

Omvallende bitjes: ons digitaal geheugen bedreigd

Veel informatie wordt tegenwoordig digitaal geproduceerd en steeds meer analoge informatie wordt digitaal opgeslagen. Organisaties worden daardoor steeds afhankelijker van digitale technologie. Doordat opslagformaten over de jaren heen verschillen, versies van toepassingsprogrammatuur elkaar snel opvolgen en sommige dragers niet langer gebruikt worden, kan informatie die een aantal jaren geleden is geproduceerd nu al niet meer goed leesbaar zijn. Jan van der Starre inventariseert de beschikbare oplossingen.

HET PROBLEEM van de digitale duurzaamheid is het eenvoudigst te schetsen aan de hand van een paar voorbeelden.

Eind jaren tachtig begon Disney met materiaal digitaal te produceren en op te slaan. Begin jaren negentig beschikte men over een systeem om dat materiaal te beheren, terug te vinden en opnieuw te gebruiken. Het oorspronkelijke medium was 8mm tape. Al gauw kwam men tot het inzicht dat tape een niet zo betrouwbaar medium voor de toekomst is. Aan de stelregel 'Keep It Forever' kon niet worden voldaan. Men stapte over op cd-rom en op dit moment denkt men vooral aan beschrijfbare dvd's, vanwege hun veel grotere capaciteit. Heeft men daarmee 'Keep It Forever' veilig gesteld?

Een farmaceutisch onderzoekslaboratorium wilde de laboratorium-aantekenboeken beschikbaar stellen aan alle medewerkers. Aanvankelijk dacht men ze geheel elektronisch te maken. Al snel kwam men tot de conclusie dat de rechtsgeldigheid problematisch zou kunnen zijn in het geval van patent-claims. Belangrijker nog: de aantekenboeken worden aangevuld met data uit diverse meetinstrumenten. Men had grote twijfel of die data in elektronische vorm de tand des tijds zouden kunnen doorstaan. Dus werd er gekozen voor scanning van de aantekenboeken en het onderbrengen van die scans in een document management systeem. Is het probleem van duurzaamheid daarmee opgelost? Uiteraard niet, maar een groter probleem is voorkomen.

Aan het begin van de jaren tachtig heb ik mijn eerste pc aangeschaft, een 8086-machine met een dubbele 5¹/₄ inch floppy drive. Het operating system was Dos 3.1, als ik me goed herinner, en mijn tekstverwerker Wordstar 3. Als ik nu nog bestanden uit die tijd zou willen lezen heb ik in ieder geval een probleem: een 5¹/₄ inch floppy drive is nergens meer te vinden, behalve wellicht in het museum. Had ik nu maar alles overgezet op 3¹/₂ inch floppies, of zelfs op cd, dan had ik dat probleem niet gehad. Of toch? Wordstar 3

bestanden zijn voor de huidige tekstverwerkers niet of niet geheel te lezen. En 'niet geheel' is in dit verband ook heel vervelend, want ik weet niet welk gedeelte niet wordt gelezen. Kortom, mijn digitaal geheugen is onvolledig, op zijn minst. En dan spreek ik nog niet eens over exotica als CP/M-machines, de Atari 512 en de Commodore 64 met cassettetapes.

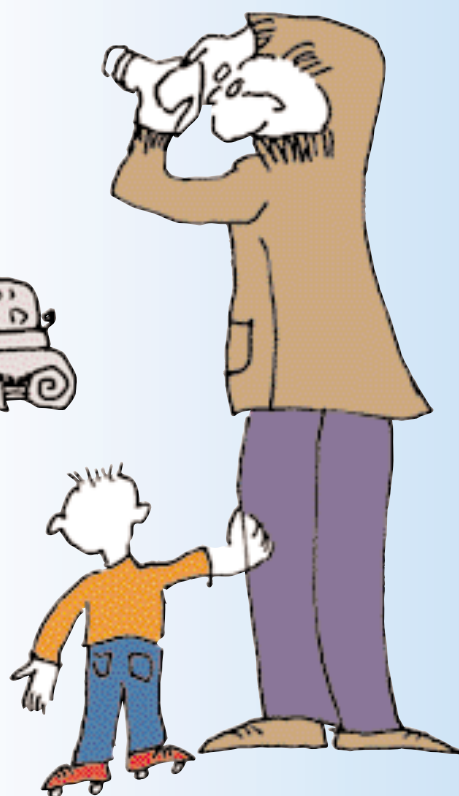
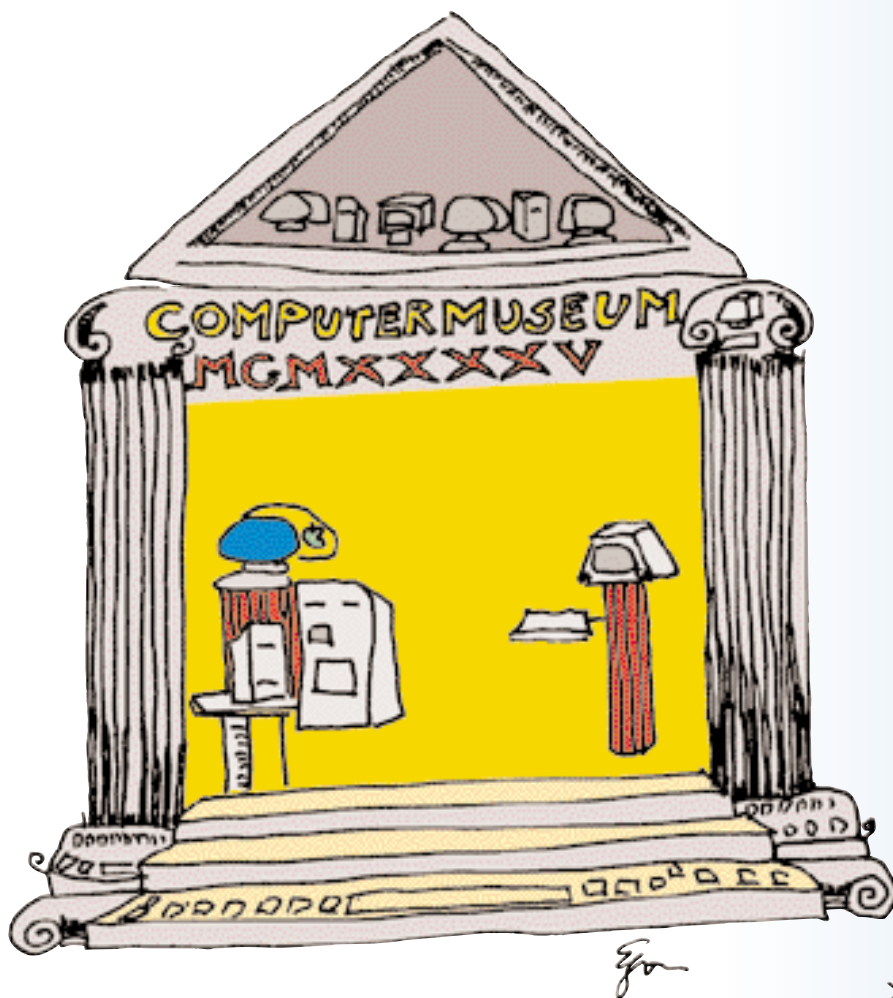
Als we het over duurzaamheid hebben, dan spreken we in het dagelijks leven al gauw over een termijn van tien tot vijftien jaar (wasmachines, koelkasten). Maar digitale duurzaamheid kent een veel langere tijdsspanne. Het archiefbeleid spreekt in dit verband van een termijn van honderd jaar leesbaarheid en toegankelijkheid.¹

Effecten

Over de effecten van bovengenoemde problemen kan in het geval van een privé-persoon wellicht nog worden gezegd: jammer, maar het is niet anders. Maar stel nu dat die persoon een belangrijk schrijver is? Een deel van onze cultuurhistorie verdwijnt op die manier. Of een wetenschapper? Wetenschapshistorie bedrijven wordt op die manier heel lastig.

Voor organisaties is het probleem nog veel ernstiger. De overheid, lokaal, provinciaal of centraal, bestaat bij de gratie van documenten, die een neerslag zijn van overheids-handelen en die gebruikt moeten kunnen worden bij de verantwoording van het gevoerde beleid. De overheid moet duidelijk kunnen maken hoe regelgeving tot stand is gekomen en hoe daar in de praktijk mee is omgegaan. Als steeds meer documenten digitaal gemaakt worden, en niet eens meer afgedrukt, dan gaat het bewaarprobleem een grote rol spelen. De overheid kan niet meer voldoen aan een van haar plichten: verantwoording.

Ook voor de particuliere sector is de beperkte houdbaarheid van digitale informatie een probleem, niet zozeer vanwege de verantwoordingsplicht, maar vanwege bewijs-



plicht. Laat het farmaceutische bedrijf maar bewijzen dat voor introductie van een nieuw medicijn voldoende en afdoende onderzoeken zijn verricht naar bijwerkingen en effecten. Als de onderzoeksrapporten en de onderliggende statistische gegevensverzamelingen niet meer of onvolledig leesbaar zijn, dan wordt het heel lastig om je tegen een schadeclaim te verdedigen.

Zowel voor de overheid als de particuliere sector speelt er nog iets. Verlies van digitaal geheugen betekent verlies van organisatiegeheugen. In een tijd waarin bij veel organisaties kennismanagement centraal staat, staat het onleesbaar worden van digitale documenten gelijk aan het wegrotten van de heipalen onder een huis: het zakt weg en stort uiteindelijk in.

Oplossingen

Kunnen we iets doen, hebben we een mogelijkheid om oudere digitale informatie leesbaar te maken, of huidige leesbaar te houden? Zowel aan de kant van de softwareproducenten als onder gebruikers is men druk doende oplossingen te verzinnen.

Allereerst zijn daar de technische oplossingen.

Afdrukken / instantiëren

Bij afdrukken of instantiëren wordt digitale informatie vastgelegd op papier of microfilm (bijvoorbeeld afdrukken

van een e-mailbericht en dat verder als archiefstuk behandelen). Het is echter geen echte oplossing voor het totale probleem, wellicht wel (voorlopig) voor onderdelen of in bepaalde situaties. Immers, digitale documenten verliezen hun eigenschappen en functionaliteit (bijvoorbeeld hypermedia), ze zijn niet langer machineleesbaar (en dus niet herbruikbaar), en documenten zijn niet langer terugzoekbaar op inhoudelijke kenmerken, tenzij die kenmerken opgenomen zijn in de beschrijving van een stuk.

Standaards

Het gebruik van standaards lijkt een oplossing te zijn voor digitale duurzaamheid. Immers, op het moment dat men weet dat een bepaald document gemaakt is volgens een bestaande standaard, is bekend hoe het is opgebouwd, en waar de relevante informatie staat. Hoewel een standaard

helpt, is het geen complete oplossing, maar hooguit een voor de korte termijn of als tussenoplossing. En nadelen zijn er ook:

- leveranciers wijken vaak af van afgesproken standaards (denk aan RDBM-systemen);
- standaards zijn niet permanent (bijvoorbeeld RDBM en OOdatabases). Wie kan mij garanderen dat huidige pdf-bestanden over honderd jaar nog leesbaar zijn (nog afgezien van de drager)?
- vertaling van de ene standaard naar een andere levert informatie- en functioneel verlies op;
- vertaling houdt tevens migratie in, hetgeen ook informatieverlies kan opleveren;
- niet alle documenten zijn onderworpen of houden zich aan standaards.

DD Scan

Wat betreft het inzetten van oplossingen: Cap Gemini Ernst & Young heeft met dit oogmerk een instrument vervaardigd om organisaties door te lichten op aspecten van digitale duurzaamheid. Deze zogenaamde DD Scan bestaat uit zes hoofdonderdelen:

1. **Bewustwording:** onderkent de organisatie het probleem van digitale duurzaamheid in voldoende mate? Vaak realiseert met name het management van een organisatie zich niet welke gevaren hen op dit terrein bedreigen, terwijl hun commitment absoluut noodzakelijk is om maatregelen te ontwerpen en te implementeren.
2. **Informatie- en functionele analyse:** welke gegevens zouden, gezien vanuit de bedrijfsdoelstellingen en -functies, in principe in aanmerking moeten komen voor langdurige bewaring? Hierbij gaat het niet alleen om elektronische documenten, maar ook om zaken als grote databases, spreadsheets, papier en microfilms.
3. **Wettelijke eisen, procedures en authenticiteit:** welke wetten zijn geldig (bijvoorbeeld de archiefwet), en welke elementen daaruit zijn van belang; welke procedures hanteert men binnen de organisatie ter waarborging van langdurige opslag en leiden die tot volledige verantwoording van keuzes/acties; welke garanties kunnen worden gegeven dat het beschikbare materiaal authentiek is.
4. **Techniek:** is de opslag en retrieval zowel op hardware- als software-gebied geregeld (wat moet worden bewaard, waar, hoe lang); in hoeverre is aanwezige hardware en software een risicofactor (houdbaarheid).
5. **Risico-analyse:** deze gaat in op zowel de organisatorische als de technische aspecten van digitale duurzaamheid. Aangegeven wordt op welke terreinen een organisatie zwak scoort en hoe dat moet worden beoordeeld. Daarbij zal een afweging gemaakt worden tussen kosten en mogelijke (im)materiële gevolgen.
6. **Oplossingsrichtingen:** op basis van de risico-analyse wordt aangegeven welke mogelijke oplossingen uit het hierboven beschreven scala toepasbaar zijn, indien gewenst voorzien van een indicatie van de kosten die daarmee gepaard gaan. In de praktijk zal het er vaak op neer komen dat – gezien de huidige stand van de techniek – een combinatie van oplossingen wordt voorgesteld.

Een van de meer recente standaards heeft wel goede papieren in het kader van digitale duurzaamheid: xml. Met behulp van xml-coderingen kan tekst worden gescheiden van opmaak, en ingedeeld naar functie. Dat levert voor 'gewone' documenten onleesbare teksten op die bol staan van de codetags, maar die kunnen worden herleid tot hun oorspronkelijk uiterlijk en bieden de mogelijkheid gecontroleerd in teksten te zoeken. De nieuwe office suite van Microsoft (XP) is voor een deel gebaseerd op xml-codering. Ook op xml gebaseerde database engines komen op de markt² en in theorie zou ook een plaatje middels xml kunnen worden gecodeerd. Of dat laatste in de praktijk haalbaar is, zal echter nog moeten blijken.

Computermusea

In computermusea worden machines op een zodanige manier bewaard, dat ze nog steeds kunnen worden gebruikt. Maar ook computermusea leveren op de langere termijn geen oplossing:

- het is onwaarschijnlijk dat machines het tot in het oneindige zullen doen, mechanische problemen zijn op den duur onvermijdelijk;
- toegang tot documenten is alleen mogelijk op een paar plaatsen;
- oudere computers kunnen geen nieuwe media lezen, terwijl die nieuwe media wel moeten worden gebruikt in verband met de houdbaarheid (conversie);
- computerchips desintegreren op termijn (fysiek verval);
- onderhoud van de machines is uitermate gespecialiseerd werk en derhalve kostbaar.

Migratie/conversie

Migratie is reeds jarenlang gebruikt als een vorm van pre-servering. Migratie (het overzetten van het ene naar het andere softwareplatform, van de ene versie van software naar de andere, en meestal in één adem genoemd met het verversen van de drager, conversie) zou een mogelijke oplossing kunnen zijn. Echter:

- het kost veel tijd (en dus geld);
- bij conversie gaat vaak informatie verloren, met het risico van corrupte data;
- conversie is niet eenmalig, maar komt regelmatig terug (dus ook de risico's);
- conversie moet altijd worden uitgevoerd op een totale set documenten.

Viewers

Met een viewer is het mogelijk oudere documenten te bekijken. Het is een programma dat in staat moet zijn alle functionaliteiten van een document te laten zien. Hoewel het een aantrekkelijke optie lijkt, zijn er toch een paar nadelen:

- het probleem van de informatiedragers wordt er niet mee opgelost;
- er ontstaat een afhankelijkheid van één programma (dat ook weer te oud kan worden);
- er is geen garantie dat alle functies van een document tot in lengte van jaren kunnen worden getoond;
- viewers werken naar alle waarschijnlijkheid alleen goed bij statische, lineaire documenten. Databases en dergelijke kunnen een probleem vormen.

Digitale archeologie

Bij het toepassen van digitale archeologie wordt ervan uitgegaan dat pas iets aan 'leesbaar maken' wordt gedaan op het moment dat dit nodig is. Het is daarmee de goedkoopste oplossing. Slechts de oorspronkelijke bitstream wordt bewaard. Indien er voldoende metadata worden bewaard die betrekking hebben op de structuur en inhoud van de bitstream is de kans van slagen wellicht groot. Aan de andere kant is er geen bewijs voor dat het zal lukken. Gezien de kosten zou dit een optie kunnen zijn voor materiaal dat slechts een jaar of tien à twintig bewaard moet blijven.

Bitstreams

Een broertje van digitale archeologie is het bewaren van bitstreams en tegelijkertijd het creëren van metadata die het decoderen mogelijk maken. Echter, het coderen van het gedrag van complexe software en documenten is bepaald geen sinecure, zal veel geld kosten en de validiteit is ook nog niet bewezen.

Emulatie

Emulatie bestaat uit een techniek waarbij documenten en de bijbehorende programmatuur worden opgeslagen, tezamen met een specificatie van het platform waarop het gedraaid heeft. Er zijn daarvoor drie sets informatie nodig:

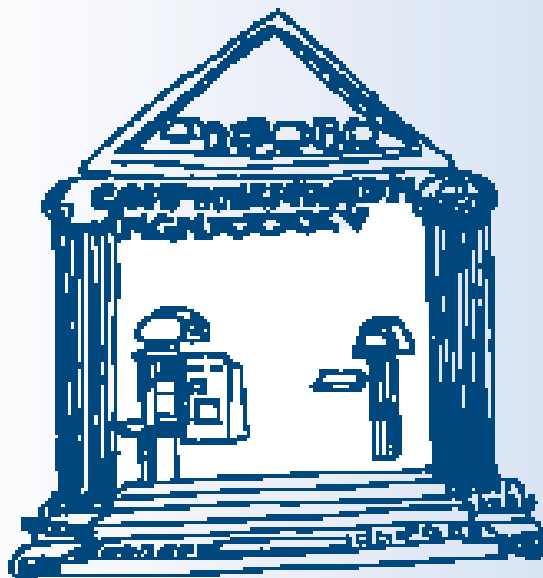
- het document zelf, met de toepassingsprogrammatuur en het operating system (alle als bitstream);
- een specificatie van de emulator voor de oorspronkelijke hardware-omgeving (nb: een specificatie, geen programma);
- metadata met betrekking tot het document, uitleg, labelinformatie etcetera. Een deel hiervan zal voor het menselijk oog leesbaar moeten zijn.

Uiteraard kan er worden gewerkt met centrale depots (bijvoorbeeld voor de emulatorspecificatie, operating system, toepassingssoftware). Deze techniek heeft zich in de praktijk nog niet echt bewezen, zij het wel op onderdelen (Apple Power PC die een 68000-processor kan emuleren). De vraag is dan ook of het een valide aanpak is. Bij de Koninklijke Bibliotheek in Den Haag worden enige tests uitgevoerd om dit vast te stellen.³

Organisatorische maatregelen

Naast de technische oplossingen is er sprake van organisatorische maatregelen om de levensduur te verlengen, of zeker te stellen dat wat in een digitaal archief zit ook echt het originele document is dat bewaard moet blijven. Men kan in dit verband denken aan beveiliging van bepaalde schijven tegen ongeautoriseerd gebruik, checksum-operaties om de authenticiteit van documenten te garanderen, het regelmatig maken van backups (uiteraard alleen valide voor bewaring op de korte termijn), en het opstellen en handhaven van richtlijnen ten aanzien van migratie en conversie.

Ook het vastleggen en bewaren van metadata is in dit verband van groot belang. Een pro-actieve houding ten aanzien van te verwerven en/of te bewaren informatie is hierbij eveneens van grote waarde. Immers, men kan beter niet verrast worden door het aanbod van een SPSS-bestand uit



de jaren tachtig, terwijl men van tevoren weet dat het eraan zal komen en wat de specificaties zijn. Voor het inventariseren van die specificaties/eigenschappen komen er ook gereedschappen beschikbaar.⁴

Kortom, een goed uitgedacht beheermechanisme kan problemen op de korte termijn voorkomen. Het moet wel worden gedragen door de organisatie en de organisatie moet zich ook realiseren dat het geld kost. Op de langere termijn is een dergelijk beheermechanisme echter weliswaar een zeer belangrijke randvoorwaarde, maar niet een oplossing op zichzelf.

Hoe inzetten?

Hoe kunnen nu de verschillende oplossingen worden ingezet? Dat kan in feite alleen op basis van drie pijlers:

- weten wat je hebt;
- weten hoe je het moet waarderen;
- weten hoe er in werkprocessen mee wordt omgegaan.

Het komt altijd neer op het maken van een afweging tussen wat haalbaar en wat noodzakelijk is, gebaseerd op kennis van wat er is, wat *moet* worden bewaard, wat *kan* worden bewaard, en de relatieve waarde van het materiaal. Niemand kan een garantie geven voor het 'Keep It Forever', maar we kunnen wel maatregelen nemen om de kans zo groot mogelijk te maken. Alleen op die manier kan worden voorkomen dat bitjes die ertoe doen omvallen.

Noten

1. De zorgdrager treft zodanige voorzieningen ten aanzien van de door hem opgemaakte archiefbescheiden die ingevolge een voor hem geldende selectielijst voor bewaring in aanmerking komen, dat bij het raadplegen van die archiefbescheiden na ten minste honderd jaar geen noemenswaardige achteruitgang zal zijn te constateren. (Archiefbesluit 1995, artikel 11, lid 1).
2. Zie bijvoorbeeld www.softwareag.com/tamino/.
3. Zie www.geheugenvannederland.nl/.
4. Volgens het pakket EDDA van Decos Software (www.decos.nl/).

Jan van der Starre is consultant bij Cap Gemini Ernst & Young, practice Documentaire Informatiesystemen en Workflow Management. Hij is lid van de themagroep Digitale Duurzaamheid van de Record Management Conventie.