

Digitale bronnen archiveren

Filip Boudrez
Expertisecentrum DAVID vzw
Antwerpen, 2008

INLEIDING

Digitaal is vandaag de dag de norm¹. Iedereen en alles gaat digitaal. Papieren en analoge informatie zijn niet meer van de ééentwintigste eeuw en moet zoveel mogelijk uit onze kantoren en huiskamers worden gebannen. Onze werkgevers voeren het cleandesk-principe en het maximaal digitaal werken in. Thuis moeten fotoalbums en vinylplaten plaats ruimen voor JPEG- en MP3-archieven op DVD of (externe) harde schijven. En binnenkort kijken we allemaal digitale televisie.

Ook onze archiefschatten ontsnappen niet aan de digitaliseringstendens. Enerzijds worden grote delen van ons documentair erfgoed met terugwerkende kracht systematisch gedigitaliseerd. Daarvan getuigen de talrijke beeldbanken die de voorbije jaren als paddenstoelen uit de grond schoten. Anderzijds worden de toekomstige archiefschatten in toenemende mate digitaal geboren. Erfgoedbeherende instellingen spelen hierop in. Het toepassingsgebied van het wettelijk depot in België werd recent uitgebreid met digitale opslagmedia. Diverse archiefinstellingen verwerven al 'digital-born' archiefdocumenten.

De gedigitaliseerde en digitale archiefschatten zijn bestemd om voortaan een digitaal bestaan te leiden en in de toekomst als bronnenmateriaal te worden gebruikt. Dat lijkt vanzelfsprekend, maar is minder evident dan het lijkt. Met name op het vlak van de leesbaarheid en de authenticiteit van digitale bronnen vergt archivering een actieve en permanente zorg. Dit is niet alleen efficiënt en kostenbesparend, maar voor digitale documenten ook een noodzakelijke voorwaarde wil men op termijn nog over leesbare en authentieke bronnen beschikken.

LEESBAARHEID

Digitale informatie overleeft minder gemakkelijk de tand des tijds dan haar analoge tegenhangers. De redenen hiervoor hoeven we niet ver te zoeken. Vernieuwing en innovatie zijn de drijvende motor van de informatica-industrie, maar ze zijn ook verantwoordelijk voor een snellere technologische veroudering dan in de analoge wereld. Die fundamentele contradictie schuilt in elk digitaliseringsverhaal, zeker daar waar het ons erfgoed betreft. Naast een grote verspreiding via het web en een betere toegankelijkheid, is het ultieme argument voor digitalisering de duurzame preservatie van analogo materiaal dat kwetsbaar is, dat onderhevig is aan gebruiksslijtage of dat van nature uit degenereert. In de algemene perceptie is digitalisering de reddingsoperatie die niet snel genoeg kan worden gestart, want digitale informatie kan onbepaald worden gekopieerd en worden gebruikt zonder dat de kwaliteit van de informatie er onder gaat leiden. De vernieuwingscyclus, en dus ook de verouderingscyclus, van digitale opslagmedia, bestandsformaten en toepassingen is echter vele malen groter dan die van analogo materiaal. De verschillende generaties digitale opslagsystemen volgen elkaar niet alleen sneller op, de impact van het generatieverlies dat kan optreden bij conversie en/of migratie is ook opmerkelijk hoger dan in de analoge wereld.

De grotere kwetsbaarheid van digitale informatie uit zich ook in de risico's die zich voordoen bij de verschillende facetten van digitale archivering. Een betrouwbare gegevensopslag alleen al vraagt een goede beheersprocedure waarbij idealiter permanent en systematisch aan

¹ Dit artikel verscheen onder de titel 'Digitale archiveren' in: *Karakter. Tijdschrift van wetenschap*, nummer 23, 2008, p. 19-21.

kwaliteitswaarborging wordt gedaan. Iedere conversie- of migratieslag ter bewaring van de leesbaarheid is een risicovol moment, dat bijgevolg zorgvuldig moet worden gepland en gecontroleerd. Het alternatief waarbij de computerbestanden niet worden gewijzigd, maar de vereiste hard- en softwarecomponenten op nieuwe platformen worden geëmuleerd, is evenmin vrij van risico's en afhankelijkheden.

Het is vanwege deze redenen niet denkbeeldig dat digitale bronnen wel eens een korter leven kunnen beschoren zijn dan analoge. En dat terwijl men algemeen verwacht dat bij digitalisering de digitale kopieën ons in het ultieme geval behoeden van definitief informatie- of bronnenverlies. Het is echter de vraag of hierover wordt nagedacht vooraleer grote volumes foto's, prentbriefkaarten en stadszichten massaal op de scanner worden gelegd. Dat de originele analoge bronnen bij de meeste cultuurhistorische digitaliseringsprojecten na scanning niet worden vernietigd, is geruststellend voor het geval een nieuwe digitalisering zich zou opdringen. In een administratieve context gaat digitalisering wel meer en meer samen met substitutie. Bij digital-born materiaal is dit opvangnet er dus niet. De situatie is dikwijls ook problematisch omdat veel digitaliseringsinitiatieven louter projectmatig worden aangepakt zonder enig perspectief op continuïteit. De afgeleide digitale raadplegingskopieën worden dan wel bijgehouden op geback-upte en gemonitorde servers, maar de digitale masters verdwijnen maar al te vaak in een schuif of een kast zonder dat ze het onderwerp zijn van de actieve preservering die digitale archivering vraagt.

Het zou natuurlijk beter zijn dat de digitalisering pas wordt gestart wanneer de organisatie de vereiste kwaliteitscontroles en procedures voor de (middel-)langetermijnarchivering van digitaal materiaal operationeel heeft en dat deze volledig in de digitaliseringsworkflow zijn ingebed. Idealiter wordt met digitalisering pas gestart wanneer in functie van een digitale bewaarstrategie meteen de juiste formaatprofielen worden toegepast en de vereiste metadata in een gedocumenteerde en archiveerbare toepassing worden geregistreerd. De archiefbeherende instellingen die eerst een bewaar- en archiveringsstrategie, inclusief een hard- en softwarearchitectuur, uitwerken voor ze met digitalisering starten, zijn echter op één hand te tellen. De langetermijnarchivering van digital-born materiaal is voor de meeste instellingen nog een open vraag of in het beste geval een 'work-in-progress'. Hier en daar gaan zelfs stemmen op om digitale informatie bestemd voor (middel-)langetermijnarchivering af te drukken of analog te bewaren. De wereld op zijn kop!

AUTHENTICITEIT

De toekomstige archiefgebruiker verwacht van zijn digitale bronnen niet alleen dat ze leesbaar en van een hoge kwaliteit zijn, maar ook dat ze authentiek en betrouwbaar zijn. Hun potentiële bruikbaarheid voor toekomstige onderzoek zal onder meer afhangen van de mate waarin authenticiteit en betrouwbaarheid kunnen worden gedemonstreerd. Vooral voor digital-born materiaal is dit een grote uitdaging. Bij digitalisering wordt substitutie nog maar in beperkte mate toegepast, zodat de originele documenten nog steeds als referentiemateriaal beschikbaar zijn.

Digitaal-zijn biedt het grote voordeel dat documenten gemakkelijk herbruikbaar zijn of kunnen worden aangepast. Hierdoor ontstaat echter een groter en fundamenteel wantrouwen ten aanzien van de betrouwbaarheid van die digitale bronnen. Digitale documenten zijn gemakkelijk aan te passen zonder dat dit achteraf gemakkelijk vast te stellen is. Daar waar in de analoge wereld algemeen wordt aangenomen dat de onbetrouwbaarheid van een document moet worden bewezen, gaat men er in de digitale wereld eerder van uit dat de betrouwbaarheid moet worden gedemonstreerd. Authenticatie of het bewijzen van de authenticiteit van digitale documenten is mogelijk, maar alleen als proactief de nodige voorzieningen hiervoor worden getroffen.

Hoe dit authenticatieproces precies moet verlopen, is momenteel nog een punt van discussie. Uiteindelijk moeten digitale bronnen beveiligd zijn zodat ze gevrijwaard blijven van onrechtmatige wijzigingen of manipulaties. Daarnaast moet hun authenticiteit bewijsbaar zijn. Diverse disciplines

buigen zich over deze problematiek, maar een algemene consensus over de geschikte oplossing is er nog niet. Zo schuiven juristen en informatici de geavanceerde digitale handtekening als authenticiteitsgarantie voor digitale informatie naar voren, maar voor archivariissen volstaat die puur technologische oplossing niet. Digitale handtekeningen zijn immers wel als digitaal object duurzaam te archiveren, maar hun validatiefunctie vervalt na verloop van tijd. Die validatiefunctie dient overigens enkel om aan te tonen dat een document achteraf al dan niet werd gewijzigd. Met een geavanceerde digitale handtekening voorkom je geen wijzigingen en kan evenmin worden opgespoord wie wat wanneer wijzigde. Andere beveiligingsmaatregelen ter preventie van wijzigingen zoals paswoorden of encryptie staan dan weer haaks op de uitgangspunten van digitale archivering waarbij afhankelijkheden en reconstructieschakels zoveel mogelijk moeten worden vermeden.

Voor archivariissen is het bewaren van de authenticiteit van digitale documenten niet louter een zaak van informatiebeveiliging. Voornamelijk gebaseerd op de diplomatiek of oorkondeleer beklemtoont de archiefwetenschap dat de authenticiteit van documenten vanuit de herkomst en de identiteit van de digitale documenten moet worden geanalyseerd. Een digitale handtekening communiceert hier echter niets over, en zorgt er evenmin voor dat die belangrijke contextgegevens worden geregistreerd of op een onlosmakelijke wijze met de documenten blijven geassocieerd. Het archiveren van authentieke digitale bronnen is pas mogelijk als de informatiesystemen waarin onze eigentijdse archiefschatten worden aangemaakt en beheerd specifiek daartoe zijn ingericht. Functionaliteiten zoals het registreren van contextuele metadata, het fixeren van de inhoud van documenten, het bijhouden van verschillende versies van documenten en hun metadata, het aanleggen van audit-trails, enz. zijn essentieel wil men authentieke bronnen archiveren. Men mag immers niet vergeten dat digitaal archiveren per slot van rekening een reconstructieproces is. Op het tijdstip van raadpleging worden digitale documenten op basis van de beschikbare gegevens en parameters opnieuw opgebouwd. Dit is eigen aan het digitaal-zijn, terwijl elke archiefgebruiker verwacht dat alle componenten van zijn bronnen statisch en gefixeerd zijn en niet meer wijzigen. Dit betekent dat alle componenten expliciet moeten zijn vastgelegd.

Aan deze aspecten van authenticiteit wordt tot op vandaag maar weinig aandacht besteed bij de implementatie van nieuwe bedrijfstoepassingen. Archiefaudits wijzen uit dat bestaande en nieuwe informatiesystemen wel een massale hoeveelheid aan digitale informatie bevatten, maar lang geen authentieke informatie. Dit is problematisch vanuit het standpunt van de bedrijfsvoering van organisaties, maar heeft ook als gevolg dat er op termijn geen authentieke digitale bronnen kunnen worden gearchiveerd. Met terugwerkende kracht authentieke digitale archiefdocumenten creëren, is immers niet mogelijk. De mate van authenticiteit en betrouwbaarheid kan bij archivering wel worden gehandhaafd, maar kan niet worden vergroot. Digitale archivering betekent in veel business cases wel een explicitering van de contextgegevens op het archiveringsmoment, maar dit kan pas wanneer een minimum aan gegevens op het moment van creatie, ontvangst of gebruik zijn vastgelegd.

Het digitaal archiveren van leesbare en authentieke informatie is perfect mogelijk wanneer met een aantal factoren rekening wordt gehouden op het tijdstip van digitaliseren of bij het implementeren van informatiesystemen. Diverse (inter)nationale onderzoeksprojecten en praktijkervaringen inzake digitale archivering wijzen dit uit. De essentiële gegevens van of over digitale documenten zullen pas aan de documenten worden toegevoegd of in hun meta-informatie worden opgenomen, als de informatiesystemen daartoe zijn ingericht.

DE DIGITALE ARCHIVARIS

De binnenlandse en buitenlandse ervaringen van de voorbije decennia leren dat de zorg voor de creatie en het beheer van leesbare en authentieke archiefdocumenten niet spontaan door proceseigenaars, informatici of de software-industrie wordt opgenomen. Archivering wordt hiervoor nog te veel geassocieerd met een nieuwe bestemming voor informatie die zijn administratief nut verloren heeft. Dit is geen wenselijke situatie, noch voor papieren informatie, noch voor digitale

informatie, maar de consequenties en het dreigende informatieverlies is manifester waar het digitale bronnen betreft. De software-industrie voelt zich evenmin geroepen om dit probleem uit zichzelf aan te pakken of hierop te anticiperen. De voorbije jaren is er misschien wel een tendens om digitale documenten in gedocumenteerde en open formaten te bewaren, de essentiële proces- en contextgegevens om documenten te interpreteren, blijven doorgaans in een toepassingsafhankelijk formaat of datamodel bewaard. Vele organisaties zijn zich niet bewust van de risico's en wapenen zich dan ook niet tegen een mogelijke vendor lock-in, terwijl het nochtans hun informatie betreft. Hierdoor lopen de meeste organisaties grote risico's.

Hier is een belangrijke taak weggelegd voor archivariissen. Zij moeten een prominente rol spelen in het creatie- en beheersproces van digitale informatie. De positie van archivaris als ontvanger, beheerder en beschrijver van informatie die zijn primair nut verloren heeft, moet deels worden verlaten. In de plaats daarvan moet hij of zij een actieve speler zijn in het records management proces en als informatiearchitect het records continuüm vorm geven. Dit vraagt natuurlijk andere competenties dan het inventariseren en beschikbaar stellen van informatie met erfgoedstatus. Een grote kennis van de informatietechnologie is een must. Inzicht in het gegevensbeheer van digitale bedrijfsapplicaties en een vertrouwdheid met normen en standaarden zijn belangrijk. Gelet op de snelle evolutie in de IT-sector is specialisering in dit vakgebied geen overbodige luxe, wel integendeel.

Ondertussen is digitale archivering al meer dan een decennium het onderwerp van allerlei studies en onderzoeken. Talloze onderzoeksprojecten hebben geleid tot allerlei best practices, richtlijnen en aanbevelingen op het vlak van digitaal documentbeheer en digitale archivering. Ook op technologisch vlak zijn de bouwstenen voor goede digitale archiveringsoplossingen voorhanden. Mits inbedding in duurzame en proactieve procedures ligt hier de sleutel voor een goed uitgebouwd digitaal archiveringsbeleid. Alleen blijft een toepassing in de praktijk nog in te veel gevallen achterwege. Hierdoor dreigt op termijn digitale amnesie en een tekort aan leesbaar en authentiek bronnenmateriaal. En dit ondanks de vloedgolf aan informatie waarmee we vandaag worden geconfronteerd.

BESLUIT

In tegenstelling tot de papieren en analoge informatie vraagt digitale archivering een heel (pro-)actief archiverings- en preserveringsbeleid. Proactieve archivering is niet alleen efficiënt en kostenbesparend, maar voor digitale documenten een noodzakelijke voorwaarde wil men op termijn over leesbare en authentieke bronnen kunnen beschikken. Zo niet zal veel digitale informatie onherroepelijk verloren gaan of niet meer bruikbaar zijn. Archivariissen kunnen deze verantwoordelijkheid opnemen, maar dienen dan een sturende rol in het archiefvormingsproces op te nemen.