



DAVID

Het digitaal archiveringssysteem:
beheersinventaris, informatielagen en beslissingsmodel
als uitgangspunt

Filip Boudrez



FACULTEIT RECHTSGELEERDHEID
INTERDISCIPLINAIR CENTRUM VOOR RECHT EN
INFORMATICA
TIENSESTRAAT 41
B-3000 LEUVEN



Stadsarchief
Stad Antwerpen

Versie 1.0

Wettelijk depot D/2001/9.213/3

Antwerpen, juni 2001

Website DAVID-project: <http://www.antwerpen.be/david>

E-mailadres: david@stad.antwerpen.be

INHOUDSTAFEL

I. INLEIDING	4
II. DE ARCHIVERINGSSTRATEGIEËN	7
II.1 <i>Hard copy</i>	7
II.2 <i>Migratie</i>	8
II.3 <i>Bewaring van de technologie</i>	10
III. BESLISSINGSMODEL VOOR HET VASTLEGGEN VAN DE ARCHIVERINGSSTRATEGIE.....	12
III.1. <i>Beheersinventaris van de digitale informatiesystemen</i>	13
III.2. <i>Informatielagen als uitgangspunt</i>	14
III.3. <i>Beslissingsmodel</i>	16
III.3.1 WAT wordt gearhiveerd ?.....	17
III.3.2 WIE beheert het digitaal archief ?.....	19
III.3.3 HOE archiveren we digitale archiefbescheiden?.....	22
III.3.4 WANNEER dragen we het digitaal archief over ?.....	26
III.3.5 TOEPASSING: Wat? Hoe? Wanneer? Wie?.....	27
IV. BESLUIT: ARCHIVEREN HERBEKEKEN	28
BIJLAGE 1: MODELFIGE INFORMATIESYSTEEM	30
BIJLAGE 2: CHECKLIST DIGITAAL ARCHIVEREN	32

I. INLEIDING

Eén van de eerste doelstellingen van het DAVID-project¹ is de opsporing en inventarisatie van de digitale archiefbescheiden die door Vlaamse administraties en archiefinstellingen worden beheerd. Een dergelijke inventaris verschaft ons inzicht in de brede waaier van digitale bestanden en informatiesystemen die door de overheid worden bijgehouden en die binnen afzienbare tijd (digitaal) worden gearhiveerd. Om een beeld te vormen welke digitale overheidsarchieven er momenteel bestaan, werd onderzocht welke geïnformatiseerde toepassingen bij de administraties van het stadsbestuur, het OCMW en het havenbedrijf van Antwerpen lopen.

De initiële bedoeling was om vervolgens een typologie van digitale archiefbescheiden uit te werken op basis waarvan voor elk type de archiveringswijze op lange termijn kan worden afgeleid. Het basiscriterium voor de typologie is de bruikbaarheid voor het vastleggen van een archiveringsstrategie. De functie van de typologie werd op dit doel afgestemd. Een eerste poging vertrok vanuit de redactionele vorm en de functie van digitale archiefdocumenten², een beschrijvings- en indelingsmethode die werd ontleend aan de papieren archiefbescheiden. Al vlug werd echter duidelijk dat dit uitgangspunt niet als basis voor digitaal archiefbeheer geschikt is, en dus ook niet voor het vastleggen van archiveringsstrategieën.

Vooreerst zijn er een aantal algemene opmerkingen tegen het benaderen van digitale archiefdocumenten vanuit het perspectief van papieren documenten. Om te beginnen gaat deze werkwijze ervan uit dat digitale archiefbescheiden als (afzonderlijke) documenten worden bijgehouden, terwijl de inhoud van elk digitaal bestand eigenlijk gewoon in de datastream van het informatiesysteem wordt bewaard. De (redactionele) vorm op scherm of papier wordt uiteindelijk bepaald door de structurele informatie en de gebruikte applicatie (zie verder). De (redactionele) vorm van digitale archiefbescheiden ligt niet altijd vast en kan zeer verscheiden zijn. Hetzelfde digitaal archiefbescheid kan in een (combinatie van) tekst-, spreadsheet-, databankbestanden worden opgeslagen. Bovendien bestaan er veel informatiesystemen³ die enkel gericht zijn op het beheren van informatie zonder dat er documenten worden gecreëerd. Ten slotte bevatten digitale informatiesystemen meer informatie dan wat in het geheel van af te drukken documenten staat. Met een documentgerichte aanpak zou daarenboven veel contextuele informatie verloren gaan, terwijl net deze gegevens belangrijk zijn voor het verzekeren van de authenticiteit van digitale archiefdocumenten.

In een documentgerichte benadering richt het digitale archiveringsbeleid zich al heel vlug op de problematiek van de bestandsformaten. Het is niet aangewezen om enkel en alleen op deze basis een

¹ Digitale Archivering in Vlaamse Instellingen en Diensten (DAVID) wordt gefinancierd door het Fonds voor Wetenschappelijk Onderzoek-Vlaanderen in het kader van het Max Wildiersfonds.

² Bewijsleverend, regelgevend, verslaggevend, berichtgevend, geheugensteunend, planning en toegangevend (*Lexicon van Nederlandse archiefvormen*, 1983).

³ “Een informatiesysteem is het geheel van gegevens, programmatuur, procedures en apparatuur dat in staat is gegevens te transformeren in informatie” (P. HORSMAN, *Archiefsystemen en kwaliteit*, in *Naar een nieuw paradigma in de archivaliek*, p. 88). Informatiesysteem wordt in dit rapport gebruikt in de betekenis van het geheel van data, structurele informatie en applicatiesoftware die binnen één systeem worden gebruikt (zie p. 15). Informatiesystemen zetten data om in informatie. Document wordt hier als papieren gegevensdrager gebruikt. Digitale archiefbestanddelen worden aangeduid met digitale archiefdocumenten.

digitaal archiveringsbeleid vast te leggen, want hierdoor wordt al voor een groot stuk aangewezen dat voor de archivering op lange termijn de migratiestrategie wordt toegepast. In een dergelijke aanpak schuilt het grote gevaar dat geen of weinig aandacht wordt besteed aan het informatiesysteem waarbinnen de bestanden worden gecreëerd. Door een archiveringsstrategie direct te koppelen aan een bestandsformaat heeft men geen oog voor de architectuur, het functioneren en de functionaliteiten van het informatiesysteem. Er wordt evenmin rekening gehouden met een mogelijke interactie tussen digitale bestanden of systemen. Door zich te focussen op het bestandsformaat waarin de digitale gegevens worden opgeslagen, is de kans groot dat afhankelijkheden van bepaalde software, dynamische componenten of externe informatiebronnen over het hoofd worden gezien. Nochtans neemt het aandeel van dergelijke dynamische en geïntegreerde informatiesystemen de laatste jaren alsnog toe. Een beslissingsmodel louter gebaseerd op bestandsformaten kan ook omwille van een aantal technische redenen worden afgewezen. Net zoals bij applicaties bestaan er van hetzelfde bestandsformaat verschillende versies die onderling van elkaar verschillen⁴. Bestandsformaten worden soms ook frequent inwendig aangepast aan het operating system of de applicatie waarbinnen ze worden gebruikt⁵. Bestandsformaten kunnen evenmin als een stabiele factor worden beschouwd, want daarvoor spelen commerciële factoren soms een te grote rol⁶. Deze werkmethode levert dus noch een systematisch overzicht, noch een stabiele basis voor het uitwerken van archiveringsstrategieën op⁷.

Ondertussen wordt wel duidelijk dat men als uitgangspunt voor digitale archivering een hoger niveau moet hanteren, met name dat van het informatiesysteem waarbinnen de digitale

⁴ Dit is onder meer het geval bij de bestandsformaten van Word97 en Word2000. Hetzelfde tekstbestand is perfect uitwisselbaar tussen beide versies van Word, waardoor men zou kunnen vermoeden dat het bestandsformaat van Word97 en Word2000 hetzelfde is. De manier waarop beide Wordversies hun bestanden samenstellen verschilt echter grondig van elkaar. Dit blijkt onder andere uit het verschil in bestandsgrootte wanneer men hetzelfde tekstbestand als een Word97- of een Word2000 bestand opslaat. Een ander gekend voorbeeld van deze problematiek is de evolutie van de grafische GIF-bestanden. Of men te maken heeft met een GIF-bestand met (GIF89a) of zonder animatie (GIF87) kan eigenlijk alleen maar achterhaald worden door het bestand in een webbrowser of viewer te openen. In de meeste gevallen zal ook de bestandsomvang een indicatie zijn (GIF87 < GIF89a).

⁵ TIFF-bestanden worden intern frequent aangepast aan één specifieke softwareapplicatie. Hierdoor bestaan er heel veel verschillende versies van TIFF-bestanden die heel vaak slechts met één bepaalde applicatie op het scherm kunnen weergegeven worden.

⁶ Het gebruik van het Graphics Interchange Format (GIF) is hier een mooie illustratie van. GIF werd in 1987 door CompuServe als een vrije en open specificatie ter beschikking gesteld. Voor de (de)compressie van de grafische data maakte het GIFformaat gebruik van het Lempel-Ziv en Welch algoritme. Hierdoor worden afbeeldingen in kleine bestanden opgeslagen, waardoor de GIF-bestanden gemakkelijk en snel kunnen worden uitgewisseld. GIF werd een standaard. CompuServe verkeerde in de mening dat het LZW-algoritme tot het publiek domein behoorde. De problemen rezen wanneer Unisys zijn patentrechten op het LZW (de)compressie-algoritme liet gelden, waardoor elke gebruiker tot eind 2003 een licentievergoeding aan Unisys moet betalen. Dit betekent dat elke archiefdienst die afbeeldingen naar GIF-bestanden migreert, in principe een licentie moet aanvragen. Uit ongenoegen met deze gang van zaken werd PNG (Portable Network Graphics; vastgelegd door W3C) gecreëerd, de opvolger van GIF die vrij is van patentrechten.

⁷ De voorbije jaren werden al enkele onderzoeken gevoerd naar de moeilijkheden die gepaard gaan met de archivering van bepaalde types bestandsformaten. Zie hiervoor: G.W. LAWRENCE, W.R. KEHOE, O.Y. RIEGER, W.H. WALTERS en A.R. KENNEY, *Risk Management of Digital Information: a file format investigation*, Washington, 2000; J.C. BENNETT, *A framework of data types and formats, and issues affecting the long term preservation of digital material*, Wetherby, 1997. Deze onderzoeken beperkten zich tot het opsommen van de voornaamste aandachtspunten bij de digitale archivering van een aantal types bestandsformaten. In *Comparison of Methods & Costs of Digital Preservation* gaat consultant T. Hendley wel een stap verder. Hij onderscheidt 10 digitale bronnen en geeft per type bron de bestandsformaten op waarnaar ze moeten gemigreerd worden. Enkel in het geval van multimediatoepassingen reikt hij emulatie als oplossing aan.

archiefdocumenten worden gevormd. Informatiesystemen zou men kunnen indelen op basis van de types bestanden die ze genereren of die er de basis van vormen (tekst-, spreadsheet-, databank-, en audiovisuele bestanden). Maar net zoals bij een documentgerichte aanpak kan voor de archivering niet zomaar van het type informatiesysteem worden uitgegaan, zonder dat opnieuw de hier boven vermelde eigenschappen van elk informatiesysteem worden onderzocht. In die zin is het opnieuw onmogelijk om zo maar een archiveringsstrategie aan één bepaald type informatiesysteem koppelen, ook al omdat in de praktijk de meeste systemen uit meerdere types bestaan en een aantal unieke kenmerken hebben. Dit vraagt een voorafgaande en grondige analyse van elk systeem die in veel gevallen zal leiden tot een systeemspecifieke oplossing. Er moet dus steeds vanuit elk informatiesysteem worden vertrokken. Zo valt bij de informatiesystemen van de stadsadministratie, het OCMW en het havenbestuur van Antwerpen het aandeel op van de informatiesystemen die hun data in de vorm van mainframedatabanken bewaren en ad hoc ontwikkelde software gebruiken. Of het volstaat om de mainframedatabanken als flat files te archiveren zoals algemeen wordt aangegeven, wordt pas duidelijk na het doorlichten van het volledige systeem. Dezelfde redenering geldt eigenlijk ook voor eenvoudige informaticatoepassingen zoals gewone tekst- en spreadsheetbestanden.

Met het oog op het uitwerken van een systeem voor het bepalen van de archiveringsstrategieën komt men ook op basis van een typologie van de informatiesystemen geen stap verder. Om vanuit de informatiesystemen zelf te kunnen afleiden welke de meest geschikte archiveringsstrategie is, kan daarentegen wel een beslissingsmodel worden gevolgd. Dit beslissingsmodel is gebaseerd op het terugkerend patroon van de factoren die bij elk informatiesysteem moeten worden onderzocht: WAT wordt gearchiveerd? WIE beheert het digitaal archief? HOE dragen we het digitaal archief over? WANNEER vindt de overdracht plaats?

In dit rapport wordt het beslissingsmodel voorgesteld dat rond deze vier vragen is opgebouwd⁸. Bij elke vraag worden kort de factoren beschreven waarmee men rekening moet houden. Het resultaat is een beslissingsmodel dat kan worden gebruikt voor het systematisch onderzoeken van de informatiesystemen. De volgorde van deze vragen is niet louter willekeurig. De vraag WAT wordt gearchiveerd, vertrekt het best uit de benadering van elk informatiesysteem als een samenstelling van meerdere informatielagen (zie figuur 2, p. 15). Welke lagen op lange termijn worden bewaard, geeft al in grote mate de richting van de archiveringsstrategie aan. De vraag aan WIE het digitaal archief wordt toevertrouwd, hangt hiermee nauw samen. De vragen HOE en WANNEER hebben al meer betrekking op de praktische uitvoering van de archivering. Dit beslissingsmodel komt in het tweede deel van dit rapport aan bod en bevat de aanzet voor het uitbouwen van een digitaal archiveringssysteem. Voorafgaand worden kort de drie archiveringsstrategieën voor digitale archiefbescheiden toegelicht.

Hoewel zijdelings elementen zoals recordness, selectie, metadata, beschrijven en ter beschikking stelling aan bod komen, is dit beslissingsmodel in de eerste plaats gericht op de allereerste uitdaging voor een archiveringssysteem: het leesbaar houden en beheren van digitale archiefbescheiden. Omwille van het leesbaar houden en het verzekeren van de authenticiteit van digitale archiefbescheiden zal het niet volstaan dat enkel computerbestanden worden gearchiveerd. Naast de bestanden zelf moeten ook emulaties, metadata, logfiles, scripts, context, beschrijvingen van procedures, enz. mee worden gearchiveerd. Dit valt echter buiten de scope van dit rapport en komt in een volgend DAVID-rapport over authenticiteit aan bod.

⁸ Met dank aan Inge Schoups en Willem Vanneste voor de correcties en suggesties.

II. DE ARCHIVERINGSSTRATEGIEËN

De kern van het probleem inzake lange termijn archiveren van digitale archiefbescheiden wordt veroorzaakt door het verouderen of het in onbruik raken van de vereiste hard-en softwareomgeving. Digitale informatie heeft nu éénmaal de eigenschap dat ze enkel door middel van IT-technologie raadpleegbaar en dat deze technologie voortdurend in evolutie is. Als archivariissen en records managers niet actief optreden, worden ze geconfronteerd met digitale archiefbescheiden die met meer recente IT-technologie niet meer consulteerbaar zijn. Er kunnen in het algemeen drie archiveringsstrategieën voor digitale archiefdocumenten worden onderscheiden⁹: hard copy, migratie en bewaring van de technologie (o.a. computermusea en emulatie). Over hard copy en computermusea is men het eens dat ze enkel in noodscenario's een rol kunnen spelen, waardoor voor het ogenblik migratie en emulatie de enige opties lijken te zijn. Over de bruikbaarheid van migratie en emulatie is men het voorlopig nog oneens. Migratie richt zich op het aanpassen van de gearchiveerde digitale bestanden aan de nieuwe technologie, terwijl emulatie tot doel heeft de oorspronkelijke technologie binnen de nieuwe omgeving te simuleren zodat men de gearchiveerde digitale bestanden in hun origineel formaat kan raadplegen.

II.1 HARD COPY

In het geval van hard copy wordt de digitale informatie gewoon afgedrukt of op microfilm geplaatst. Het enige voordeel van deze piste is dat de informatie op een duurzame drager wordt geplaatst en dat heikele problemen zoals hard -en softwareondersteuning worden vermeden. Het nadeel is natuurlijk dat de functionaliteit en elk voordeel van digitale informatie verloren gaat en dat deze oplossing niet op alle multimedia bestanden kan worden toegepast. Bewegende beelden of geluid kunnen niet op deze wijze worden vastgelegd. Hard copy levert ook problemen op voor het verzekeren van de authenticiteit, want heel veel originele eigenschappen gaan verloren (bijv. vorm, structuur, context, ...). Overigens wees onderzoek van het NARA uit dat digitale archivering goedkoper is dan alle digitale informatie op papier af te drukken¹⁰.

⁹ K. THIBOUDEAU, *The unsteady state of the art of preserving electronic records*, Lezing gehouden tijdens VI^e Europees Archiefcongres, Firenze, 31 mei 2001.

Voor sommige auteurs is het gebruik van standaarden een vierde strategie. In de praktijk hangt deze piste nauw samen met het migreren van digitale informatie. De levensduur van standaarden is immers eveneens in tijd beperkt. Bij kleine toepassingen is standaardisatie vooral een kwestie van bestandsformaten. Bij grote informatiesystemen (bijv. mainframes) is het soort databanktoepassing van belang.

¹⁰ J. HOFMAN, *Het 'papier' tijdperk voorbij. Beleid voor een digitaal geheugen van onze samenleving*, Den Haag, 1995, p. 24.

11.2 MIGRATIE

Migratie kan algemeen omschreven worden als de archiveringsstrategie waarbij de bestanden naar een andere omgeving worden omgezet zodat ze met een nieuwe of andere computerconfiguratie compatibel zijn. Migratie heeft in zijn ruime betekenis betrekking op de hardware, het besturingssysteem en de applicatiesoftware. De bestanden worden met andere woorden aangepast aan de nieuwe omgeving waarbinnen ze in de toekomst worden gebruikt. Met migratie wordt soms ook refreshen (overzetten naar een andere drager) of conversie (overzetten naar een andere versie van dezelfde applicatie) bedoeld. Migratie heeft dan in engere zin betrekking op het omzetten naar een andere besturingssysteem¹¹.

Bij het overzetten van gearchiveerde digitale informatie staat de keuze van het bestandsformaat centraal. De bestanden worden doorgaans omgezet in een formaat dat duurzamer is of beter geschikt is voor archivering op lange termijn. Indien mogelijk dan gaat de voorkeur naar een gestandaardiseerd bestandsformaat uit en dit omwille van een aantal samenhangende factoren. Een eerste reden is dat softwareproducenten standaarden implementeren en er bij de ontwikkeling van hun applicaties rekening mee houden. Dit maakt de uitwisseling van computerbestanden mogelijk. Ten tweede kunnen digitale gegevens opgeslagen in een standaardformaat in principe door meerdere applicaties worden ingelezen. De kans is bijgevolg ook kleiner dat deze bestandsformaten moeten gemigreerd worden wanneer één bepaalde applicatie in onbruik raakt of niet meer wordt ondersteund. De duurzaamheid en bruikbaarheid van een gestandaardiseerd bestandsformaat is groter dan van een niet-gestandaardiseerd formaat. Toch is ook enige waakzaamheid geboden. Softwareproducenten voegen dikwijls applicatiegebonden eigenschappen aan de standaarden toe om zo hun marktaandeel te bevestigen of te vergroten. Standaarden hebben evenmin een onbepaalde levensduur en zullen in de toekomst ook gemigreerd worden, alleen zal dit minder frequent gebeuren dan bij niet-gestandaardiseerde formaten. Ten slotte is de bruikbaarheid van een standaard afhankelijk van hun toepassing in de praktijk. Er kan eigenlijk echt sprake zijn van standaardisatie wanneer producenten en eindgebruikers volop een standaard toepassen.

De bestandsformaten kunnen hiërarchisch worden ingedeeld. Bovenaan staan de officiële standaarden. Deze zijn vastgelegd door *officiële* standaardiserende instanties zoals *ISO (International Standard Organisation)* al dan niet in samenwerking met *IEC (International Electrotechnical Commission)* of *ITU (International Telecommunications Union)*. Het is geen toeval dat de formaten bestemd voor de uitwisseling van data de grootste groep van deze officiële standaardformaten vormen. De bekendste voorbeelden zijn ASCII, Unicode en SGML. *ISO*-standaarden zijn niet gratis verkrijgbaar en niet alle officiële standaarden kennen een grote verspreiding.

In de hiërarchie volgen na de officiële standaarden de *de facto* standaarden. Hoewel de term ‘standaarden’ eigenlijk is voorbehouden voor (inter)gouvernementele organisaties (*ISO, CEN, ETSI, ANSI, enz.*) wordt met *de facto* standaarden de bestandsformaten aangeduid die door hun

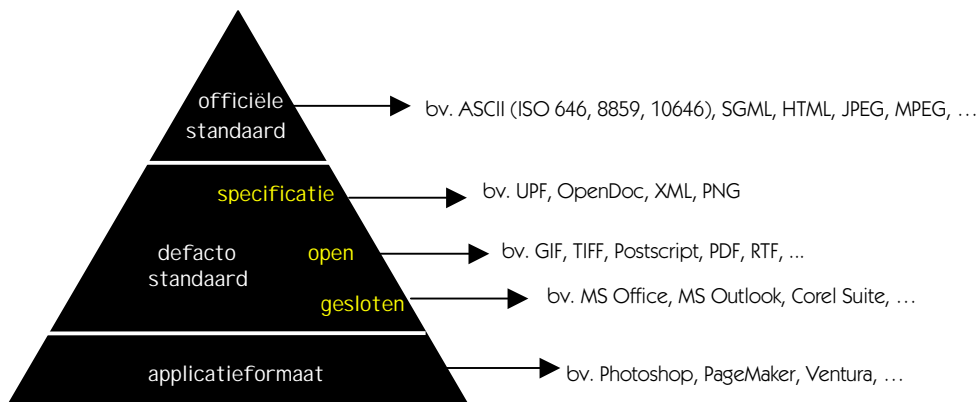
¹¹ Conversie en migratie worden frequent door elkaar gebruikt en er is nog geen consensus over een begripsafbakening. Zo wordt conversie soms ook gebruikt voor het overzetten van gegevens naar een andere softwareapplicatie (bijv. Lotus → Excel) en migratie dan voor het overzetten naar een ander platform (bijv. Unix → WinNT). (*Begrippenlijst digitale duurzaamheid; Ontwerp-Regeling geordende en toegankelijke staat archiefbescheiden 2000*: art. 1; C. DOLLAR, *Authentic Electronic Records: Strategies for Long-Term Access*, p. 29-31;). Anderen gebruiken conversie enkel voor het bewaren van data in een andere versie van bestandsformaat en migratie voor het overzetten naar een ander bestandsformaat (G.-J. VAN BUSSEL, *De opmaat voor hoe het niet moet*, in *Archos Magazine*, 11 (2000), p. 5).

weidverspreidheid normatief zijn geworden. De defacto standaarden kunnen in drie groepen worden onderverdeeld: de specificaties of aanbevelingen vastgelegd door normerende organisaties, de open bestandsformaten afhankelijk van één producent en de gesloten bestandsformaten.

De specificaties of aanbevelingen zijn het resultaat van samenwerkingsinitiatieven met normering of standaardisatie als doel. Eén van de bekendste voorbeelden op dit ogenblik is *W3C (World Wide Web Consortium)*. Net zoals bij officiële standaarden wordt voor het opstellen van de specificatie van deze bestandsformaten een hele procedure gevolgd. Samen met het feit dat meerdere partijen (softwareproducenten, universiteiten, consumenten) bij deze initiatieven zijn betrokken, wordt hierdoor een stukje stabiliteit gewaarborgd. De impact van deze specificaties mag niet worden onderschat. Vanuit zuiver juridisch perspectief staan de officiële standaarden hoger aangeschreven, maar gezien de betrokkenheid en de belangen van de grote softwareproducenten is de kans groot dat een aantal van deze specificaties een ruimere toepassing kennen dan een officiële standaard (bijv. XML versus SGML).

De open en gesloten bestandsformaten hebben als gemeenschappelijk element dat ze eigendom zijn van één producent. Het verschil tussen beide groepen is dat in het geval van de open bestandsformaten de specificatie is vrijgegeven en in het geval van de gesloten bestandsformaten niet. Helemaal onderaan in de hiërarchie staan ten slotte de bestandsformaten van weinig voorkomende commerciële toepassingen of ad hoc ontwikkelde toepassingen. Deze formaten zijn niet gemakkelijk uitwisselbaar en zijn afhankelijk van één applicatie.

Figuur 1: Hiërarchie van de bestandsformaten



De moeilijkheid bij migratie is dat er niet altijd geschikte formaten voor handen zijn. Archivarissen kunnen slechts de formaten gebruiken die op de markt beschikbaar zijn, en dit is afhankelijk van privé-initiatieven en van standaardisatieprojecten. In veel gevallen gaat migratie met verlies samen (bijv. lay-out, functies, ...) en moeten er keuzes worden gemaakt. Het komt er dan op aan dat op voorhand de belangrijkste elementen van de digitale archiefdocumenten (bijv. inhoud, structuur en context)

worden vastgelegd en dat het nieuwe formaat deze archiveert. Deze werkwijze werd bij de digitale archivering van het kiezersregister toegepast¹².

De migratie van gearchiveerde bestanden naar een geschikt archiveringsformaat zal maar zelden een éénmalige operatie zijn. Vooral bij het gebruik van open en gesloten defacto formaten en van applicatieformaten is de kans groot dat migraties zich zullen herhalen. Elke migratie of conversie moet gedocumenteerd worden (oorspronkelijk bestandsformaat, wijze, toetsing, informatieverlies, enz).

11.3 BEWARING VAN DE TECHNOLOGIE

Als de bestandsformaten niet aan de nieuwe technologie worden aangepast, moet de originele technologie op één of andere manier worden bewaard of gesimuleerd. Technisch gezien zijn er hiervoor meerdere mogelijkheden zoals het bewaren van originele hard- en/of software en het gebruik van configureerbare chips of virtuele machines. Emulatie is wellicht het bekendste voorbeeld om in de toekomst verouderde technologie te gebruiken.

De strategie van computermusea is het bewaren van de originele hard- en/of software die vereist is om de gearchiveerde bestanden te raadplegen. Deze oplossing heeft de voordelen dat zowel de archiefbescheiden als de nodige hard- en softwareondersteuning in hun oorspronkelijke en originele vorm worden behouden. De archiefbescheiden blijven in digitale vorm beschikbaar, zonder dat de archivaris zich om migratie of emulatie moet bekommeren. Het regelmatig overzetten van de bestanden naar andere dragers zou de voornaamste beheers- en onderhoudstaak zijn. De kostprijs, de vereiste informaticakennis, de beperkte levensduur van schijven en chips en het gebrek aan commerciële ondersteuning maken deze oplossing minder plausibel, zoniet onrealistisch.

Emulatie is de archiveringsstrategie waarbij bestanden in hun origineel formaat worden bewaard en waarbij de vereiste hard- en softwareomgeving wordt nagebootst op een “host”-systeem¹³. Emulatie is toepasbaar op hardware, software of op beide samen. Eén van de eenvoudigste toepassingen is de emulatie van de hardware en het besturingssysteem. In dit geval bestaat de emulatiestrategie uit volgende stappen:

- ✓ opstellen van een emulatie specificatie voor elk platform. Deze specificatie bevat informatie over de hardware- en softwarevereisten. Hierin wordt ondermeer het register, de jumpersettings, de operandi, de bestandsformaten, enz. gespecificeerd. De specificatie wordt in een bepaalde taal (*emulator specification language*) vastgelegd.
- ✓ creëren van een interpreter geschikt voor de emulatiespecificatie en aangepast aan de virtuele machine. De interpreter is een programma die draait op een virtuele machine (VM) en die de emulatie van het oude platform creëert. Voor elke specificatietaal is er een andere interpreter

¹² Het DAVID-rapport over de digitale archivering van het kiezersregister is beschikbaar op de DAVID-website (onder publicaties → rapporten).

¹³ <http://129.11.152.25/CAMiLEON/dh/ep5.html>; J. ROTHENBERG, *Avoiding technological quicksand*, 1998; J. ROTHENBERG, *An experiment in using emulation to preserve digital publications*, april 2000; S. GILHEANY, *Preservation Information Forever and a Call for Emulators*, Singapore, 1998.

vereist. Een interpreter kan worden vergeleken met een compiler die leesbare broncode in machinetaal omzet.

- ✓ de oorspronkelijke applicatie wordt op de emulator opgestart.
- ✓ de gearcheeerde digitale bestanden worden in hun (oorspronkelijke) applicatie opgevraagd. Het origineel digitaal bestand en de originele toepassing of de vereiste software voor het renderen van het document moeten samen met de specificatie en de interpreter worden gearcheeerd.
- ✓ wanneer de VM verouderd raakt, dan moet er een specificatie en interpreter voor een nieuwe VM worden geschreven, die de oude VM kan simuleren. De specificatie en interpreter van de oude emulator kunnen vervolgens op de nieuwe VM draaien. De recentste VM moet in staat zijn om alle vorige versies van VM's te emuleren. De raadpleging van een gearcheeerd digitaal bestand verloopt dan als volgt: 1. Opstarten van de interpreter van de overeenstemmende VM op de huidige VM, 2. Opstarten van de interpreter van de passende applicatie, 3. Openen van het gearcheeerde bestand.

In een dergelijke emulatietoepassing moeten naast de gearcheeerde bestanden ook de applicatiesoftware en de emulatiespecificaties worden bewaard. Metadata spelen bij emulatie een heel belangrijke rol. Naast de gebruikelijke metadata moet immers ook de emulatiespecificatie worden bijgehouden. Omdat die informatie ten allen tijde moet beschikbaar zijn, wordt zelfs aangeraden om de specificatie op papier af te drukken. Bij emulatie is het in de eerste plaats de bedoeling dat de basisfunctionaliteiten van het informatiesysteem kunnen uitgevoerd worden. Aanvullende of niet essentiële functies van het informatiesysteem hoeven in principe niet geëmuleerd te worden. Toch kan dit betekenen dat voor het werken met een geëmuleerde toepassing specifieke kennis is vereist, waardoor de bewaring van gebruikershandleidingen nodig is.

De programmeertalen gebaseerd op C/C++ zijn de meest gebruikte ontwikkelomgeving voor emulaties. Emulatie wordt al algemeen toegepast bij het operationeel houden van video -en computerspelletjes. Men is het vooralsnog oneens over het tijdstip waarop de emulatiesoftware moet worden gecreëerd. Rothenberg meent dat er kan gewacht worden totdat het platform in onbruik is geraakt, terwijl anderen zoals Holdsworth en Wheatley het belangrijk vinden dat er vroeg genoeg wordt gestart met de ontwikkeling van emulatieprogramma's.

III. BESLISSINGSMODEL VOOR HET VASTLEGGEN VAN DE ARCHIVERINGSSTRATEGIE

Beide archiveringsstrategieën hebben voor- en tegenstanders. Voorstanders van emulatie (bijv. J. Rothenberg en S. Gilheany) halen als argumenten aan dat emulatie goedkoper en minder arbeidsintensief zou zijn. Door bestanden in hun origineel formaat op te slaan, behouden ze hun functionaliteit, hun vorm en hun originele “look and feel”. Deze elementen worden gebruikt om de authenticiteit en integriteit aan te tonen en gaan in veel gevallen bij migratie (deels) verloren. Tegenstanders van emulatie wijzen op de noodzakelijkheid om VM’s te migreren en stellen de technische haalbaarheid in vraag. De eerste resultaten van emulatietests lijken hen gelijk te geven, maar verder onderzoek blijft noodzakelijk vooraleer men een definitief oordeel over emulatie kan vellen. Voorstanders van migratie (C.M. Dollar, D. Bearman, e.a.) zijn zich bewust van de tekortkomingen van de migratiestrategie en geven toe dat er situaties bestaan waarin het technisch onhaalbaar is om met migraties de functionaliteit en de integriteit te bewaren. In deze gevallen moet overwogen worden een andere piste te volgen (emulatie, hard copy, ...) of moet de archivaris keuzes maken en enkel de essentiële gegevens archiveren¹⁴.

Vanuit beheersmatig standpunt lijkt het ideaal dat de afhankelijkheid van hard -en software tot een minimum wordt herleid en dat er zo softwareloos mogelijk wordt gearchiveerd. De migratiepiste waarbij bestanden worden overgezet naar een standaardformaat benadert dit ideaal beter dan emulatie, maar op basis van de inventaris van digitale informatiesystemen van de stad Antwerpen lijkt het niet opportuun om een principiële beleidskeuze tussen migratie of emulatie als archiveringsstrategie te maken. In de praktijk moet men rekening houden met het feit dat de elke opeenvolging van bits slechts door middel van gebruik van de gepaste software betekenis krijgt. De software-afhankelijkheid verschilt van informatiesysteem tot informatiesysteem. Cruciaal is de vraag of de digitale bestanden kunnen worden gearchiveerd zonder de softwareomgeving waarbinnen ze zijn gecreëerd. Voor het ene informatiesysteem zal emulatie de meest aangewezen oplossing zijn, voor een ander zal dat migratie zijn. Migratie kan bijvoorbeeld worden gevolgd bij de archivering van het kiezers-en bevolkingsregister. GIS-toepassingen zijn zodanig sterk in elkaar geïntegreerd en softwareafhankelijk dat emulatie de beste oplossing lijkt te zijn. Dynamisch webpagina’s die uit ASP- of JAVAscripts bestaan, kunnen enkel met de gepaste server- of clientsoftware in een webbrowser worden bekeken. Bepalende factoren in de keuze tussen beide strategieën zijn ongetwijfeld de toekomstige functionaliteit van de gearchiveerde data en de afhankelijkheid van bestanden. Digitale bestanden kunnen immers afhankelijk zijn van bepaalde softwareapplicaties en van externe bestanden. Overigens is het niet uitgesloten dat de levenscyclus van gearchiveerde digitale bestanden uit een opeenvolgende migratie- en emulatiefase (of omgekeerd) kan bestaan.

Algemeen zou men kunnen stellen dat migratie de beste oplossing is, wanneer het volstaat dat de datastream wordt gearchiveerd en wanneer de functionaliteit beperkt blijft tot het opvragen en raadplegen van vastgelegde informatie. Van zodra ook de structurele laag of tools mee moeten worden

¹⁴ C.M. DOLLAR, *Authentic Electronic Records: Strategies for Long-Term Access*, Chicago, 2000, p. 72-74; D. BEARMAN, *Reality and chimeras in the preservation of electronic records*, in *D-Lib Magazine*, april 1999.

bewaard omwille van de functionaliteit, wint emulatie aan belang. Migratie heeft in ieder geval op emulatie voor dat het technisch minder complex is en eventueel door archiefdiensten kan worden uitgevoerd. Emulatie daarentegen is werk voor specialisten en houdt in dat naast de gearchiveerde bestanden ook emulaties (specificaties, interpreters en VM's) moeten worden bewaard.

III.1 BEHEERSINVENTARIS VAN DE DIGITALE INFORMATIESYSTEMEN

Bij het uitstippelen van een archiveringsstrategie wordt dus best uitgegaan van het informatiesysteem. Een eerste stap is natuurlijk dat de archivaris moet weten welke digitale archiefbescheiden de administratie vormt. Een systematisch opgebouwd overzicht zoals een inventaris van de digitale informatiesystemen is omwille van diverse redenen van belang. Het is belangrijk dat de archivaris op de hoogte blijft van alle informatiesystemen die bij de archiefvormer lopen en welke IT-toepassingen hiervoor worden gebruikt. Zo weet hij welke informatie op welke lokatie in digitale vorm aanwezig is. Op basis van de gegevens die over elk informatiesysteem worden bijgehouden, moet hij de archiefwaarde kunnen afleiden. De gegevens moeten hem ook in staat stellen om een archiveringsbeleid uit te stippelen voor de informatiesystemen met archiefwaarde. Een dergelijke inventaris kan als meta-informatiesysteem dienen, waaruit men later ook de nodige metadata kan overnemen. Eén van de belangrijkste elementen die ook in de inventaris moet worden beschreven is de context van het archiefdocument en de relatie met andere papieren of digitale archiefdocumenten. Het is één van de weinige plaatsen waar men deze gegevens kan vastleggen en deze gegevens zijn belangrijk voor de efficiënte archivering van authentieke archiefdocumenten.

Bij het verzamelen van gegevens over de informatiesystemen van de stadsadministratie, het OCMW en het havenbedrijf van Antwerpen viel eens te meer op dat de systemen in de eerste plaats gericht zijn op het informatiseren van werkprocessen en handelingen, waardoor een functioneel opgebouwde inventaris voor de hand lag. Deze contextualisering en een korte beschrijving van de functie van het systeem maken het mogelijk dat de archivaris de archiefwaarde kan inschatten. Het opstellen en bijhouden van een beheersinventaris van de digitale documenten die door een administratie worden gevormd, is dan ook een sleutelinstrument in het digitaal archiefbeleid van een archiefdienst. Deze inventaris is niet in dit rapport opgenomen, maar vormt een afzonderlijk werkdocument dat continue opvolging behoeft¹⁵.

Het opstellen en bijhouden van een inventaris van de informatiesystemen is geen gemakkelijke opgave. Uit Nederlandse en eigen ervaring blijkt dat de basisinformatie voor een dergelijke inventaris maar zelden op een systematische wijze bij de administratie of de informaticadienst aanwezig is. Uit deze vaststelling groeide de idee om het bijhouden van een inventaris van de digitale informatiesystemen formeel op te leggen¹⁶. Een bijkomende moeilijkheid bij het verzamelen van

¹⁵ De inventaris van digitale informatiesystemen is raadpleegbaar op de DAVID-site (onder 'publicaties' en vervolgens 'overige publicaties').

¹⁶ Een dergelijke inventaris wordt bijvoorbeeld voorgeschreven in het *Besluit Informatiebeheer Provincie Zeeland 1997*, art. 12: "Het hoofd van de beheerseenheid ziet erop toe, dat van informatiebestanden een inventaris wordt aangelegd en bijgehouden, waarin de informatiebestanden worden beschreven en in verband kunnen worden gebracht met de verschillende werkprocessen en taken". Deze verplichting is ook opgenomen in de *Ontwerp-Regeling geordende en toegankelijke archiefbescheiden 2000* (art. 11) en is een mijlpaal in de *Modernising Government 2004 Requirement* van het Public Record Office.

gegevens over het digitaal archief van een organisatie of instelling is de problematiek inzake de harde schijven van lokale werkstations. De C- en D-schijven zijn in veel gevallen de opslagplaats van digitale archiefdocumenten zonder enige vorm van controle of goed beheer. Wellicht kan het bijhouden van een inventaris bijdragen tot de sensibilisering van de administratie.

III.2 INFORMATIELAGEN ALS UITGANGSPUNT

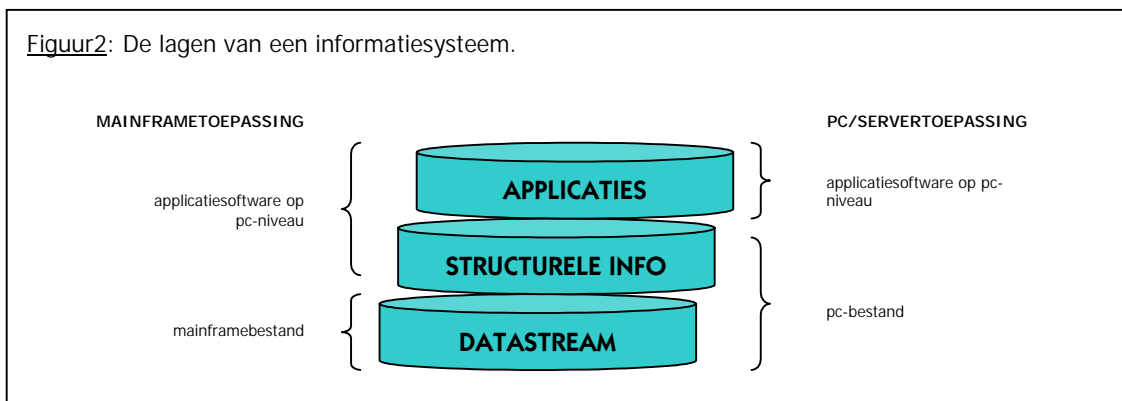
Eén van de redenen waarom er geen typologie als basis voor het vastleggen van archiveringsstrategieën kan worden opgesteld, is de grote diversiteit in het geheel van informatiesystemen. De bestandsformaten waarin de digitale informatie wordt opgeslagen, de configuraties, de gebruikte softwaretoepassingen, de aard van de informatie, de functionaliteiten en afhankelijkheden verschillen van systeem tot systeem. Desondanks zijn er een aantal gemeenschappelijke elementen vast te stellen, waardoor toch een richting voor de besluitvorming inzake het uitbouwen van een archiveringsstrategie kan worden aangegeven:

- ✓ De informatiesystemen zijn gericht op het automatiseren van functies van de overheid. Voor de grote of belangrijke functies van de overheid zijn er specifieke informatiesystemen ontwikkeld. De bestanden (tekst, databank, spreadsheets) die uit één functioneel werkproces voortvloeien worden binnen hetzelfde systeem gegenereerd en in veel gevallen aan elkaar gekoppeld. Ook informatiesystemen zijn soms aan elkaar gekoppeld en mogen niet zomaar als geïsoleerde entiteiten worden gearchiveerd. Deze vaststelling leidde tot de overtuiging dat men de informatiesystemen zelf als uitgangspunt moet hanteren, en niet hun type. Aangezien men een duidelijke samenhang tussen de informatiesystemen en de functies van de archiefvormer vaststelt, is een systematisch overzicht van de informatiesystemen het gemakkelijkst functioneel in te delen. De koppeling aan functies is van belang voor de contextualisering en om de archiefwaarde vast te leggen.
- ✓ Digitale gegevens worden in bestanden opgeslagen. De bestanden zijn opgebouwd volgens een codering en indeling die eigen is aan de gebruikte applicatie. Elk bestandsformaat heeft zijn unieke code, indeling en ordening van de bits. In het algemeen bestaat de indeling uit twee componenten: de structurele- of logicalaag enerzijds en de datastream anderzijds. De structurele of logica elementen zijn de gegevens over de data die de applicatie gebruikt bij het uitvoeren van haar functionaliteiten en zijn dus ook eigen aan de applicatie en het bestandsformaat. De datastream is de bitreeks met de ingevoerde of gegenereerde data. Dit gedeelte van de bitstream kan van alles betekenen: tekst, geluid, afbeelding, video, enz.¹⁷. Zonder de structurele informatie kan de datastream niet (correct) weergegeven worden, flat

¹⁷ Ter illustratie: een structureel element van een spreadsheet of een databank is een header die informatie bevat over respectievelijk de celindeling en formules of de velden (naam, lengte, datatype, ...) en index. Een tekstverwerkingsbestand zoals dat van MS Word is een Object Linking and Embedding applicatie die multistreamfiles genereert. Deze streamfiles zijn aan elkaar gekoppeld en vormen één bestand. Een MS Word97 bestand bestaat uit een main stream (o.a. header, formatting information), summary information stream, tablestream (o.a. plcf's), datastream en 0 of meerdere objectstreams. De vier bitreeksen naast de datastream zijn structurele elementen. Een GIF-file bestaat uit een header, een global en een local colour table (beiden optioneel), een local image descriptor en de image data. PDF-files zijn als volgt intern gestructureerd: header, body, cross reference table en trailer.

files uitgezonderd maar dan moet wel vastliggen dat de bits als ASCII-karakters moeten worden weergegeven¹⁸.

- ✓ Tekstbestanden, spreadsheets en databanken op pc's en servers hebben als gemeenschappelijke factor dat de bits van hun datastream ASCII- of Unicodekarakters weerspiegelen. Hun structurele elementen verschillen, want deze is afhankelijk van de soort applicatie, van de gebruikte applicatie (produkt) en van de versie. De datastream vormt de kern van elk bestand of informatiesysteem en is hetgeen wat eerst en vooral moet worden gearchieveerd. Of de structurele elementen mee moeten gearchieveerd worden, is afhankelijk van de toekomstige functionaliteit van de gearchieveerde data elementen en van de handelingen die nog moeten worden uitgevoerd. Wanneer digitale gegevens bij opslag worden gecomprimeerd moet men er over waken dat ook de geschikte tool voor decompressie voor handen is, anders wordt latere raadpleging onmogelijk. Op die manier kunnen we elk informatiesysteem in lagen opsplitsen. Het onderzoek van het informatiesysteem moet dan uitwijzen welke lagen al dan niet samen met de datastream worden gearchieveerd. Als er geen bijzondere functionaliteiten zijn en geen dynamische componenten mee worden gearchieveerd, dan richt de digitale archivering zich op de datastream. In het andere geval moet samen met de datastream structurele informatie en eventueel tools worden bewaard.



- ✓ De meeste informatiesystemen bevatten enkel tekstuele informatie die in de vorm van databanken wordt bijgehouden. Deze datastream op zich kan door meerdere applicaties worden ingelezen. Van de 59 informatiesystemen van de stedelijke administratie en het OCMW van Antwerpen waarvan de soort toepassing kon achterhaald worden, vormen voor 51 systemen databanken (86 %) de kern van de toepassing. Databanken zijn algemeen de meest toegepaste informaticatoepassingen en vormen ook de grootste groep informatiesystemen met archiefwaarde¹⁹. Wanneer het volstaat dat enkel de datastream wordt

¹⁸ In het OAIS-model worden er vijf lagen in een informatiesysteem onderscheiden: fysieke, binaire, structuur, object en applicatie. De drie lagen die wij onderscheiden stemmen overeen met de lagen structuur, object en applicatie. De fysieke en binaire laag zijn wel van belang maar laten wij hier buiten beschouwing omdat de keuze van de drager (fysieke laag) in principe vrij is en het bestandensysteem (binaire laag) samenhangt met het besturingssysteem.

¹⁹ Dit grote aandeel van databanktoepassingen in het geheel van informatiesystemen wordt bevestigd door gegevens over de aard van de computertoepassingen bij Nederlandse en Canadese overheden (A.A.C. JANSSEN, *MLG's geteld en gewogen*, in J. HOFMAN (red.), *Het papieren tijdperk voorbij. Beleid voor een*

gearchiveerd, kunnen de bestanden als flat file worden bewaard. In het geval van pc/servertoepassingen wordt de datastream als het ware uit het bestand gefilterd en ontdaan van de structurele laag. Het resultaat is een gewoon ASCII- of SGML/XML-bestand. Voor mainframebestanden wordt in dit geval een file transfer uitgevoerd waarbij de EBCDIC-karakters naar ASCII worden omgezet.

EBCDIC (Extended Binary Coded Decimal Interchange Code) is IBM's 8-bit uitbreiding van de 4 bits Binary Coded Decimal codetabel en vormt de basis van de codetabellen die tot op de dag van vandaag door mainframes worden gebruikt. Er bestaan verschillende codetabellen. De codetabellen voor België zijn 274 (Belgian) of 500 (Belgian New). Het is eveneens mogelijk een eigen codetabel op te stellen. Zo wordt in Antwerpen voor BEAM een codetabel gebruikt die op nr. 274 is gebaseerd, maar die extra karakters bevat om een aantal vreemde tekens toe te laten. De mainframecodetabel wordt dan doorgaans vanuit de applicatie gestuurd. Mainframebestanden kunnen naar server- of pcniveau getransfereerd worden via een gewone FTP- of SNA- file transfer. Er zijn hier eveneens een aantal specifieke tools voor beschikbaar. De EBCDIC-code wordt hierbij in ASCII omgezet. Deze omzetting gebeurt bijna nooit foutloos. Vooral de omzetting van diakritische tekens levert nogal wat problemen op. Bij een filetransfer van EBCDIC naar ASCII moet men beschikken over een COBOL copy-book (*.LAY: data lay-outfile) die de record lay-out (structuur, lengte van de velden, enz.) van de bestanden beschrijft. Het resultaat van een dergelijke file transfer is een gewone flat file.

Bij de omzetting van mainframebestanden naar server- of pcniveau moet niet enkel rekening worden gehouden met de diakritische tekens, maar ook met de bestandsomvang. Mainframebestanden zijn doorgaans te groot om als één server- of pcbestand gebruikt te kunnen worden. Een opsplitsing in kleinere bestanden zal zich in veel gevallen opdringen.

III.3. BESLISSINGSMODEL

Het onderscheiden van de lagen in een informatiesysteem is een goede vertrekbasis om te onderzoeken welke archiveringsmethode het best geschikt is voor een bepaald informatiesysteem. In een mainframetoepassing zijn de lagen duidelijker van elkaar gescheiden dan in een gewone pc-applicatie. Het ziet er overigens naar uit dat in de toekomst bij gewone pc-toepassingen het onderscheid tussen de data laag en de structurele laag duidelijker wordt. Het gebruik van SGML/XML als bestandsformaat en SQL-databanken illustreert alvast deze evolutie. De eerste vraag heeft hierop betrekking.

- **Wat** archiveren we: enkel de datastream? de datastream in combinatie met de structurele laag? de applicatietool? kan er geselecteerd worden? worden er verschillende versies overgedragen?
- **Wie** beheert het digitaal archief: de archiefvormer, de informaticadienst of de archiefdienst?

- **Hoe** dragen we het digitaal archief over: in welk bestandsformaat? op welke manier komt de archiefdienst in het bezit van het archief?
- **Wanneer** wordt het digitaal archief overgedragen: na het verstrijken van de administratieve bewaartermijn? na een upgrade? na het bereiken van een bepaalde bestandsomvang? met welke periodiciteit?

Op basis van deze vragen en hun antwoorden kunnen de grote lijnen van de archiveringsstrategie worden uitgetekend. Samen vormen ze de bouwstenen van het beslissingsmodel voor de wijze waarop digitale bestanden worden gearcheeerd. Bij wijze van voorbeeld wordt soms verwezen naar informatiesystemen die bij de administratie van het stadsbestuur, van het OCMW of van het havenbedrijf lopen. Meer informatie over deze systemen is in de *Beheersinventaris van digitale informatiesystemen* opgenomen (zie DAVID website).

III.3.1 WAT wordt gearcheeerd ?

Hoewel bij elke vraag het informatiesysteem zelf steeds het vertrekpunt is, wordt hiermee niet aangegeven dat het volledige systeem met alle lagen wordt gearcheeerd. Dit kan bij bepaalde informatiesystemen het geval zijn, maar in andere gevallen zal het volstaan dat enkel de data of althans een gedeelte daarvan wordt gearcheeerd. In de praktijk moeten er ook nog andere gegevens samen met de gearcheeerde data worden vastgelegd, maar hier wordt de focus enkel op te archiveren digitale archiefdocumenten gericht.

Bij een overdracht van papieren bescheiden wordt doorgaans al een eerste selectie doorgevoerd. De te archiveren archiefbestanddelen worden gescheiden van de bestanddelen die vernietigd mogen worden. Het onderscheid tussen bewaring en vernietiging wordt hoofdzakelijk bepaald door het juridisch en historisch belang van de stukken. In een digitale omgeving is de keuze niet alleen afhankelijk van de archiefwaarde, maar ook van de vraag of het technisch wel mogelijk is om die digitale gegevens te vernietigen. Afzonderlijke of op zichzelf staande bestanden (bijv. tekstverwerkingsbestanden) kunnen probleemloos worden vernietigd, zonder dat het informatiesysteem daar hinder van ondervindt.

Moeilijker wordt het wanneer slechts bepaalde gegevens van grote of geïntegreerde informatiesystemen permanent bewaard hoeven te worden. Blijft het permanent te bewaren gedeelte van het bestand intact wanneer het geselecteerde deel wordt vernietigd? Kan de toepassing verder functioneren zonder de vernietigde informatie? Kan het permanent te bewaren gedeelte dan nog zonder problemen worden geconsulteerd? Hoe zit het dan met de integriteit van het bestand? Een voorbeeld hiervan zou FIN2000, een mainframetoepassing voor het bijhouden van de stadsboekhouding, kunnen zijn. Met behulp van FIN2000 worden zowel permanent (jaarrekening, grootboek, ...) als tijdelijk (dagboek) te bewaren archiefstukken aangelegd. Kan het dagboek worden vernietigd zonder de raadpleegbaarheid van de jaarrekening of het functioneren van de hele toepassing in gevaar te brengen?

Bij digitale gegevens moet dus niet alleen nagegaan worden welke digitale bestanden een archiefwaarde hebben, maar moet ook onderzocht worden of het technisch mogelijk is permanent te bewaren informatie te scheiden van de informatie die voor vernietiging in aanmerking komt. In de

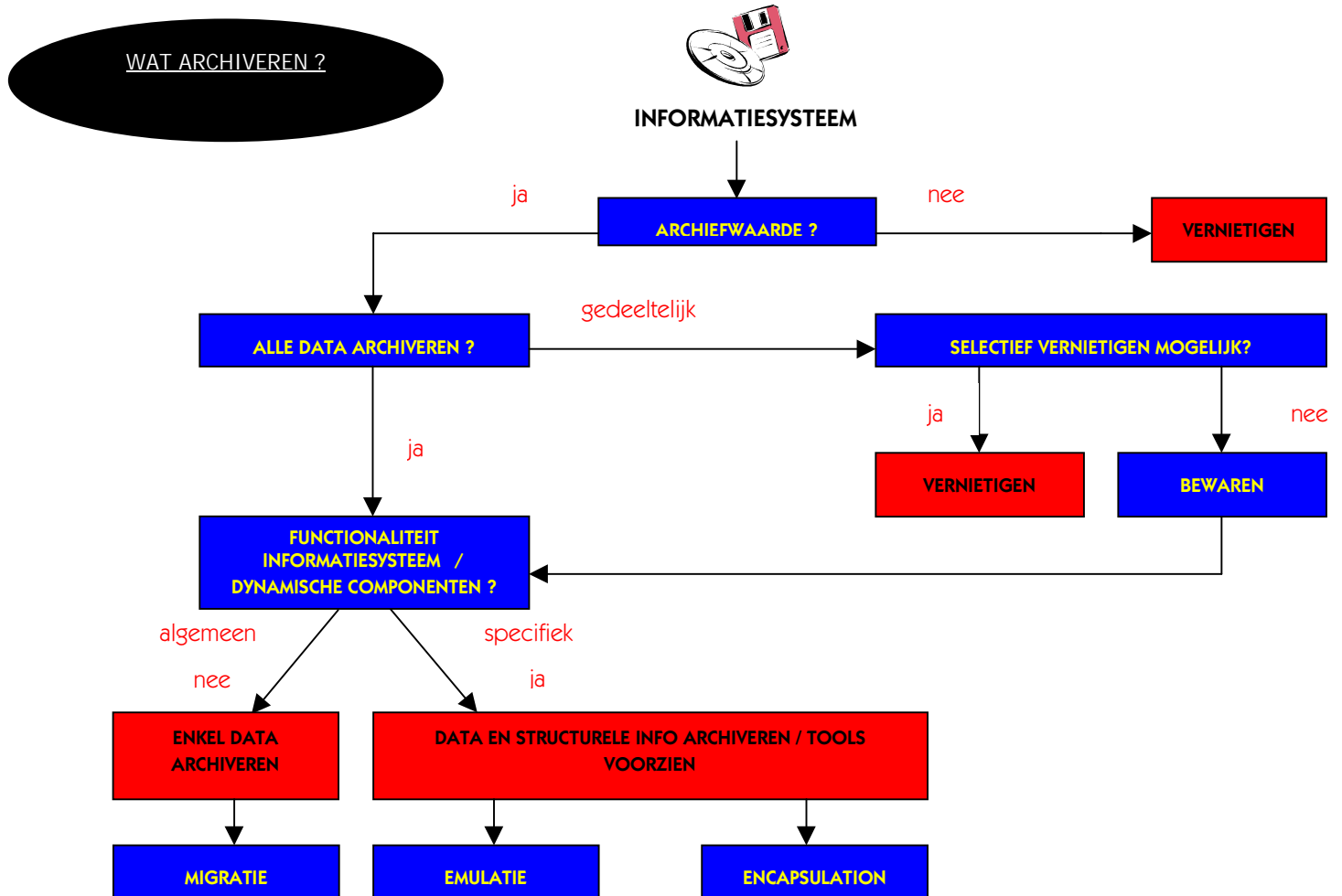
meeste gevallen zal het wellicht niet mogelijk zijn om bijvoorbeeld van een bestand of informatiesysteem enkel dat deel met een archiefwaarde te archiveren en het andere deel te vernietigen. De keuze zal in deze gevallen beperkt worden tot het integraal bewaren of integraal vernietigen van het bestand of systeem. De selectie zal met andere woorden niet meer op het niveau van de digitale archiefdocumenten gebeuren, maar veeleer ook op het niveau van het informatiesysteem. Zo kan het voorvallen dat in de toekomst informatie die in een papieren omgeving werd vernietigd digitaal wel bewaard wordt omdat het onlosmakelijk verbonden is met permanent te bewaren informatie of met het informatiesysteem. Dit heeft voor gevolg dat selectielijsten opgesteld voor papieren bescheiden niet zomaar toepasbaar zijn op digitale archiefbescheiden.

Een andere aanpak kan inhouden dat het permanent te bewaren gedeelte van de digitale informatie uit het systeem wordt geëxporteerd en uiteindelijk dan toch afzonderlijk wordt gearcheveerd. Volgt men deze weg, dan archiveert men die data buiten het oorspronkelijke informatiesysteem en moet men elementen zoals de context op een andere manier vastleggen, bijv. omwille van authenticiteitsdoeleinden. Welke bijkomende gegevens samen met de digitale data worden gearcheveerd, wordt momenteel nog volop onderzocht.

Ten tweede rijst de vraag of het volstaat om de data en structurele informatie van een toepassing te archiveren en of men geen bijzondere tools of software mee moet archiveren. Digitale informatie heeft immers als kenmerk dat de informatie slechts met behulp van technologie kan worden geraadpleegd. Voor de data met een algemene en weidverspreide functionaliteit zoals een tekst- of gewoon databestand waarvoor de functie van de tool beperkt blijft tot het opvragen en raadplegen van informatie is dat overbodig en volstaat het enkel de data te archiveren. Er zullen in de toekomst voldoende applicaties zijn die deze basisfuncties ondersteunen. Bij de beschrijving van de typen informaticasystemen viel echter op dat voor veel systemen specifieke op maat geprogrammeerde tools waren ontworpen. Een voorbeeld hiervan is de GIS-viewer KAVIA die door Telepolis werd ontwikkeld. Kan de informatie van deze toepassingen (bijv. kadastrale informatie gekoppeld aan de GIS-kaart) nog worden geraadpleegd, opgezocht of samengesteld zonder de tool KAVIA of wanneer andere tools worden ingeschakeld? Als dit niet het geval is dan moet die specifieke functionaliteit van de applicatie en de nodige software mee worden gearcheveerd. Overigens stelt men vast dat de afhankelijkheid van digitale bestanden van specifieke softwarepakketten de laatste jaren sterk is toegenomen. Dit is een evolutie die samengaat met de alsmaar grotere verspreiding van multimediatoepassingen en de verder gaande integratie van informatiesystemen. Hierdoor stelt de archivering van een informatiesysteem als GIS zoveel problemen. De diverse GIS-componenten zijn sterk in elkaar verweven en afhankelijk van specifieke softwareapplicaties. De archivering van gecomprimeerde bestanden stelt een analoog probleem. Op basis van de functionaliteit van de bestanden of van het informatiesysteem moet beslist worden of enkel de bestanden met de informatie ofwel (een deel van) het informatiesysteem moet worden gearcheveerd. Wanneer (een deel van) het informatiesysteem wordt gearcheveerd, betekent dit dat er samen met de gegevensbestanden software wordt opgeslagen. In een aantal gevallen kan dat zelfs leiden tot het mee archiveren van de oorspronkelijke applicatie. Wanneer met de data ook bepaalde tools worden gearcheveerd, dan moet men er over waken dat de tool verder kan worden gebruikt en dat de vereiste hardwareconfiguratie en het passende besturingssysteem is voorzien. Volgt men een dergelijke piste, dan zal men vroeg of laat een emulator van het vereiste platform moeten voorzien. Veel zal ook afhangen van de manier waarop de digitale informatie beschikbaar wordt gesteld.

Wanneer digitale informatie samen met een applicatie aan de archiefdienst wordt overgedragen, moet er ook gedacht worden aan een softwarelicentie voor de archiefdienst.

Over de gearchiveerde digitale informatie en eventueel bijhorende tools moeten er ook metadata worden bijgehouden door de archiefdienst. De metadata hebben doorgaans betrekking op de technische eigenschappen en de contextgegevens van de gearchiveerde digitale informatie. De emulatiespecificatie kan eveneens tot de metadata behoren. Metadata moeten ten allen tijde leesbaar zijn. Zij worden bijgevolg als ASCII-bestand of op papier bewaard. Metadata worden bij voorkeur gestandaardiseerd. Zij worden in grote mate door de archiefvormer en de informaticadienst verstrekt.



III.3.2 WIE beheert het digitaal archief ?

Papieren archiefbescheiden met archiefwaarde kan men na het verstrijken van de administratieve bewaartermijn aan de archiefdienst overdragen. De archiefdienst is vanaf dan volledig verantwoordelijk voor het archiefbeheer. Bij digitale archiefbestanden is het niet altijd vanzelfsprekend dat de permanent te bewaren bescheiden naar de archiefdienst worden overgedragen. In een digitale omgeving kunnen zowel de archiefdienst, de informaticadienst als de archiefvormer of zelfs alle drie samen de gearchiveerde bestanden beheren.

Non-custodial

De Australische archiefdiensten hangen de *non-custodial* opvatting aan. Zij vertrouwen het beheer van digitaal archief in eerste instantie aan de archiefvormende diensten toe. De archiefvormers beschikken immers over de nodige technische infrastructuur en kennis om de digitale systemen of documenten te beheren. De taken van de archiefdiensten beperken zich in dit concept voornamelijk tot selectie en ontsluiting. Het Australisch concept acht het niet haalbaar dat de archiefdienst over elk platform beschikt dat binnen de administratie wordt gebruikt. Deze visie hangt samen met hun opinie dat de goedkoopste en beste oplossing voor de digitale archivering een zo lang mogelijk beheer binnen de originele hard -en softwareconfiguratie is. Zolang deze operationeel is hoeven de digitale bestanden niet gemigreerd te worden. De applicatie moet wel voorzien zijn op het beheren en toegankelijk maken van de gearchiveerde bestanden. De meeste systemen voldoen nog niet aan deze eis, zodat het bij een dergelijke archiveringswijze des te belangrijker zal zijn dat de archivaris van bij de ontwikkeling van een systeem daarop toeziet. Met deze aanpak hopen Australische archivariissen een aantal migratiestappen uit te sparen, wat werk- en kostenbesparend is. Bovendien blijft de toegankelijkheid en de authenticiteit van de bestanden verzekerd, want ze blijven bewaard in hetzelfde informatiesysteem waarin ze zijn aangelegd²⁰. Een dergelijke oplossing kan handig zijn voor applicaties die op specifieke platformen (bijv. mainframe, Unix, ...) worden bijgehouden of waarvoor bijzondere applicaties zijn vereist en waarover de archiefdienst niet beschikt. Binnen de Antwerpse stadsadministratie worden veel grote bestanden op mainframeniveau bijgehouden, die niet als dusdanig door de archiefdienst ter beschikking kunnen worden gesteld.

Beheer binnen de oorspronkelijke omgeving kan ook maar duren zolang de softwareomgeving operationeel is of wordt ondersteund. Indien dit niet langer het geval is, dan moet er een keuze worden gemaakt. Ofwel blijven de gearchiveerde bestanden onder het beheer van de archiefvormende dienst en worden ze naar de nieuwe versie of applicatie overgezet zodat men ze verder met behulp van de actieve applicatie kan beheren en raadplegen. Ofwel worden ze overgedragen aan de archiefdienst die ze migreert naar een formaat dat wel wordt ondersteund of voorziet men in een emulatie van het oorspronkelijk systeem. Een moeilijkheid in een *non-custodial* archiveringsbeleid is het ter beschikking stellen van de gearchiveerde data aan vorsers. Zij moeten zich tot de administratie wenden om toegang te krijgen tot deze digitale archiefbestanden, tenzij de archiefdienst toegang tot het informatiesysteem verschaft.

Een toepassing van deze opvatting is bij object-geörienteerde informatiesystemen niet zo veraf. In een dergelijke toepassing kan er van elk object een historiek worden bijgehouden. Op die manier worden ‘verouderde’ of ‘te archiveren’ gegevens binnen het oorspronkelijke systeem beheerd.

Custodial

De Amerikaanse en Canadese archiefdiensten hanteren de *custodial* opvatting, wat betekent dat zij zelf verantwoordelijk zijn voor het beheer van de digitale archieven. Een basisvoorwaarde voor deze optie is dat de archiefdiensten de nodige technische ondersteuning kunnen bieden. Veel hangt af van de informaticavoorzieningen waarover de archiefdienst beschikt en de manier waarop het digitaal depot is georganiseerd.

Vooraleer digitale bestanden onder het beheer van de archiefdienst komen, moet onderzocht worden of de archiefdienst beschikt over de hardwareconfiguratie, het besturingssysteem en de

²⁰ http://www.naa.gov.au/recordkeeping/er/keeping_er/ ; *Managing electronic records issues. A discussion paper*, april 1998, p. 31-32.

applicatiesoftware die de bestanden vereisen. De technische ondersteuning die een archiefdienst kan bieden, is hoe dan ook beperkt. Anders dreigt de archiefdienst in de richting van een computermuseum te evolueren. De praktijk wijst evenwel uit dat er bij administraties informatiesystemen worden gebruikt met de meest uiteenlopende technische vereisten op het vlak van soort toepassingen (mainframe, server, pc), soort platformen (mainframe, WINNT, Unix, Novel, ...), ad hoc ontwikkelde applicatiesoftware. Dit heeft voor gevolg dat de bestanden die worden overgedragen, moeten aangepast worden aan de standaarden die de archiefdienst hanteert (migratie). Deze standaarden kunnen betrekking hebben op platformen, bestandsformaten, codetabellen en eventueel dragers en worden best in een lijst met archiveringsstandaarden vastgelegd. Deze lijst is samen met de beheersinventaris van de informatiesystemen een tweede belangrijk beleidsdocument voor het digitaal archiefbeheer. De lijst moet regelmatig aan de IT-evolutie worden aangepast en bevat eigenlijk de kwaliteitseisen waaraan de digitale archiefdocumenten bij overdracht moeten aan beantwoorden²¹. Een andere mogelijkheid is dat de archiefdienst in de mogelijkheid voorziet dat hun infrastructuur wordt aangepast zodanig dat de originele bestanden kunnen worden ondersteund (emulatie). Op die manier kunnen eventueel hardwareconfiguraties, besturingssystemen of zelfs applicaties worden gesimuleerd die normaal gezien niet tot de infrastructuur van de archiefdienst behoren.

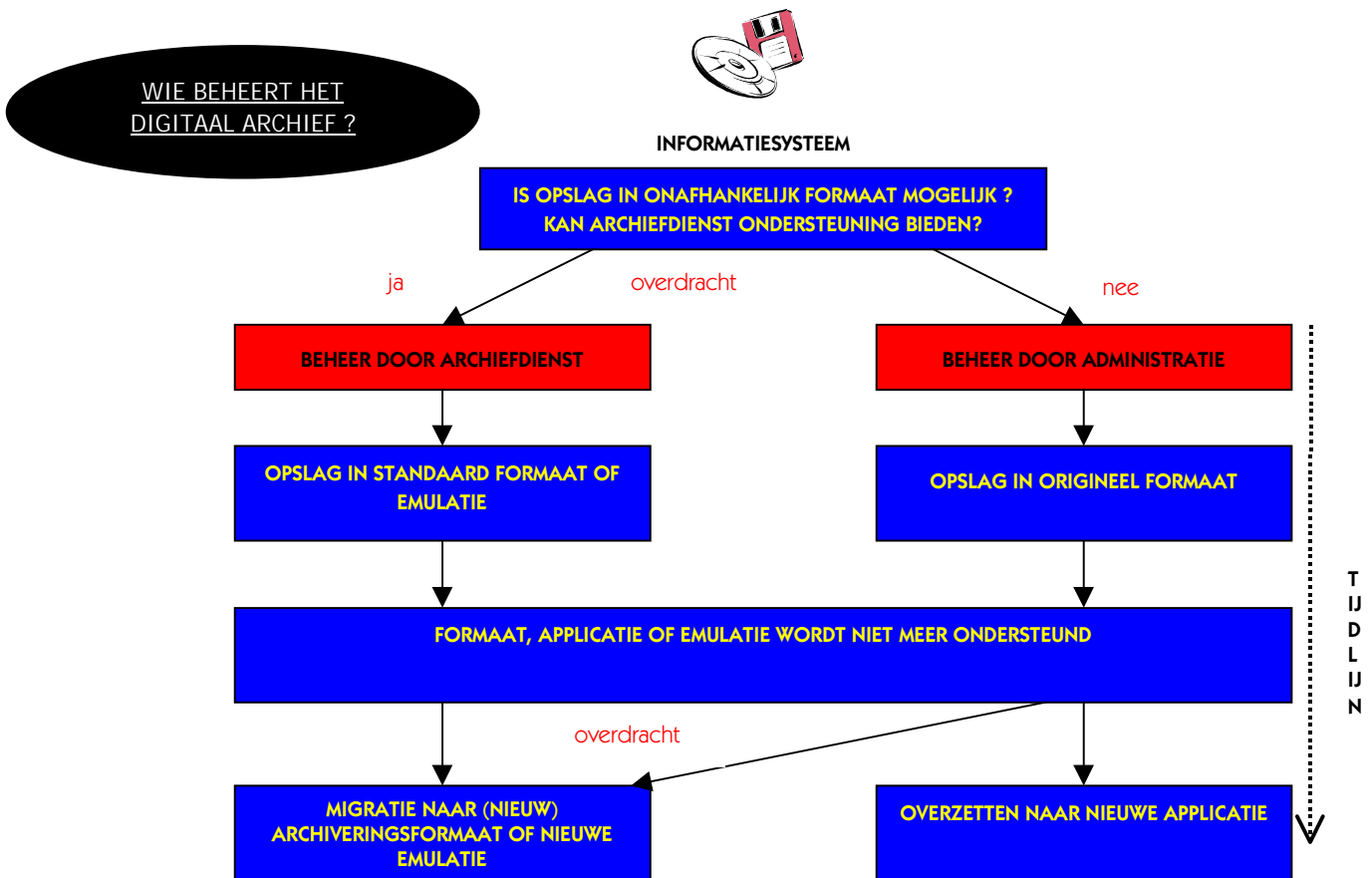
Het ligt voor de hand dat het voor de archiefdienst gemakkelijker is wanneer bij de administratie van de archiefvormer niet te veel informatiesystemen met uiteenlopende technische vereisten worden gebruikt, dan wanneer dat wel het geval is. Enige standaardisatie op dit vlak vergemakkelijkt de taak van de archiefdienst wanneer deze met de zorg over het digitaal archief wordt belast. De archivaris die betrokken is bij het ontwikkelen van een nieuw informatiesysteem moet hierop toezien. Het ideale ware natuurlijk dat standaarden in de informatiesystemen worden geïmplementeerd zodat de te archiveren bestanden met een minimum aan bewerkingen kunnen worden neergelegd.

Digitaal archiveren betekent ook dat de overdrachtswijze moet herbekeken worden. De archiefbescheiden kunnen overgedragen worden op een drager (diskette, cd-rom, dvd, tape, enz.). In dat geval moet de archiefdienst over de nodige apparatuur en drivers beschikken om de bestanden te kunnen inlezen. Een andere mogelijkheid is dat digitale bestanden over een netwerk worden getransfereerd. Hierdoor wordt het probleem van het kunnen lezen van de drager vermeden, al kan deze oplossing voor heel grote bestanden problemen opleveren. De problematiek van het uitbouwen van een digitaal depot hangt hiermee samen. De archiefdienst kan een fysiek gescheiden digitaal depot uitbouwen dat vergelijkbaar is met een apart domein. De fileservers en de jukeboxen moeten over de nodige opslagcapaciteit beschikken. Meer dan 90 % van de opstellingen zijn vandaag de dag DAS (Direct Attached Storage)-toepassingen waarbij schijven en RAID-systemen rechtstreeks aan servers en clients zijn verbonden. Voor de toekomst wordt inzake dataopslag meer gedacht aan de ontwikkeling van SAN- en NAS-toepassingen. In een SAN-configuratie (Storage Area Networks) zijn de schijven en RAID-systemen via een netwerk aan de servers verbonden. In een NAS-opstelling vormen de storage devices en de server één geheel (Network Attached Storage). Voor de archivering van digitale documenten lijkt een NAS-opstelling een meerwaarde te bieden omdat heterogene platformen data kunnen uitwisselen en bewaren op dezelfde NAS-server²². Binnen opstellingen zoals

²¹ Voorbeelden van dergelijke lijsten zijn: http://www.naa.gov.au/recordkeeping/er/keeping_er/append_a.html of bijlage 1 bij de *Ontwerp-Regeling geordende en toegankelijke staat archiefbescheiden 2000*.

²² G.-J. VAN BUSSEL, *Toekomst in opslag: NAS en SAN*, in: *Archos Magazine*, 7/8, 2000, p. 4; *Software Development Network & Storage Architecture Guide* en *Storage Architecture Guide* (<http://www.auspex.com>). Volgens Auspex, producent van NAS-servers en -technologie, zijn NAS-toepassingen te verkiezen boven SAN's omdat NAS zowel het file system van Sun Microsystems (NFS: Network File System) en van IBM en Microsoft (CIFS: Common Internet File System) ondersteunt. Dit is

NAS en SAN wordt er geen onderscheid gemaakt inzake fysieke bewaarplaats van dynamische en statische bestanden. Een overdracht naar de archiefdienst situeert zich hier eerder op het niveau van de toegekende gebruikersrechten of kan uitgevoerd worden door de gearchiveerde digitale bestanden achter de firewall van de archiefdienst te plaatsen. In een soortgelijk digitaal depot is de problematiek van de drager waarop digitale bestanden worden neergelegd niet aan de orde.



III.3.3 HOE archiveren we digitale archiefbescheiden?

Hoewel de archiefbeheerstaken voor digitale archiefbescheiden overdracht, beheer, ontsluiting en ter beschikking stelling omvatten, wordt de problematiek van digitaal archiefbeheer veelal toegespitst op de vraag in welk formaat de digitale informatie wordt opgeslagen en hoe het aan de archiefdienst wordt overgedragen. De manier waarop we digitale gegevens opslagen moet in principe aan een aantal criteria beantwoorden: leesbaar, gebruiksvriendelijk, onafhankelijk, duurzaam, betrouwbaar, enz. De

mogelijk omdat de applicatiesoftware (client) en het filesystem (server) van elkaar gescheiden zijn, wat in een SAN niet het geval is. In een NAS zijn de opslagapparaten rechtstreeks verbonden aan de servers, waardoor de I/O performantie ook hoger is.

keuze van het bestandsformaat is hier één van de belangrijkste zaken, tenzij voor emulatie als archiveringsstrategie wordt gekozen want dan worden de bestanden doorgaans in hun origineel formaat aan de archiefdienst overgedragen.

Het bestandsformaat waarin data wordt opgeslagen is afhankelijk van de applicatie waarbinnen het bestand is aangelegd of wordt geraadpleegd. De gebruikte applicatie wordt doorgaans gekozen op basis van haar functionaliteit. Bij het vastleggen van de archiveringswijze van informatie in digitale vorm moet eigenlijk een zelfde afweging worden gemaakt: Welke bewerkingen of handelingen moeten we nog met de data kunnen uitvoeren? Welke acties voert het systeem uit wanneer die informatie wordt opgevraagd? Het antwoord op deze vraag zal bepalen in welk formaat de data wordt gearchiveerd. Bij het onderzoeken van de functionaliteit van de gearchiveerde bestanden is het ook belangrijk om te onderzoeken of de bestanden statisch of dynamisch zijn en in welke mate ze afhankelijk zijn van bepaalde applicaties.

Tekstbestanden

Tekstbestanden waarvan de inhoud primeert kunnen als ASCII- of SGML/XML-bestand worden opgeslagen. Is de presentatie van het tekstbestand belangrijk dan kan het SGML/XML-document van een stylesheet worden voorzien of kan een bewaring als PDF-document²³ worden overwogen. Wanneer het tekstbestand daarentegen een aantal dynamische elementen bevat die functioneel moeten blijven (bv. macro, index, inhoudsopgave, links, OLE-object, enz.), dan kan er enkel een applicatieformaat worden gebruikt. Microsoft Corporation heeft voor de uitwisseling van tekstbestanden met opmaak Rich Text Format (RTF) ontworpen. De specificatie van RTF is open. Door een tekstbestand als RTF te bewaren, kunnen zowel de datastream als de opmaakgegevens naar een ander systeem worden overgezet. RTF wordt door veel applicaties ondersteund. Theoretici die het element vorm als één van de belangrijke kenmerken omschrijven, pleitten voor emulatie als archiveringsstrategie want bij migratie van een applicatieformaat naar ASCII gaat de originele vorm bijvoorbeeld verloren. In RTF wordt de lay-out doorgaans bewaard, al kunnen er wel verschuivingen optreden.

Audiovisuele bestanden

Bij audiovisuele bestanden moet men een tool voorzien waarmee de data correct kunnen worden gedecomprimeerd. Elk bestandsformaat waarin audiovisueel bestanden worden bewaard, gebruikt een bepaalde compressiemethode om de bestandsgrootte in de hand te houden en uitwisseling mogelijk te maken. Voor het openen van de bestanden moet de applicatie de overeenstemmende decompressie uitvoeren. Het bestandsformaat waarin digitale data wordt gearchiveerd en de drager waarop data wordt opgeslagen, moet door de archiefdienst worden ondersteund. Het omzetten naar een standaard of open formaat van applicatiegebonden bestanden heeft meestal voor gevolg dat er een aantal functionaliteiten of gegevens (bijv. kleuren) verloren gaan.

²³ De keuze voor PDF mag zeker niet worden gemotiveerd omwille van de vermeende “onwizigbaarheid” van de informatie. Er zijn immers voldoende applicaties voor handen die het mogelijk maken om de inhoud van een PDF-bestand te wijzigen of opnieuw te bewerken in een teksteditor (bijv. Photoshop, Ghostscript, PDF2RFT, PDF2TXT, BCLDrake, enz.)

Databanken

Hiërarchische databanken kunnen relatief gemakkelijk als flat file (ASCII, XML) worden gearcheveerd, zonder dat er informatieverlies is. Men moet er enkel over waken dat de structuur (root-parent-child) mee wordt gearcheveerd. XML is hier zeer geschikt voor. De archivering van relationele en object-geörienteerde databanken is gecompliceerder. Een relationele databanken in flat tables opslaan heeft voor gevolg dat de relaties tussen de tabellen, de index(en) en de sleutels verloren gaan. Bij object-geörienteerde databanken zou één object als één XML-element kunnen gedefinieerd worden. De historiek van het data-object kan één subelement vormen, die op zijn beurt uit meerdere elementen bestaat. Beide databanksystemen kunnen hun data als XML-bestanden wegschrijven en beheren, maar daarvoor moet er een softwarelaag boven de XML-bestanden worden geplaatst die de data transferreert en bewerkt. In die softwarelaag kunnen voor relationele databanken de sleutels, relaties en index(en) en voor object-geörienteerde databanken de interacties worden vastgelegd. Of de softwarelaag mee wordt bewaard, is afhankelijk van de functionaliteit van de gearcheveerde data. Queries worden bij voorkeur in SQL of een andere gestandaardiseerde querytaal vastgelegd.

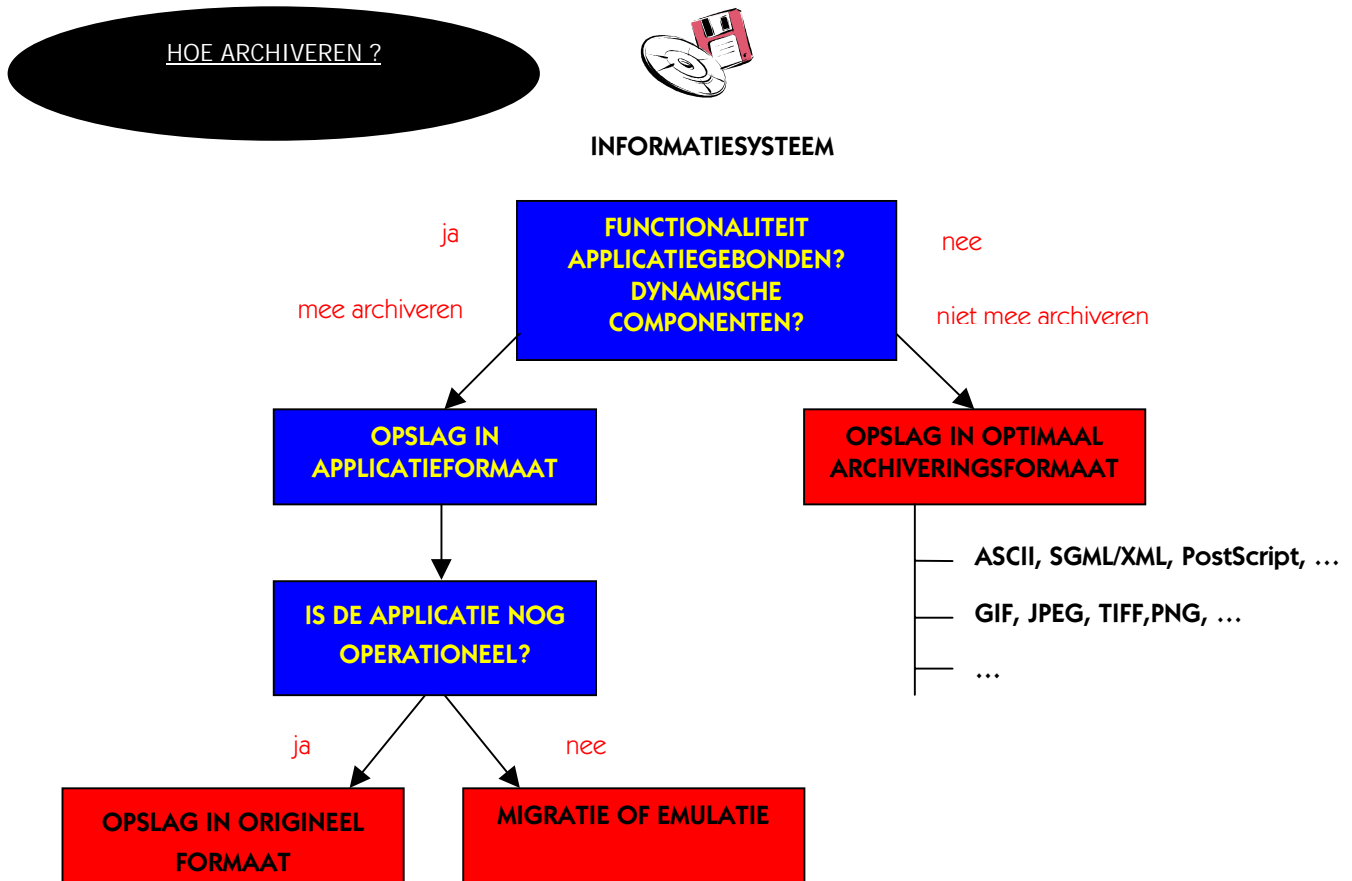
De keuze van het bestandsformaat wordt uiteindelijk bepaald door dezelfde twee factoren die bij het ontleden van een informatiesysteem centraal staan: wat is de functionaliteit van de gearcheveerde data en bevatten de bestanden nog dynamische elementen? In het geval van statische bestanden waarvan de functionaliteit door meerdere applicaties wordt ondersteund wordt er best een officiële of defacto standaard gekozen. De toekomstige beheersactiviteiten zullen dan voornamelijk migraties zijn. In de *custodial* traditie is het aan te raden dat de archiefdiensten een overzicht van standaard opslagformaten opstellen. Hierin sommen ze de formaten op die ze ondersteunen en waarin men de bestanden moet opslagen bij hun overdracht naar de archiefdienst.

Onder invloed van de toenemende impact van internet en andere netwerktoepassingen hebben hedendaagse informatiesystemen steeds meer een open en gescheiden structuur. De data van het informatiesysteem zijn bijvoorbeeld opgeslagen in SQL-databanken of XML-bestanden die vanuit verschillende applicaties raadpleegbaar zijn. Dit hangt samen met een duidelijke scheiding tussen de lagen van het informatiesysteem, wat archivering op lange termijn gemakkelijker maakt. De gelijkenis tussen de wijze waarop we momenteel digitaal archiefdocumenten migreren naar een archiveringsformaat en de architectuur van dergelijke open informatiesystemen is treffend. Wanneer de functionaliteit van deze gearcheveerde informatie beperkt blijft tot ontsluiting en raadpleging kunnen deze digitale data relatief gemakkelijk in een archiefbeheerssysteem worden opgenomen.

Wanneer bestanden daarentegen ingesloten programma's, macro's of andere actieve componenten bevatten, sterk platformafhankelijk zijn of hun specifieke originele functionaliteit moeten behouden, dan zal in de meeste gevallen de voorkeur worden gegeven aan de archivering binnen de oorspronkelijke omgeving. Tenzij er gemakkelijk kan gemigreerd worden naar een formaat dat beter geschikt is voor archivering, maar dit zal echter meer uitzondering dan regel zijn. In veel gevallen zal de keuze van bestandsformaat beperkt blijven tot het oorspronkelijke formaat (*native format*) of een meer geschikt archiveringsformaat dat niet alle elementen overneemt. Men moet er ook over waken dat samen met de gearcheveerde bestanden de passende launch of viewer mee wordt gearcheveerd (*logical encapsulation*²⁴). Vervolgens volgt men de emulatiepiste voor de viewer, want deze zal slechts

²⁴ Encapsulation kan ook betekenen dat alle structurele informatie met de datastream samen in één bestand wordt bewaard: physical encapsulation.

op een specifiek besturingssysteem kunnen functioneren. Het spreekt voor zich dat tussenoplossingen altijd mogelijk zijn of dat er overgeschakeld wordt van emulatie naar migratie, of omgekeerd.



De vraag hoe digitale archiefbescheiden worden gearchiveerd, heeft ook betrekking op de manier waarop ze aan het beheer van de archiefdienst worden toevertrouwd (*custodial*). Hiervoor zijn meerdere scenario's mogelijk, die in essentie enkel een kwestie van praktische organisatie zijn. De diensten kunnen de bestanden manueel neerleggen door ze op een drager te plaatsen en deze aan de archivaris te overhandigen. De archiefdienst legt vast op welke dragers het digitaal archief wordt neergelegd. Gezien het digitale karakter van deze archiefbescheiden kunnen ze ook worden doorgemailed of naar een andere locatie worden gekopieerd. Het probleem van de drager stelt zich hier niet, maar er komen wel een aantal andere aspecten bij kijken zoals de zekerheid over de identiteit van de afzender en de integriteit van de data.

III.3.4 WANNEER dragen we het digitaal archief over ?

Het tijdstip vanaf wanneer men normaliter papieren archiefbescheiden aan de archiefdienst kan overdragen is afhankelijk van de administratieve bewaartermijn. Na het verstrijken van de administratieve bewaartermijn kunnen de papieren bescheiden worden vernietigd of voor bewaring bij de archiefdienst worden neergelegd. In een digitale context kan het tijdstip van overdracht niet zomaar samenvallen met het einde van de administratieve bewaartermijn. Voor digitale archieven zal het tijdstip van overdracht in de eerste plaats worden bepaald door wie of welke dienst ze zal beheren. In de *non-custodial* opvatting zal overdracht op een later tijdstip plaats vinden dan in de *custodial* opvatting. Wanneer de archiefvormers zelf hun digitaal archief beheren zal een overdracht van bestanden overwogen worden wanneer ze niet meer compatibel zijn met nieuwe versies of toepassingen of wanneer er problemen rijzen in verband met de bestandsomvang en de performantie van het systeem. Op dit tijdstip zal wellicht eerst een selectie plaats vinden. Wanneer een nieuw systeem of een nieuwe versie in gebruik wordt genomen, moet de keuze gemaakt worden tussen migratie naar een softwareomgeving die de archiefdienst ondersteunt en tussen een migratie naar de nieuwe toepassing(versie). In bepaalde gevallen zullen de bestanden misschien wel voor vernietiging in aanmerking komen. Is de archiefdienst verantwoordelijk voor de zorg over de bestanden vanaf het ogenblik dat de administratieve bewaartermijn is verstreken, dan vindt de overdracht veel vroeger plaats.

Het tijdstip is ook afhankelijk van de aard van het digitaal archiefdocument en de manier waarop verouderde gegevens worden bijgehouden. Wordt het bestand voortdurend bijgewerkt of aangepast waarbij de oude informatie overschreven wordt en wordt het belangrijk geacht dat deze verouderde gegevens bewaard blijven, dan zullen ze waarschijnlijk met een bepaalde frequentie worden gearchiveerd. Op die manier zullen er meerdere versies (snapshots) worden bewaard. Dit kan bijvoorbeeld het geval zijn bij actieve databanken waarin informatie permanent wordt veranderd, aangevuld of verwijderd. Anders loopt men het gevaar dat men niet over momentopnames beschikt. De opslagcapaciteit moet voorzien zijn op de archivering van meerdere versie van hetzelfde bestand. Wanneer de inhoud van een bestand vastligt en niet meer verandert, dan hoeft in de meeste gevallen uiteindelijk slechts het origineel of de definitieve versie gearchiveerd te worden.

De toenemende verspreiding van object-geörienteerde informaticatoepassingen maakt het mogelijk dat historische informatie gemakkelijker in de informatiesystemen zelf kan worden beheerd. De historiek van een object wordt als het ware een apart object binnen het object zelf, zodat het niet meer nodig zal zijn meerdere versies van dezelfde digitale informatie te bewaren om de historische evolutie te schetsen. Op basis van de data waaruit de historiek van een object bestaat, kan een stand van zaken op een bepaald ogenblik in het verleden worden gereconstrueerd. Dit heeft voor gevolg dat historische informatie veel gemakkelijker in het oorspronkelijke informatiesysteem kan worden beheerd (*non-custodial opvatting*). Een voorbeeld van een dergelijke toepassing is bijvoorbeeld een digitale stadsplattegrond die voortdurend wordt bijgewerkt. In de plaats van met een bepaalde frequentie snapshots van de volledige plattegrond te nemen, wordt van elk object data bijgehouden. De gebruiker hoeft slechts een datum op te geven, en de plattegrond wordt samengesteld met alle objecten die toen in de stad stonden.

III.3.5 TOEPASSING: Wat? Hoe? Wanneer? Wie?

Bij de archivering van BEAM door het stadsarchief Antwerpen kwamen al deze vragen met betrekking tot de overdracht aan de orde. BEAM is de naam voor de applicatie waarmee de dienst bevolking van de bedrijfseenheid Burgerzaken de bevolkingshuishouding bijhoudt. De eigenlijke gegevens van het bevolkingsregister werden in een hiërarchische IMS-databank op mainframe bijgehouden. In december 1999 werd een nieuwe applicatie in gebruik genomen die gebaseerd is op een relationale DB2-databank. De gegevens uit de IMS-databank werden overgezet naar de DB2-databank. De gegevens van de mensen die vóór 1983 werden uitgeschreven, werden niet mee overgezet. Het betreft de gegevens van ca. 83000 mensen. De functionaliteit van deze gearchiveerde data blijft in de toekomst beperkt tot opvraging en raadpleging. Het archiveren van de oorspronkelijke applicatie (bovenste laag) is dus overbodig. Er moeten geen gegevens meer ingevoerd of gewijzigd worden, er moeten geen documenten meer worden gegenereerd, er moeten geen handelingen meer worden uitgevoerd en voor de raadpleging kunnen er andere tools (viewer, editors, browsers, XQuery-tool, enz.) worden gebruikt. De datastream is evenmin van de structurele laag (middelste laag) of andere informatiesystemen afhankelijk. Het lag dus voor de hand dat het bestand werd overgezet naar een platform dat door de archiefdienst kan worden ondersteund. Bovendien bouwde de informaticadienst het gebruik en de ondersteuning van mainframe IMS-databanken af. Uiteindelijk werd er voor een XML-bestandsformaat gekozen omdat het een vendor- en applicatieonafhankelijk standaardformaat is dat de semantiek en de data van het bestand samen archiveert. De migratie naar XML verliep in verschillende stappen. Om te beginnen moest er een file transfer worden uitgevoerd waarbij de karakters werden omgezet van EBCDIC-codering (mainframe) naar een ASCII-karakterset (server/pc). Het resultaat was een flat file (ASCII) waar de begin-en eindtags van de elementen werd aan toegevoegd. Op die manier werd het ASCII-bestand naar XML omgezet. Omwille van de grote bestandsomvang (ca. 350 megabyte) was het nodig om het mainframebestand in kleinere bestanden op te splitsen. De namen werden alfabetisch gesorteerd en alle namen die met dezelfde letter beginnen werden in één bestand geplaatst²⁵.

²⁵ Meer uitleg (o.a. gebruikte DTD en problematiek van diakritische tekens) is beschikbaar op de DAVID-website (cases → gegevens uit het bevolkingsregister).

IV. BESLUIT: ARCHIVEREN HERBEKEKEN

Voor digitaal archiefbeheer wordt best uit het informatiesysteem zelf vertrokken. Noch een typologie gebaseerd op de functie van de gegenereerde bescheiden, noch een typologie op basis van het type computerbestand of -toepassing kan gebruikt worden voor het uitstippelen van een archiveringsstrategie. Het type bestand (tekst-, spreadsheet en databankbestand) of toepassing (mainframe-, server- of clienttoepassing), de vorm, de functionaliteit en het bestandsformaat van een digitaal archiefdocument is variabel. De functie van een digitaal document in het bedrijfsproces ligt wel vast, maar kan niet gebruikt worden als basis voor de ontwikkeling van een archiveringsstrategie.

Voor de archivering van digitale bestanden zijn migratie en emulatie de opties die momenteel worden verkend en toegepast. Welke strategie het meest geschikt is, is afhankelijk van de karakteristieken van het informatiesysteem en van de toekomstige toepassingen met de gearchiveerde data. Om deze factoren te achterhalen onderscheidt men best de verschillende lagen van een informatiesysteem (datastream, structurele informatie, applicatie). Migratie richt zich in de eerste plaats op de bewaring van de datastream en indien nodig op de omzetting van de structurele informatie. Wanneer structurele informatie en/of tools moeten gearchiveerd worden, neemt het aandeel van de emulatiepiste toe. Een cruciale factor in de keuze tussen migratie en emulatie is de toekomstige actie die met de gearchiveerde data wordt uitgevoerd en de vraag of de functionaliteit door een onafhankelijk platform kan worden ondersteund. Aangezien dit voor elk informatiesysteem afzonderlijk moet worden bestudeerd, is het des te belangrijker dat er over elk informatiesysteem documentatie wordt bijgehouden. Van elk informatiesysteem moeten op zijn minst de gegevens worden bijgehouden die het mogelijk maken om de archiefwaarde ervan vast te leggen. Van de informatiesystemen met een archiefwaarde wordt vervolgens een uitgebreidere technische fiche bijgehouden (zie bijlage 1). De nodige metadata kunnen later uit deze beheersinventaris worden overgenomen.

Migratie en emulatie hebben elk hun voor- en nadelen. In de internationale literatuur spreken bepaalde auteurs zich duidelijk uit voor of tegen een bepaalde strategie. Voor een archiefdienst die de meest uiteenlopende informatiesystemen moet archiveren, is een dergelijke discussie niet relevant. Beide strategieën zijn geschikt voor de archivering van digitale informatie uit een bepaald informatiesysteem. Algemeen kan worden gesteld dat migratie kan toegepast worden wanneer de inhoud van een digitaal bestand vast ligt en wanneer de functionaliteit door meerdere applicaties kan worden ondersteund. Bij migratie primeert de opslag in een zo optimaal mogelijk archiveringsformaat. In de gevallen waarin de digitale informatie afhankelijk is van een specifieke applicatie en waarbij dynamische eigenschappen mee moeten gearchiveerd worden, wint emulatie aan belang.

Bij digitale archivering gaat het niet enkel om het duurzaam vastleggen van de bitreeksen die de data van een informatiesysteem bevatten. Die data worden binnen een informatiesysteem gevormd. Over dit informatiesysteem en de data zelf moeten de nodige gegevens worden bewaard, want deze zijn niet of nauwelijks af te leiden uit de digitale archiefdocumenten zelf. In die zin worden er niet louter archiefdocumenten gearchiveerd, maar ook informatie.

Doordat de digitale archivering niet op het document is gebaseerd, moeten ook nog andere archiefbeheerstaken zoals beschrijven, bewaren in goede staat, selectie en ter beschikkingstelling opnieuw bekeken worden. Beschrijven van digitale bestanden is op basis van de huidige

beschrijvingstechniek en -terminologie slechts evident in de gevallen waarin de digitale informatie een papieren equivalent heeft. Een tekstverwerkingsbestand dat een brief of jaarverslag bevat, kunnen we op dezelfde manier beschrijven als een papieren brief of jaarverslag. Het enige wat dan in de beschrijving moet aangepast worden is zijn materiële vorm met eventueel de aanvulling van een aantal technische gegevens. Beschrijven wordt moeilijker wanneer er geen papieren equivalent bestaat, een heel informatiesysteem met meerdere types documenten wordt gearcheveerd of wanneer er geen documenten worden gegenereerd en het systeem enkel gericht is op het beheren van informatie. Een mogelijke oplossing voor deze gevallen is een beschrijving op basis van het type bestand (tekst-, spreadsheet-, databank-, grafisch-, muziek-, videobestand) met daaraan gekoppeld een inhoudelijke omschrijving. Het werkproces zal een belangrijk onderdeel zijn in de beschrijving van een informatiesysteem, wat opnieuw het belang van dit gegevensveld in een beheersinventaris van digitale informatiesystemen onderstreept.

Bij selectie van digitale bestanden moet ook de vraag worden gesteld of het technisch wel mogelijk is om (dat gedeelte van) de digitale informatie te vernietigen zonder het informatiesysteem niet meer functioneert of het permanent te bewaren gedeelte verder kan worden geraadpleegd. Papieren bescheiden zijn fysiek van elkaar gescheiden, terwijl digitale informatie dikwijls aan elkaar is gekoppeld. Vooraleer over te gaan tot selectie komen dezelfde vragen aan de orde als bij de keuze van wat er wordt bewaard. Dit zal voor gevolg hebben dat in een digitale context selectie meer op het niveau van het informatiesysteem zal worden toegepast. Voor archiefdocumenten afkomstig uit hetzelfde digitale informatiesysteem zal men zelden een verschillende bewaartermijn kunnen hanteren. Bijgevolg zullen de selectielijsten een aanpassing moeten ondergaan.

Tenslotte moet men ook rekening houden met het feit dat gearcheveerde digitale bestanden slechts ter beschikking kunnen worden gesteld door middel van hard- en software. De leeszaal moet met de nodige apparatuur en programmatuur worden uitgerust. Naast de gearcheveerde bestanden moeten ook de geschikte zoek-en viewapplicaties worden voorzien. Wanneer de archiefvormer zijn digitale informatie beheert, moet in de mogelijkheid worden voorzien dat onderzoekers daartoe toegang hebben.

Bij digitaal archiefbeheer is een beleidsmatige en gestructureerde aanpak des te belangrijker. In dit rapport werden drie kapstukken voor een digitaal archiefbeleid aangereikt. In de beheersinventaris wordt een overzicht van de digitale informatiesystemen bijgehouden waarin de nodige metadata worden vastgelegd en waaruit de archiefwaarde kan worden afgeleid. In elk informatiesysteem kunnen drie lagen worden onderscheiden. Op basis van het beslissingsmodel weet men welke informatielagen worden gearcheveerd, wie het digitaal archief beheert, wanneer de overdracht plaats vindt en hoe er wordt gearcheveerd.

BIJLAGE 1: MODELICHE INFORMATIESYSTEEM

<p><u>NAAM</u> Naam van de applicatie (+ betekenis afkorting)</p>	
<p><u>DOEL</u> Omschrijf het doel en het werkproces waarbinnen het systeem wordt gebruikt.</p>	
<p><u>FUNCTIES</u> Hoe functioneert het informatiesysteem? Wat zijn de voornaamste functies? Welke dynamische componenten bevatten de computerbestanden?</p>	
<p><u>AFHANKELIJKHEDEN: DATA</u> Is het systeem gekoppeld aan één of meerdere andere systemen? Welke?</p> <p><u>AFHANKELIJKHEDEN: TOOLS</u> Is het systeem afhankelijk van bepaalde tools? Wat is de functie van de tools? Zijn er andere tools die deze functionaliteit ondersteunen? Welke?</p>	
<p><u>PLATFORM</u> Wat is de hardware en besturingssysteem van het informatiesysteem?</p>	

<p><u>HISTORIEK</u></p> <p>Voorgaande informatiesystemen?</p> <p>Datum van - ingebruikname? - afsluiting/overzetting?</p> <p>Versies?</p> <p>Migraties van de digitale informatie?</p> <p>Opvolgende informatiesystemen?</p>	
<p><u>DIENSTEN</u></p> <p>Welke diensten beheren/werken met het informatiesysteem?</p>	
<p><u>INFORMATIE</u></p> <p>Welke digitale informatie wordt door het systeem gegenereerd of beheerd?</p> <p>Openbaarheid van de informatie?</p>	
<p><u>BESTANDEN</u></p> <p>Welke opslagformaten worden er gebruikt?</p> <p>Volume?</p>	
<p><u>FUNCTIONALITEIT VAN DE GEARCHIVEERDE INFORMATIE</u></p> <p>Welke handelingen of bewerkingen moeten bij raadpleging mogelijk zijn?</p>	
<p><u>ARCHIEFWAARDE</u></p> <p>Wat is de archiefwaarde van de digitale informatie?</p>	

BIJLAGE 2: CHECKLIST DIGITAAL ARCHIVEREN

Aan de archivering van digitale archiefdocumenten gaat een heel beslissingsproces vooraf. Een checklist van voornaamste aandachtspunten kan een interessant hulpmiddel zijn bij de voorbereiding van de digitale archivering of het opzetten van een digitaal archiveringssysteem. Hieronder sommen we een aantal elementen op waarmee rekening moet worden gehouden bij:

➤ Ontwikkeling en gebruik van digitale informatiesystemen

- ✓ Wordt er op toegezien dat er zoveel mogelijk standaarden worden geïmplementeerd bij de ontwikkeling van digitale informatiesystemen?
- ✓ Hebben de informatiesystemen een open structuur? Kunnen data en structurele informatie van elkaar worden gescheiden?
- ✓ Wordt de archiefdienst op de hoogte gebracht van de ontwikkeling van nieuwe informatiesystemen of de aanpassing van bestaande informatiesystemen (bijv. meldingsplicht)? Wordt de archiefdienst geraadpleegd?
- ✓ Wordt het beheer van de digitale informatiesystemen gedocumenteerd (metadata)? Wordt er een overzicht van de digitale informatiesystemen bijgehouden?
- ✓ Is er bij de diensten een richtlijn over het omgaan met digitale (archief)documenten (opslag, vernietiging, wijziging, enz.)? Is er duidelijk vastgelegd wat archief is en wat geen archief is?

➤ Digitale archivering door de archiefvormer:

- ✓ Kan het digitale informatiesysteem de gearcheeerde informatie beheren, ontsluiten en ter beschikking stellen? Worden hiervoor bijzondere functionaliteiten of modules voorzien? Is selectie mogelijk?
- ✓ Wie is er verantwoordelijk over de gearcheeerde digitale archiefdocumenten?
- ✓ Zijn de digitale archiefdocumenten door archiefgebruikers consulteerbaar?
- ✓ Zijn er maatregelen genomen om de authenticiteit en integriteit te waarborgen?
- ✓ Is er een procedure voorzien in het geval het informatiesysteem aanpassingen ondergaat?
- ✓ Is er een procedure voor het maken en beheren van veiligheidskopieën?
- ✓ Zijn er afspraken gemaakt over het tijdstip van neerlegging (evt. afhankelijk van bestandsomvang, upgrades, wegvallen ondersteuning, ...)?

- ✓ Worden er snapshots genomen wanneer de oude gegevens systematisch worden overschreven? Met welke periodiciteit worden er snapshots genomen? Worden de snapshots aan de archiefdienst overgedragen?

➤ Digitale archivering door de archiefdienst?

- ✓ Is er een overzicht van de kwaliteitseisen waaraan de neergelegde digitale archiefdocumenten moeten voldoen (bestandsformaat, drager, bestandensysteem, bestandsomvang, enz.)?
- ✓ Beschikt de archiefdienst over de nodige hard -en software?
- ✓ Zijn de digitale archiefdocumenten bewaard in een bestandsformaat en op een drager die door de archiefdienst wordt aanvaard? Is er een emulatieprogramma voorzien? Vormt de bestandsgrootte geen probleem voor server, pc of netwerk?
- ✓ Zijn de nodige queries, stylesheets en logfiles neergelegd? Moeten er toegangen worden voorzien?
- ✓ Is er een procedure voor de neerlegging?
- ✓ Zijn de vereiste metadata aanwezig?
- ✓ Worden de migraties gedocumenteerd?
- ✓ Beschikt de archiefdienst of -instelling over een gebruikershandleiding in het geval van niet-courante informatiesystemen?
- ✓ Kunnen gecomprimeerde bestanden correct worden gedecomprimeerd? Is de vereiste software hiervoor aanwezig?
- ✓ Is de archiefdienst of -instelling in het bezit van de nodige softwarelicenties?
- ✓ Kan de gearhiveerde digitale informatie in het archiveringssysteem worden opgenomen?
- ✓ Zijn er maatregelen genomen om de authenticiteit en integriteit te waarborgen?
- ✓ Is er een procedure voor het maken en beheren van veiligheidskopieën? Worden de digitale bestanden regelmatig naar een andere drager overgezet (refreshen)?