

De legpuzzel van digitale duurzaamheid

Barbara Sierman

Vooraf

De wereld van digitale duurzaamheid wordt gekenmerkt door Engelse termen, vertaling daarvan in het Nederlands maakt het moeilijker het internationale jargon te blijven volgen en verlaagt de herkenbaarheid. Ik heb hier zoveel mogelijk geprobeerd het Nederlands te hanteren, maar koos voor de Engelse term als dit meer gebruikelijk is.

Inleiding

In 1999 was Jef Rothenberg een van de eersten om het vraagstuk van digitale duurzaamheid op de agenda te plaatsen.¹ Hij schreef een artikel waarin hij zich voorstelde dat zijn kleinkinderen van hem een cd-rom ontvingen. In een begeleidende brief vertelde hij over het fortuin dat hij verstopt had. De cd-rom bevatte aanwijzingen hoe die schat te vinden. Op het moment dat ze de cd-rom in handen kregen, zou het echter het jaar 2045 zijn. Hij vroeg zich óf en hoe ze de inhoud van dat schijfje konden lezen.

Vele organisaties herkenden dit probleem. Niet alleen erfgoedinstellingen als bibliotheken, archieven en musea, maar ook ruimtevaartinstellingen, onderzoeksinstituten, enzovoort hebben immers naast hun analoge bestanden steeds meer digitaal materiaal. Vele artikelen, conferenties en presentaties volgden en er ontstond een nieuw onderzoeksgebied, dat van digitale duurzaamheid of *digital preservation*. Inmiddels wordt er wereldwijd op grote schaal onderzoek gedaan. Nationale overheden stellen geld beschikbaar, de Europese Commissie draagt bij met speciale programma's² en de Amerikaanse overheid startte het project NDIIP³. Op internet is



Figuur 1. De inhoud van een cd-rom zal in 2045 niet meer leesbaar zijn. Foto: www.flickr.com, didmyself

¹ 'Ensuring the Longevity of Digital Information', *Scientific American Magazine*, januari 1995.

² In het zesde en zevende Kaderprogramma van de Europese Commissie is speciale aandacht voor digitale duurzaamheid. Projecten als Planets, Caspar, DPE, Driver en Parse-Insight zijn daarin ondergebracht. In Canada financiert The Social Sciences and Humanities Research Council of Canada's Community-University Research Alliances (SSHRC-CURA) het InterPARES project dat sinds 1991 loopt, zie: www.interpares.org.

³ NDIIP: National Digital Information Infrastructure and Preservation, zie voor meer informatie op de website van de Library of Congress: www.digitalpreservation.gov.

een overweldigende hoeveelheid literatuur over digitale duurzaamheid te vinden. Ondanks al deze inspanningen zijn we er nog niet. Je zou de vergelijking met een enorme puzzel kunnen maken. Op de doos van de puzzel staat een fraaie afbeelding van een groep toekomstige gebruikers, die de door hun (groot)ouders gemaakte digitale objecten bekijken, ermee spelen, ze afluisteren of nalezen. In de doos zitten de duizenden stukjes die na veel inspanning en doorzettingsvermogen, eindelijk één geheel zullen vormen. Het complete plaatje is dan zichtbaar. Digitale duurzaamheid is te vergelijken met de inhoud van die doos. Vele instellingen werken aan één of meer van die stukjes. Sommige zoeken eerst alle stukjes die gezamenlijk de rand van de puzzel zullen vormen. Andere zoeken alle blauwe stukjes bij elkaar om de lucht compleet te maken. Veel acties vinden gefragmenteerd plaats en zullen, willen ze op grote schaal bruikbaar zijn, geïntegreerd moeten worden. De praktijkhandleiding is nog niet geschreven.

Maar wanneer we uiteindelijk gezamenlijk in staat zullen zijn de puzzel compleet te maken, toont dat aan dat we ons digitaal erfgoed goed kunnen beheeren en toegankelijk kunnen houden. Dat is de taak waar erfgoedinstellingen zich nu voor gesteld zien. In dit artikel wil ik een overzicht geven van wat er zoal in die doos met puzzelstukjes zit.

Waarom digitale duurzaamheid?

Er zijn vele definities van digitale duurzaamheid in omloop. De omschrijving van de ISO-standaard op dit gebied, het OAIS-model (zie verder) is “the act of maintaining information, in a correct and independently understandable form, over the long term”.⁴ Andere definities variëren hierop, zoals de definitie van Jones en Beagrie, die spreken van “series of managed activities necessary to ensure continued access to digital materials”.⁵ Twee belangrijke activiteiten springen naar voren: het managen van de data én het toegankelijk houden van het materiaal.

Waarom is digitale duurzaamheid nodig?

In toenemende mate verschijnen onze uitingen in digitale vorm en dat zal ertoe leiden dat instellingen die waken over ons cultureel erfgoed, geconfronteerd worden met het vraagstuk hoe deze digitale informatie voor de toekomst te bewaren. En dan spreken we nog niet over hoe ver die toekomst ligt, tientallen of mogelijk zelfs honderden jaren (wanneer je het vergelijkt met wat er nu in de analoge collecties de tijd heeft weten te doorstaan).

In 2003 nam de UNESCO het *Charter on the Preservation of Digital Heritage*⁶ aan, waarin de uitgangspunten voor het bewaren van digitaal erfgoed staan geformuleerd. Lidstaten onderschreven deze uitgangspunten en gaven daarmee aan hun

4 CCSDS, Reference model for an Open Archival Information System (OAIS) (2202), 1-11, zie: <http://public.ccsds.org/publications/archive/650xob1.pdf>.

5 Maggie Jones & Neil Beagrie, *The Digital Preservation Coalition, The Preservation Management of Digital Material Handbook* (onlinetekst): www.dpconline.org/graphics/handbook.

6 Zie: http://portal.unesco.org/ci/en/files/13367/1070015911Charter_en.pdf/Charter_en.pdf.

verantwoordelijkheid op dit gebied te accepteren. Niet alleen bibliotheken, archieven en musea werden betrokken, maar ook universiteiten en onderzoeksinstituten. Deze instellingen beheren allemaal digitaal erfgoed, of het nu teksten, kunstwerken of wetenschappelijke publicaties zijn. Met dit UNESCO-charter werd internationaal erkend dat digitaal materiaal niet vanzelfsprekend wel zal overleven, maar dat er actief mee omgegaan moet worden. Om het populair te zeggen: “negeren is geen optie”.

Wat is er dan zo speciaal?

Waarom is digitaal materiaal zo apart dat het een speciale behandeling nodig heeft? Voor het maken van een digitaal object, zoals een tekstdocument, een website of een database, gebruikt men software. Deze software draait alleen in een specifieke omgeving, bijvoorbeeld een bepaald besturingssysteem dat op bepaalde apparatuur (hardware) draait. Het resultaat, het digitale bestand, bestaat uit nullen en enen, die door een mens niet gelezen kunnen worden. Er is weer software, een besturingssysteem en hardware voor nodig om de nullen en enen in het oorspronkelijke document om te zetten. Neem een tekstdocument in MS-Word®. Je kunt deze file openen zonder het MS-Word-programma, bijvoorbeeld in Kladblok. In het gunstige geval krijg je dan wel de tekst te lezen, maar alle speciale tekens en opmaak zijn verdwenen. Het document is dus niet meer gelijk aan het document zoals de maker het opleverde. Dat kun je alleen nog zien wanneer je het document weer in MS-Word leest, of een speciale MS-Word reader. Om de toegankelijkheid te handhaven, heb je dus bijbehorende software en hardware nodig.

Soft- en hardwareontwikkelaars zitten niet stil en brengen regelmatig nieuwe versies van hun producten uit. Niet altijd ondersteunt de nieuwe software het gebruik van files die in vorige versies zijn aangemaakt (compatibiliteit). Soms verdwijnen softwarepakketten van de markt, zoals met tekstverwerkers uit de jaren zeventig van de vorige eeuw gebeurde.

Opslag, ofwel *storage*, heeft de afgelopen jaren een onstuimige groei doorgemaakt, waarbij de leveranciers van opslagsystemen de capaciteit wisten uit te breiden van megabytes naar gigabytes en petabytes. Zij ontwikkelden een reeks aan opslagmedia, van tape tot optische schijven en de holografische disk. Intelligente opslagsystemen monitoren de opgeslagen informatie en waarschuwen als zaken fout gaan. Desondanks verdwijnt er nog steeds informatie door de kwetsbaarheid van magnetische en optische schijven en de fysieke achteruitgang die plaatsvindt.

Ook hardware vernieuwt voortdurend, floppy disks van 5.25 inch werden opgevolgd door die van 3.5 inch en vandaag de dag heeft een nieuwe computer geen floppydisk-drive meer. Deze hardware is gewoon verdwenen en vervangen door bijvoorbeeld usb-poorten.



Behalve technische aspecten speelt ook de context van digitale objecten een belangrijke rol. Welke informatie heeft de toekomstige gebruiker nodig om het digitale object in zijn oorspronkelijke omgeving te kunnen plaatsen en hoe geven we deze informatie aan het object mee?

De natuurlijke achteruitgang en de permanente ontwikkelingen van de digitale omgeving, of het nu software of hardware is, vormen een bedreiging voor de duurzame toegankelijkheid van digitaal materiaal. Er zijn speciale acties nodig om de gevolgen op te vangen. Digitale duurzaamheid gaat over deze activiteiten.

Natuurlijk zijn er ook gevallen bekend van digitaal materiaal dat behoed is voor verdwijning. Een van de meest bekende voorbeelden is het *Domesday Book*. Bij dit project van de BBC werd de viering van het 900-jarige bestaan van het *Domesday Book* gevierd door inwoners te vragen mee te helpen om een nieuwe, digitale versie te maken. De resultaten werden opgeslagen op zeer geavanceerde schijven. Enkele jaren later bleek dat deze geavanceerde schijven niet langer leesbaar waren en de gegevens bleken ontoegankelijk. Met veel geld en inspanning is er een geslaagde reddingspoging gedaan.⁷ Eén conclusie die uit dit voorbeeld getrokken kan worden, is dat digitale duurzaamheid goed te realiseren is, als er genoeg geld beschikbaar is. De praktijk is echter anders.

Een organisatie die haar digitale bezit toegankelijk wil houden voor toekomstig gebruik, zal voorbereidingen treffen om, vanuit haar beleid (de *preservation policy*) tijdig de juiste maatregelen te nemen (de *preservation actions*). Bij de aanpak kunnen verschillende overwegingen een rol spelen. Bijvoorbeeld de mate waarin we rekening willen houden met deze toekomstige gebruikers. Uiteraard moeten zij erop kunnen vertrouwen dat het digitale object niet ongecontroleerd gewijzigd is, en dat het nog steeds weergeeft wat het voordoet te zijn (*authenticity*), iets dat met name bij archieven een rol speelt. Maar willen de toekomstige gebruikers wel de oorspronkelijke verschijningsvorm, de *'original look and feel'* ervaren? Of gaat hun voorkeur uit naar het ervaren van de digitale objecten in hun eigen omgeving, bijvoorbeeld op de mobiele telefoon? In beide gevallen zul je specifieke informatie vastleggen, om later, aan de hand van die informatie, de gewenste weergave te kunnen realiseren.

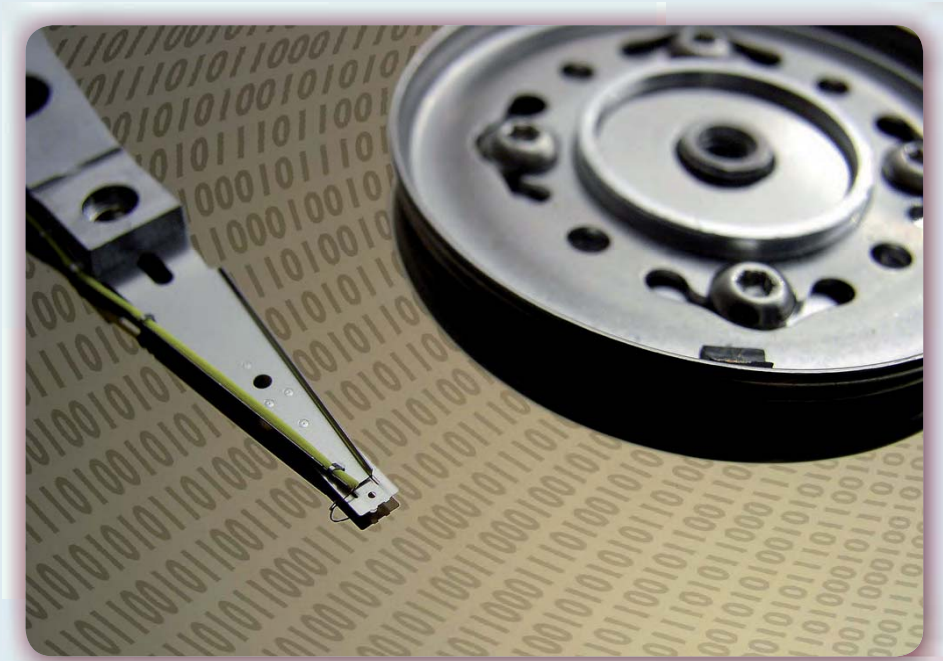
Het ontstaan van een digitaal object

Bij digitale objecten onderscheid men *'born digital material'*, dat vanaf het moment van creatie al digitaal was, zoals een digitale foto of een rapport dat met een tekstverwerker op de computer is gemaakt. Daarnaast kennen we materiaal dat als analog object begon, maar waarvan door digitalisering een digitale versie is gemaakt: *'digitized material'*.

Tussen beide objecten zit vanuit een duurzaamheidsstandpunt één groot verschil: in het laatste geval is er een analog object, dat als uitgangspunt kan dienen. De eigen-

7 Phil Mellor, 'CAMiLEON: Emulation and BBC Domesday', *RLG Digitnews*, 7/2 (2003), <http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file3600.html#feature3>.

DE LEGPUZZEL VAN DIGITALE DUURZAAMHEID



127

Figuur 2. Binaire code ofwel de nullen en enen vormen de essentie van digitale duurzaamheid.
Foto: www.flickr.com, wilhei

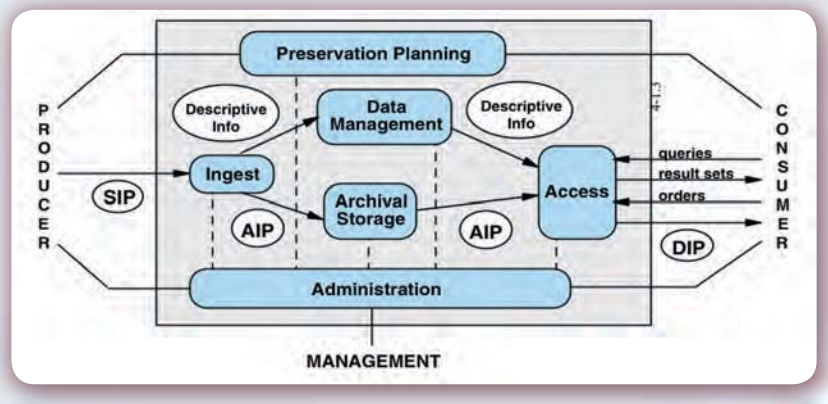
schappen van dit analoge object, zoals kleur, toepassingsmogelijkheden, grootte en omvang, zijn het ijkpunt waaraan de digitale variant idealiter zou moeten voldoen. Een gedigitaliseerd boek zou net zoveel pagina's moeten hebben als de originele versie (en in dezelfde volgorde), de illustraties dienen dezelfde kleurschakeringen en resolutie te bevatten, en ook het omslag waarmee het boek oorspronkelijk verscheen, dient onderdeel te zijn van de gedigitaliseerde versie. Wanneer er na digitalisering andere acties op het object gaan plaatsvinden, bijvoorbeeld een migratie, dan vormen deze eigenschappen de uitgangspunten voor de actie en de karakteristieken die bewaard moeten blijven. De *borndigital* objecten ontberen deze voorganger. Ze kunnen wel vele malen worden gekopieerd, maar niet opnieuw worden gegenereerd. Zowel bij *borndigital*- als bij gedigitaliseerde objecten start de activiteit om het digitale object duurzaam te bewaren, vanaf het prille (digitale) begin. Immers, op het moment van creatie worden keuzes gemaakt die van grote invloed zijn op de overlevingskans van het digitale object, zoals de keuze voor bestandsformaat, wachtwoorden, fonts, et cetera.

Bij de term 'digitaal object' moet men overigens niet te strikt denken aan één bestand of bitstream. Veelal bestaat het te bewaren geheel, ook wel de '*intellectual entity*' genoemd, uit een combinatie van meerdere bestanden, of uit meerdere objecten.

De OAIS-standaard

In de wereld van digitale duurzaamheid maakt men zoveel mogelijk gebruik van algemeen aanvaarde standaarden. Een van de belangrijke standaarden is het *Reference Model for an Open Archival Information System* (OAIS), opgesteld door de Consultative Committee for Space Data Systems (CCSDS) en in 2003 tot ISO-norm uitgeroepen.

Het OAIS-model is een referentiemodel: het beschrijft op conceptueel niveau een digitaal archief voor langetermijnopslag en de bijbehorende verantwoordelijkheden, zonder op de feitelijke uitwerking in te gaan. Het is aan de organisaties zelf dit conceptuele model te vertalen naar de praktijk.



Figuur 3. Functionele entiteiten van het OAIS-model

Het OAIS-model beschrijft een zestal functionele entiteiten van een digitaal archief (zie Figuur 3). Het digitale object, het *information package*, krijgt in verschillende stadia van verwerking, zoals ontvangst, opslag en teruggave, een andere naam.

Informatie komt binnen als een *submission information package* (SIP), afkomstig van een *producer*. Deze producer kan een uitgever zijn, of een lokaal archief, of een universiteitsbibliotheek die haar digitale bestanden ter bewaring geeft aan een speciaal daarvoor ingericht langetermijnarchief. Dat kan op vrijwillige basis of omdat dit wettelijk verplicht is, bijvoorbeeld bij archieven.

Na verschillende controles is deze SIP gereed voor de volgende stap, *ingest*, waarbij het object in de uiteindelijke opslag, de *archival storage*, als *archival information package* (AIP) komt. De bijbehorende metadata komen terecht in *data management*. Wanneer een *consumer* (de eindgebruiker) het object via *access* wil bekijken, krijgt hij een *dissimination information package* (DIP), waarin het digitale object is opgenomen om het te tonen.

Tijdens de levensduur van een object in het archief vindt er *preservation planning* plaats. Hierin vinden de noodzakelijke activiteiten plaats om het object toegankelijk te houden (de *preservation actions*). Uiteraard zijn er ook activiteiten met betrekking

tot monitoring van deze processen en van de soft- en hardware van het archief zelf, deze vallen onder *administration*.

Het model gaat verder uitgebreid in op de informatie die samen met het digitale object (de *content information*) moet worden opgeslagen in de *preservation description information*. Deze extra informatie omhelst gegevens als beschrijvende metadata zoals titel- en auteurgegevens, informatie over de herkomst (*provenance*) van het object, verwijzingen naar gerelateerde objecten in het archief en technische informatie die aangeeft of het object niet onbedoeld is gewijzigd tijdens de opslag (*fixity*). Behalve de aandacht die het model besteedt aan de inrichting van een archief en de informatie die over een object moet worden opgeslagen, speelt de toekomstige gebruiker, de *designated community*, een belangrijke rol. Deze zal immers het opgeslagen object niet alleen gebruiken, maar ook zelfstandig moeten kunnen begrijpen. Het OAIS-model besteedt ook aandacht aan de verantwoordelijkheden die een organisatie met een langetermijnarchief draagt. Deze uitgangspunten zijn in verschillende methodes vertaald naar praktische toepassingen (zie verder de paragraaf 'Digitale duurzaamheid en de organisatie'). Voor elke organisatie geldt echter dat een kant-en-klaaroplossing nog niet voorhanden is. Het vereist veel inspanning om de organisatie zó in te richten dat aan de OAIS-eisen kan worden voldaan.

Door de uitgebreidheid van het model, functioneert het binnen de digitaleduurzaamheidsgemeenschap als een belangrijke referentie. Een groot voordeel van het OAIS-model is dat men wereldwijd dezelfde terminologie kan hanteren, of men nu in Europa, Azië of Amerika over digitale duurzaamheid spreekt.

Bewaren van digitale informatie

Niveaus van bewaren (preservation levels)

Niet alle instellingen die digitale objecten beheren, kunnen garanderen dat voor de 'eeuwigheid' te blijven doen – omdat ze daar de capaciteit niet toe hebben of omdat het hun taak niet is. Zij zullen digitale objecten 'gebruiken' binnen hun organisatie, voordat zij deze overhandigen aan een van de organisaties die meer toegerust zijn voor het bewaren op langere termijn. Het landschap van instellingen die deze taak op zich zullen nemen, is nog niet volledig in kaart gebracht, maar zal, in navolging van de analoge wereld, waarschijnlijk vooral bestaan uit cultureel-erfgoedinstellingen, zoals bibliotheken, archieven, musea, datacentra, ... Mogelijk dat op termijn ook commerciële instellingen deelnemen.

Door middel van *preservation levels* geeft een organisatie aan in welke mate het object is voorbereid voor langdurige bewaring. De meest eenvoudige vorm van bewaren is *bit level preservation*. Daarbij zullen de bits zo goed mogelijk bewaard blijven, maar is bijvoorbeeld minder aandacht besteed aan het verzamelen van metadata over de context van het object. Zo zijn er meer varianten in *preservation levels*, bij-



voorbeeld *access preservation* waarbij de nadruk ligt op de toegankelijkheid van het object of *representation preservation*, waarbij vooral aandacht wordt geschonken aan de weergave van het object.

Bestandsformaten

Er is een ruime variatie aan programma's beschikbaar om digitale objecten te maken. Het eindproduct kan bestaan uit één of meerdere bestandsformaten. Het aantal verschillende bestandsformaten dat bekend is, loopt nu in de duizenden.

Een instelling met een langetermijnarchief zal vanuit beheersoogpunt het liefst het aantal bestandsformaten beperkt houden. Het bijhouden van kennis over zoveel verschillende bestandsformaten is een kostbare taak. Deze keuze is niet altijd mogelijk: archieven en bibliotheken krijgen vaak materiaal binnen zonder dat zij de keuze hebben om deze bestandsformaten te weigeren.

Maar als een instelling zelf digitale objecten creëert, bijvoorbeeld bij digitalisering, is de keuze van bestandsformaat, met daarbij aandacht voor de duurzaamheid ervan, het overwegen waard. Er zijn inmiddels een aantal algemeen aanvaarde criteria voor 'duurzame' bestandsformaten. Bijvoorbeeld: is het een open standaard, dus zijn de specificaties van het formaat openbaar toegankelijk? Wordt het bestandsformaat op grote schaal gebruikt? De veronderstelling hierachter is dat bij een grote gebruikersgroep de leverancier, of anderen, er heil in zien om het bestandsformaat in productie te houden. Of om oplossingen te verzinnen, als de leverancier afhaakt, juist omdat zoveel mensen er belang bij hebben. Een andere overweging kan zijn of je de inhoud van het bestand ook gewoon kunt lezen. Dit geldt bijvoorbeeld voor XML-bestanden, daar kun je de inhoud van lezen, hoewel echt begrijpen pas kan als je ook de bijbehorende uitleg van de verschillende eenheden hebt.

Een aantal bestandsformaten verdient vanuit duurzaamheidsstandpunt de voorkeur. Bijvoorbeeld PDF/A van Adobe, dat speciaal voor de archiefwereld is ontwikkeld. Bepaalde functionaliteiten, die problemen voor duurzame bewaring kunnen opleveren, zoals wachtwoorden in de bestanden, zijn hierin niet toegestaan.

Het vaststellen van een bestandsformaat

De schrijver, fotograaf of websitebouwer die een digitaal object creëert, krijgt met de ontwikkelingen in software steeds meer mogelijkheden tot zijn beschikking: tekstbestanden kunnen plaatjes bevatten, aan een presentatie kun je geluiden toevoegen, er kunnen verschillende lettertypes in een tekst worden gebruikt, op websites verwijst men naar andere websites door middel van links, en men kan zich beveiligen door het document van een wachtwoord te voorzien. Voor het duurzaam bewaren van objecten is het minimaal nodig te weten welk bestandsformaat en bijbehorende versie het object heeft.

DE LEGPUZZEL VAN DIGITALE DUURZAAMHEID

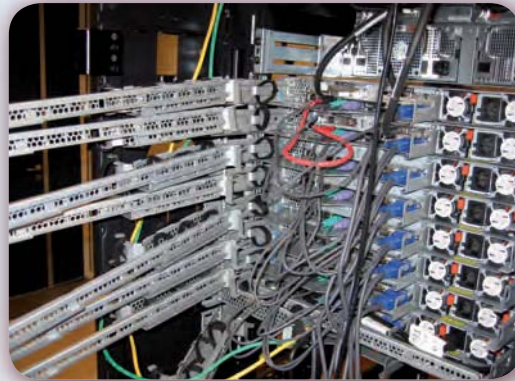
Bestanden worden aangeleverd met een zogenaamde *file extension*, de drieletterige afkorting na de punt, zoals .pdf, .doc of .txt. Maar helaas zeggen deze drielettercodes niet veel. Een doc-bestand kan met vele programma's zijn aangemaakt en de auteur kan zelf de extensie wijzigen. Daarenboven geeft de afkorting niet aan welke versie van het bestandsformaat van toepassing is. En dus zijn er speciale programma's gemaakt, die deze informatie uit het bestand delven, zodat de gegevens als metadata aan het object kunnen worden toegevoegd.

Een van de programma's is Digital Record Object Identification (DROID)⁸, dat op basis van informatie in het bestand de *file signature* leest en aan de hand daarvan vaststelt wat het bestandsformaat is. DROID kan dit voor een zeer groot aantal bestandsformaten.

Meer functionaliteit heeft de tool JSTOR/ Harvard Object Validation Environment (JHOVE)⁹, een opensourcetool die in staat is het object te *identificeren*, dat wil zeggen, vast te stellen welk bestandsformaat en versie het object heeft. Daarnaast kan JHOVE het object *valideren*: voldoet het object aan de specificaties van dat bestandsformaat, syntactisch en semantisch? En JHOVE kan het object *karacteriseren* door te analyseren welke toepassingen en eigenschappen van dat bestandsformaat gebruikt zijn in het object. Zijn er bijvoorbeeld speciale fonts gebruikt? En zijn die dan opgenomen in het bestand

(*embedded*)? Fonts zijn een speciaal probleem binnen duurzaamheid. Om de oorspronkelijke weergave te bereiken, wil je ook de oorspronkelijke fonts laten zien. Dit kan alleen wanneer deze bij het object worden opgeslagen, anders kan het object, getoond op een andere computer zonder die fonts, het object foutief weergeven, met verschoven regels, formules die niet meer juist zijn of tabellen met verkeerde inhoud in cellen.

Voor een tiental van de meest gangbare bestandsformaten, kun je JHOVE goed gebruiken om het document te analyseren. In het JHOVE2-project (2008-2010) zullen de tekortkomingen van de reeds ontwikkelde tools verbeterd worden en komt er uitgebreide documentatie, een betere foutafhandeling en een mogelijkheid om samengestelde objecten (dus met meerdere verschillende bestanden in één object, bijvoorbeeld websites) te verwerken.



Servers: een blik achter de schermen van digitaal archiveren. Foto: www.flickr.com, jemimus

⁸ Zie voor meer informatie: <http://droid.sourceforge.net/wiki/index.php/Introduction>.

⁹ Zie: <http://hul.harvard.edu/jhove>.

Metadata

Metadata worden veelal gedefinieerd als 'data over data'. Bij digitale objecten zijn metadata nodig om informatie mee te geven die essentieel is voor het toegankelijk houden van het object. In welke omgeving is een object gemaakt? Welke software is er nodig om het te tonen? Welke structuur heeft een object? Welk bestandsformaat? Enzoverder.

Er zijn verschillende soorten metadata en de inhoud van elke soort is niet strikt afgebakend, maar we kunnen een onderverdeling maken op basis van hun functie. Er zijn metadata die nodig zijn om het object te identificeren en te vinden, ook wel beschrijvende of bibliografische metadata (*descriptive metadata*) genoemd, bijvoorbeeld Dublin Core, ISAD(G) of MARC. Daarnaast zijn er metadata nodig om de toegankelijkheid te regelen: structuurinformatie over het object, toegangsinformatie en allerlei informatie die nodig is om het object juist weer te geven. Hiervoor worden onder meer standaarden als MODS, METS en MPEG21 gebruikt.

Een derde groep metadata is nodig om het object langdurig te bewaren: *preservation metadata*. In 2005 zag de *Premis Data Dictionary* het licht, waaraan verschillende internationale erfgoedinstellingen een bijdrage leverden. Inmiddels is er een tweede versie verschenen en wordt de *Premis Data Dictionary* als een de facto standaard gezien. Het bevat een set aan metadata ten behoeve van digitale duurzaamheid. Bij elk element wordt een redenering gegeven waarom deze metadata belangrijk zijn voor het object. Niet elk element hoeft bij elk object te worden vastgelegd. Het is aan de bewarende instelling om te bepalen welke gegevens waar worden vastgelegd, maar de lijst helpt bij het vaststellen van "the information a repository uses to support the digital preservation process"¹⁰. Sommige elementen horen bij een object, andere elementen zijn van toepassing op een reeks van objecten en zouden dus ook op één centrale plaats bewaard kunnen worden. Een zogenaamd *registry*, dat deze informatie bewaart en bijhoudt en internationaal toegankelijk houdt, zou deze taak van een instelling kunnen overnemen. Eén voorbeeld van een dergelijk *registry* voor onder meer bestandsformaatinformatie is PRONOM¹¹ van The National Archives in Londen.

Hoewel metadata voor het duurzaam bewaren van objecten van vitaal belang zijn, is het ook een kostbare aangelegenheid de juiste metadata te verzamelen. Toevoegen van kwalitatief goede metadata vraagt om opgeleid personeel. Bij grote hoeveelheden digitaal materiaal kan dit eenvoudigweg niet meer handmatig. Programma's als DROID en JHOVE helpen deze gegevens automatisch te genereren.

Het beste moment om metadata te creëren is bij het ontstaan van het object, maar de auteur of schepper is zich hier nog te weinig van bewust. Veel software geeft de auteur bijvoorbeeld de mogelijkheid om direct al metadata mee te geven via de 'eigenschappen' van het document.

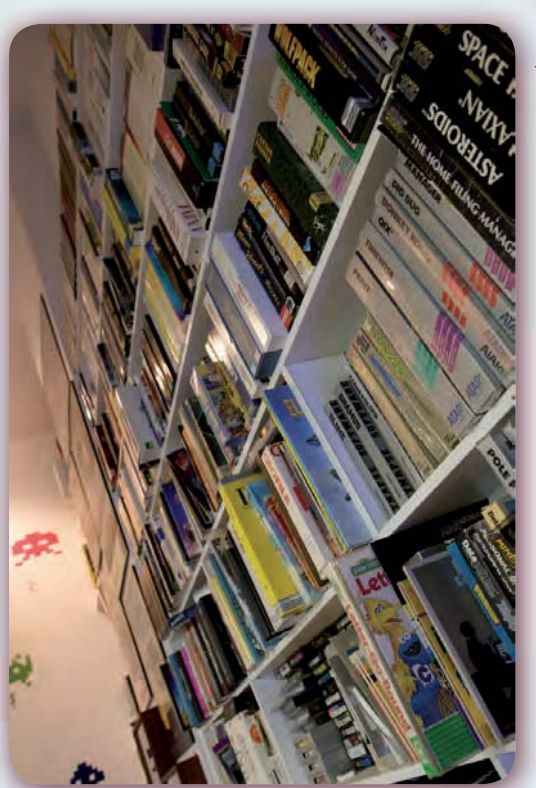
¹⁰ Voor meer informatie over Premis, zie: www.loc.gov/standards/premis, waar ook de integrale tekst te vinden is. Het citaat staat op p. 3.

¹¹ Zie: www.nationalarchives.gov.uk/pronom.

Karakteristieken van een digitaal object

Elke actie die op een digitaal object plaatsvindt om de duurzaamheid te waarborgen, heeft informatie nodig over eigenschappen over het oorspronkelijke digitale object. Deze oorspronkelijke eigenschappen (ook wel karakteristieken of *significant properties* genoemd) vormen een ijkpunt. Doordat deze karakteristieken zijn vastgelegd, kan men ook bepalen of de actie nog wel recht doet aan deze karakteristieken of dat enkele karakteristieken na de actie verdwenen zijn. Ook voor de juiste presentatie van digitale objecten kunnen karakteristieken een rol spelen, denk maar aan een kleurenmonitor versus een zwart-witmonitor.

Voor een organisatie is het belangrijk te bepalen welke karakteristieken voor een object of een collectie van belang zijn. Sommige objecten vereisen bijvoorbeeld een zeer nauwkeurige kleurenweergave omdat, als de kleuren niet goed worden weergegeven, een deel van de betekenis verloren gaat, zoals bij het gebruik van kleuren in tabellen of kunstobjecten.



133

Het bewaren van alle informatie over software is een must. Foto: www.flickr.com, mwichary

Duurzaamheidsstrategieën

Duurzaamheidsstrategieën, ofwel *permanent access strategies*, zijn bedoeld om digitale objecten toegankelijk te houden. Een reden om een dergelijke strategie toe te passen kan zijn dat een bestandsformaat door de leverancier niet langer onderhouden wordt en *obsolete* dreigt te worden. In de toekomst zullen onder meer registries als die van PRONOM op basis van de verzamelde informatie waarschuwen als dit het geval is.

De keuze van de te volgen strategie is afhankelijk van verschillende factoren, zoals de karakteristieken van het digitale object zelf, het beschikbare budget en welke dienst men de toekomstige gebruiker wil aanbieden. In het algemeen onderscheidt men drie strategieën: het bewaren van technologie, migratie en emulatie.

Technologie bewaren

Er zijn verschillende organisaties die hardware en software bewaren en werkende houden. In Nederland is er bijvoorbeeld een computermuseum bij de Universiteit van Amsterdam waar dit gebeurt. Hoewel deze verzamelingen van belang kunnen zijn om in te kunnen spelen bij ad-hocrampen, is het materiaal te kwetsbaar (plastics verouderen, onderdelen zijn aan slijtage onderhevig en niet meer te vervangen) om de miljoenen digitale objecten die bewaard moeten blijven, aan toe te vertrouwen. Het wordt in het algemeen eerder gezien als een aanvullende strategie en de aandacht gaat dan ook meer uit naar migratie en emulatie als blijvende oplossingen.

Migratie

Wanneer dreigt dat een digitaal object niet meer getoond kan worden omdat het bestandsformaat niet langer ondersteund wordt, kan de beheerder besluiten het digitale object te migreren naar een ander bestandsformaat, dat bestendiger of actueler is. Hoewel migratie een algemeen geaccepteerde strategie is, zitten er haken en ogen aan. In de eerste plaats is het geen blijvende strategie, je zult het omzetten naar een meer gangbaar formaat steeds moeten herhalen op het moment dat het huidige formaat geen toegang meer biedt. Daarnaast loop je elke keer het risico dat de migratie kleine fouten introduceert, die in de loop der tijd cumulatief tot grotere fouten kunnen leiden. Testen zou een oplossing kunnen zijn. Maar het testen van gemigreerde objecten ten opzichte van het origineel is niet eenvoudig. Waar moet je op testen? Wat zijn de cruciale eigenschappen van een document, die je wenste te behouden en die dus getest kunnen worden? Een hulpmiddel om dit vast te stellen, kan de eerder genoemde JHOVE-programmatuur zijn.

Er zijn verschillende vormen van migratie:

- Van een bestandsformaat naar hetzelfde formaat, maar dan een hogere versie.
- Van een bestandsformaat naar een ander bestandsformaat dat meer geschikt is voor langdurige opslag. Bijvoorbeeld van een MS-Wordformaat naar PDF/A-formaat. Een dergelijke migratie heet 'normalisatie' wanneer deze conversie plaatsvindt vóórdat het object wordt opgeslagen in het archief. Het is ook mogelijk deze activiteit achteraf uit te voeren op objecten die al zijn opgeslagen, via een *batch* migratie.
- Migratie op verzoek (*migration on request*) is een methodiek waarbij men het object opslaat en pas maatregelen neemt wanneer het object weer wordt opgevraagd; als men dan constateert dat het object niet langer leesbaar is in de opvraagomgeving, past men een migratie toe die het weergeven ervan voor de gebruiker weer mogelijk maakt.¹²

¹² Zie hiervoor Mellor, Wheatley & Sergeant, 2002.

DE LEGPUZZEL VAN DIGITALE DUURZAAMHEID

Overigens is het bij migratie gangbaar om naast de gemigreerde versie altijd het origineel te bewaren. Er is immers een kans dat technologische ontwikkelingen in de toekomst het mogelijk maken deze originele file weer te lezen. Emulatie biedt daartoe een kans.

135

Emulatie

Bij emulatie blijft het oorspronkelijke object ongewijzigd, maar recreëert men de oorspronkelijke omgeving waarin het object functioneerde, bovenop een bestaande, eigentijdse computeromgeving. Deze nieuwe omgeving werkt 'virtueel' als de oorspronkelijke omgeving.

Hoewel emulatie zowel op applicatieniveau en besturingssysteemniveau als op hardwareniveau kan plaatsvinden, is binnen digitale duurzaamheid vooral de emulatie van de oorspronkelijke hardware door middel van software (de zogenaamde *software emulation of hardware*) het meest gebruikelijk. De Koninklijke Bibliotheek in Den Haag ontwikkelde samen met het Nationaal Archief de modulaire emulator DIOSCURI¹³. Dit programma kun je installeren op je huidige computer, waarmee je een zogenaamde X86-computer op je eigen computer nabootst. Mits je de juiste software hebt, kun je de bestanden die in die omgeving draaiden, weer gebruiken, bijvoorbeeld MS-DOS-software. Emulatie heeft als voordeel dat je het digitale object ongewijzigd laat, wat niet het geval is bij migratie. Met behulp van de originele software is het immers weer mogelijk het object in zijn originele vorm te tonen. Maar hiermee is tevens een van de nadelen aangegeven, namelijk het bewaren van de originele software en de bijbehorende uitleg over de werking ervan. Software is immers vaak door rechten beschermd, die het bewaren ervan in bijvoorbeeld een 'softwarearchief' waarschijnlijk niet toestaan. Daarnaast is het bouwen van een emulator een gecompliceerde activiteit. Voor sommige digitale objecten echter, zoals websites, zou emulatie wel eens de enige mogelijkheid kunnen zijn, omdat migratie hier vanwege de grote hoeveelheden verschillende bestandsformaten waarschijnlijk geen optie is.



Door emulatie kan men een oorspronkelijke computeromgeving reconstrueren.

Foto: www.flickr.com, blakespot

¹³ Zie voor meer informatie: <http://dioscuri.sourceforge.net>.

Digitale duurzaamheid en de organisatie

Richtlijnen

136

Objecten voor lange termijn bewaren stelt veel eisen aan een organisatie. In het OAIS-model zijn deze op globaal niveau geformuleerd. Verschillende initiatieven zijn gestart om deze eisen om te zetten naar praktische richtlijnen, om te komen tot een duidelijke omschrijving van een *trusted* of een *trustworthy repository*, een betrouwbaar digitaal archief, waaraan je als organisatie met een gerust hart je digitale objecten aan kunt toevertrouwen.



Computermusea bewaren de technologie en de hardware. Foto: www.flickr.com, berl

Deze richtlijnen zijn niet alleen van belang voor degenen die een dergelijk archief willen opzetten, zoals cultureel-erfgoedinstellingen die (al dan niet wettelijk verplicht) digitale collecties voor lange tijd bewaren, maar ze zijn ook van belang voor organisaties die deze activiteit willen uitbesteden. De afgelopen jaren zijn wereldwijd verschillende initiatieven gestart om deze eisen op te stellen. De verwachting is dat er op termijn auditcommissies ontstaan die een digitaal archief zullen certificeren, maar op dit moment bestaan dergelijke instellingen nog niet.

De Research Libraries Group (RLG) en OCLC publiceerden in 2007 *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, als opvolger van de eerste versie uit 2002.¹⁴ Het enthousiasme waarmee deze set aan aanwijzingen werd ontvangen, en het gebruik ervan bij instellingen als 'checklist', leidde tot het initiatief om deze publicatie voor te dragen voor ISO-certificatie. Een internationale werkgroep waaraan iedereen uit het veld kan deelnemen, publiceert regelmatig over haar vorderingen. In Nederland is dit document voor de archiefwereld bewerkt als *ED3 Duurzaam Digitaal Depot*.¹⁵ Het project Digital Preservation Europe DPE benadert het probleem vanuit het standpunt van risicobeheersing en stelde in samenwerking met het Digital Curation Centre (DCC) de tool DRAMBORA samen.¹⁶ De Duitse werkgroep Nestor publiceerde een eigen lijst van criteria, de *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive*.¹⁷

De drie initiatieven werken gelukkig samen en dat leidde in 2007 tot de definitie van de *Ten core principles of trust*, die de leidraad voor digitale archieven zouden moeten

14 Beide publicaties zijn te vinden op www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91. De internationale werkgroep publiceert haar vorderingen op <http://wiki.digitalrepositoryauditandcertification.org/bin/view>.

15 Het document is te downloaden vanaf http://lopai.nl/pdf/ED3_v1.pdf.

16 Deze tool is te downloaden vanaf www.repositoryaudit.eu.

17 Deze is vertaald in het Engels in 2007 en is te downloaden van www.langzeitarchivierung.de/modules.php?op=modload&name=PagEd&file=index&page_id=18#section8.

zijn. Voor beginnende archieven zijn deze uitgangspunten vertaald in *PLATTER*¹⁸, een publicatie met bijbehorende tools om organisaties die beginnen met een digitaal archief, te helpen deze principes te implementeren en op termijn voor certificering in aanmerking te komen.

Kosten van digitale duurzaamheid

Instellingen die een digitaal archief voor de lange termijn opzetten, gaan een verplichting voor vele jaren aan. Maar welke kosten gemaakt moeten worden, is afhankelijk van vele factoren: welk soort bestandsformaten zijn opgeslagen, of de software om *preservation actions* uit te voeren beschikbaar is of dat die nog ontwikkeld moet worden, hoeveel metadata je nog aan de objecten moet toevoegen, enzoverder. Het LIFE (LIFECycle Information for E-literature)¹⁹-project, ontstaan uit een samenwerking tussen de British Library en University College London, heeft een kostenmodel opgezet om hier meer inzage in te kunnen krijgen. Op basis van de verschillende stadia in de levenscyclus van een digitaal object – zoals het ontstaan, de verwerving, het toekennen van metadata, de opslag, de toegang, tot en met eventuele vernietiging – is in kaart gebracht welke kostenposten elk stadium met zich meebrengt.

Samenwerking

Als één ding voor alle betrokkenen duidelijk is, dan is het wel dat digitale duurzaamheid een activiteit is die alleen in een gezamenlijke aanpak tot goede resultaten kan leiden: het vergt simpelweg te veel verschillende kwaliteiten om door één instelling gedaan te worden. Doordat het digitale speelveld voortdurend wijzigt, is blijvend onderzoek nodig.

Uitwisseling van kennis en ervaring vindt op grote schaal plaats, en internet is een dankbare bron voor iedereen die zijn eigen collecties op verantwoorde wijze wil beheren. Behalve instellingen als nationale bibliotheken en archieven, die over hun vorderingen op hun website publiceren, zijn ook de resultaten van verschillende Europese projecten interessant.

Eén van de belangrijkste projecten op het gebied van digitale duurzaamheid is Preservation and Long-term Access through Networked Services (Planets)²⁰, dat loopt van 2007 tot 2010. Met de financiële steun van de Europese Commissie werken achttien instellingen (archieven, bibliotheken, onderzoeksinstituten en commerciële instellingen) aan een praktische ondersteuning van digitale duurzaamheid. Door het Planets-project zullen verschillende functies worden opgeleverd, zoals een tool

18 Zie de website www.digitalpreservationeurope.eu.

19 De projectresultaten zijn te vinden op www.life.ac.uk.

20 Verschillende projectresultaten zijn te vinden op www.planets-project.eu.

DE LEGPUZZEL VAN DIGITALE DUURZAAMHEID

138

die helpt bij het beslissen welke *preservation action* te kiezen, en een testbed om deze acties uit te proberen en ervaringen van anderen te zien. Daarnaast zijn er verschillende bestanden (*registries*) in ontwikkeling, bijvoorbeeld met informatie over te gebruiken tools. Er is een samenwerking met PRONOM. Daarnaast verschijnen verschillende studies over Preservation Planning en andere deelonderwerpen. Ook Digital Preservation Europe (DPE)²¹, eveneens een Europees project, geeft op de website veel praktische informatie.

In Nederland is in 2008 – in navolging van bijvoorbeeld Engeland (Digital Preservation Coalition) en Duitsland (Nestor Kompetenznetzwerk Langzeitarchivierung) – de Nationale Coalitie Digitale Duurzaamheid (NCDD) opgericht. De NCDD zal werken aan een technische en organisatorische infrastructuur voor digitale informatie in Nederland, waarbij samenwerking tussen alle betrokken partners hoge prioriteit heeft.²² Erfgoedinstellingen kunnen op de website van DEN (Digitaal Erfgoed Nederland) in hun reeks 'DE BASIS', waarin richtlijnen voor digitalisering worden gegeven, ook aanwijzingen voor digitale duurzaamheid vinden.²³

Zoals gezegd, digitale duurzaamheid is volop in beweging. Over alle puzzelstukjes is wel iets bekend en op internet terug te vinden. Zo langzamerhand passen ook steeds meer stukjes bij elkaar. Het zal nog wel enige tijd duren, maar de verwachting is dat we op termijn de puzzel compleet krijgen en het digitaal erfgoed voor komende generaties kunnen behouden.

Barbara Sierman studeerde Nederlandse letterkunde aan de Universiteit van Amsterdam. Zij begon haar loopbaan in bibliotheek-automatisering bij Pica (nu onderdeel van OCLC) en werkte daarna bij verschillende IT-bedrijven. In 2005 werd zij *digital preservation officer* bij de Koninklijke Bibliotheek (KB) in Den Haag (de Nationale Bibliotheek), waar zij nu teamleider van de afdeling Onderzoek Digitale Duurzaamheid is. Zij publiceert regelmatig over digitale duurzaamheid en over de resultaten van de internationale projecten waar de KB aan deelneemt, zoals Planets, Driver 2, ParseInsight en KEEP.

²¹ Zie noot 17.

²² Zie de website van de NCDD: www.ncdd.nl.

²³ Zie: http://wiki.den.nl/DE_BASIS_voor_duurzaamheid.

FARO. Vlaams steunpunt voor cultureel erfgoed vzw, Priemstraat 51, 1000 Brussel (<http://www.faronet.be>), geeft het bijgaande onderdeel van het werk *Erfgoed 2.0, Nieuwe perspectieven voor digitaal erfgoed* vrij voor verspreiding onder een Creative Commons Naamsvermelding-Geen Afgeleide werken 2.0 België Licentie. (1 januari 2010)



De Creative Commons Naamsvermelding-Geen Afgeleide werken 2.0 België Licentie is van toepassing op dit werk. Ga naar <http://creativecommons.org/licenses/by-nd/2.0/be/> of stuur een brief naar Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, VS om deze licentie te bekijken.

De gebruiker mag het werk kopiëren, verspreiden en doorgeven, onder de volgende voorwaarden:

- **Naamsvermelding:** de gebruiker dient bij het werk de door de maker of de licentiegever aangegeven naam te vermelden (maar niet zodanig dat de indruk gewekt wordt dat zij daarmee instemmen met uw werk of uw gebruik van het werk).
- **Geen afgeleide werken:** de gebruiker mag het werk niet bewerken.

Bij hergebruik of verspreiding dient de gebruiker de licentievoorwaarden van dit werk kenbaar te maken aan derden.

De gebruiker mag afstand doen van een of meerdere van deze voorwaarden met voorafgaande toestemming van de rechthebbende.

Niets in deze licentie strekt ertoe afbreuk te doen aan de morele rechten van de auteur, of deze te beperken.

Het voorgaande laat de wettelijke beperkingen op de intellectuele eigendomsrechten onverlet.

De tekst van de volledige licentie is beschikbaar op de website van Creative Commons:
<http://creativecommons.org/licenses/by-nd/2.0/be/legalcode.nl>