breaker CB-101. Verify power is removed to VHF-1 system.
		2. Remove and Replace VHF-1 transeiver, Part Number TR-12345-6
		3. Close circuit breaker CB-101.
		4. Perform operational test of system.

Parts Required
TR-12345-6

Circuit Breakers
CB-101

Structure

title

equiptype

skillcode

area

reldate

keyword

description

warning

step

cb

pnr

Afbeelding: Bill Wheat, Delta Airlines © Suite Solutions

```

<taskcard>
  <title>Sample Task Card</title>
  <equiptype>A-123</equiptype>
  <skillcode>06</skillcode>
  <area>02</area>
  <reldate>Sept. 11, 2007</reldate>
  <keyword>VHF Transceiver</keyword>
  <description>Remove and replace the Nbr. 1
    VHF Radio Transceiver.</description>
  <step signoff="m">
    <warning>Power must be removed before
      performing maintenance to this
      system.</warning>
    Open circuit breaker <cb>CB-101</cb>.
    Verify power is removed to VHF-1
  </step>
  <step signoff="m">Remove and Replace VHF-1
    transceiver, Part Number <pnr>TR-12345-
    6</pnr></step>
  <step signoff="m">Close circuit breaker
    <cb>CB-01</cb>.</step>
  <step signoff="mi">Perform operational test of
    system.</step>
</taskcard>
    
```

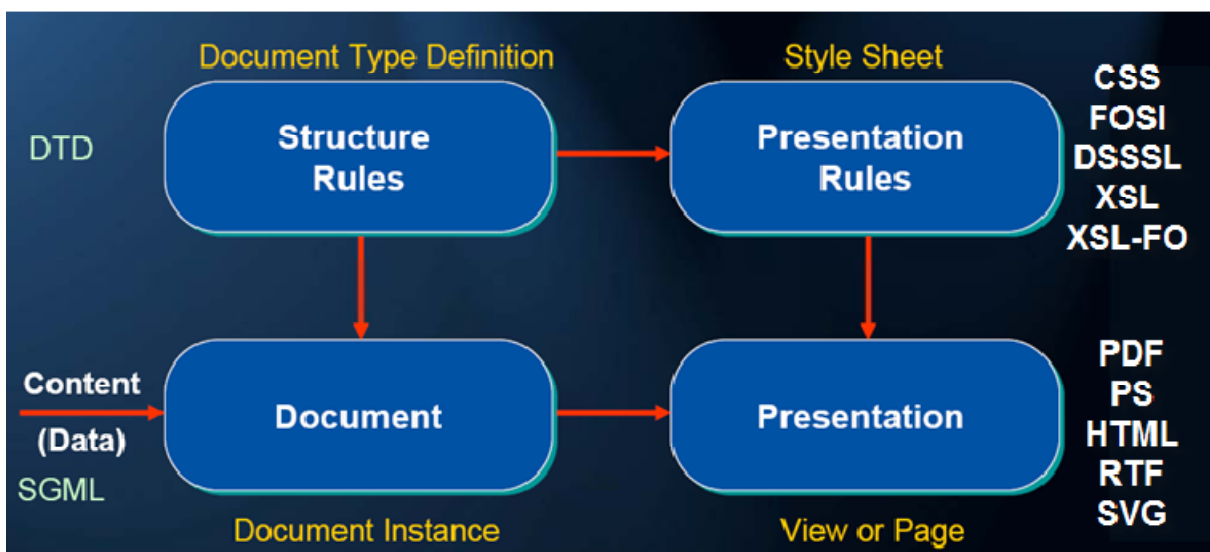
Afbeelding: Bill Wheat, Delta Airlines © Suite Solutions



Bij een typische XML werkwijze ontstaat er een samenwerking tussen de verschillende XML-instanties: de *Document Instance*, de DTD, de XSL-template of *Style Sheet* en de presentatiepagina waarop de informatie wordt vormgegeven.

Inhoud wordt in de vorm van gegevens toegevoegd aan een document, de *XML Document Instance*. Inhoud kan bestaan uit tekst, afbeeldingen, koppelingen of verwijzingen, eventueel aangevuld

met computerdata en verrijkt met metagegevens. De structuurregels, vastgelegd in de DTD of Schema, zorgen voor de juiste samenhang. De presentatieregels in de Style Sheets baseren zich op de structuur en voegen formattering of opmaak toe aan de inhoud, filteren de inhoud en voegen desgevraagd extra informatie toe die nodig is voor de gewenste presentatie. De publicatiesoftware tenslotte vertaalt de XML naar een weergave met gewenste opmaak voor het juiste presentatiekanaal.



Afbeelding: Bill Wheat, Delta Airlines © Suite Solutions

De belofte van XML

XML was veelbelovend en leek de oplossing voor de problemen waar documentatieafdelingen zich voor gesteld zagen. Single sourcing, hergebruik, betekenisvolle mark-up en uitwisselbaarheid – dat alles kan met XML.

Toch bleek het mogelijk om XML net zo complex en onwerkbaar te maken als voorheen SGML

Het IBM-ontwikkelteam besloot daarom XML te combineren met de principes van minimalism en taak-georiënteerd schrijven en ontwikkelde een systeem waarmee zowel inhoud (gestructureerd, topic-georiënteerd), ontwerp (maps, architectuur) en verwerking konden worden beheerd. Dit leidde uiteindelijk tot de presentatie van een XML-architectuur die voorzag in een standaard DTD, uitbreidbaarheid binnen de standaard en zodanig gestructureerd dat het modulair (ver)werken van informatie wordt afgedwongen: DITA.

Bij het ontwikkelen van de DITA-standaard is het een belangrijk uitgangspunt geweest om er voor te zorgen dat naast de DTD ook alle andere tools ontwikkeld worden die het mogelijk maken het proces van inhoudscreatie tot presentatie uit te voeren. Zo ontstond de DITA Open Toolkit: DITA-OT (<http://dita-ot.sourceforge.net/>).

De Open Toolkit is een Java-implementatie van de DITA DTD en werkt onder alle bekende besturingssystemen. De meeste DITA-ondersteunende XML-bewerkers zoals XmetaL, oXygen, FrameMaker, hebben inmiddels de Open



Toolkit toegevoegd aan hun standaard installatie en ondersteunen hiermee ook de uitvoer van DITA via de Open Toolkit processen.

Bronnen:

[An XML Architecture for Technical Documentation: The Darwin Information Typing Architecture](#)

[History of DITA by Bob Doyle](#)

iDTP – international Desk Top Publishing

iDTP is een klein bedrijf dat ondersteuning verzorgt bij alle vormen van technische documentatie en verwerken van informatie. Dat kan door u werk uit handen te nemen (DTP-werk, dataconversie), door advies te geven bij het selecteren van een CMS en leveranciers voor automatisering, door het geven van training en werkplaatsondersteuning en door het maken van documentontwerpen en sjablonen in alle gewenste programma's.

Zelfs, of juist wanneer u nog niet weet of en hoe u uw informatiestroom efficiënter kunt maken kan iDTP u helpen met het doen van analyses en u informeren over mogelijkheden. iDTP verzorgt trainingen om u en uw medewerkers vertrouwd te maken met XML, met DITA of met Framemaker en de combinatie daarvan. Dit zowel voor beginners als gevorderden.

Wilt u meer weten, aarzel niet en neem vrijblijvend contact op.

e-mail: **info@idtp.eu**
website: **www.idtp.eu**

telefoon: **0652036811**

Met vriendelijke groet,
Wim Hooghwinkel