



Universiteit Twente
de ondernemende universiteit



Digitale duurzaamheid van archieven in het CDD+

Digital preservation of archives in the CDD+

Auteur:	R.H. Berentsen	Datum:	28-09-2007
Studentnummer:	0023434		
Status:	Definitief	Versie:	2
In opdracht van:	Justitiële Informatiedienst (JustID)		
Bedrijfsbegeleiding:	Drs. D. Venema (JustID) Ing. E.C. Groen (Deloitte)	Opleidingsinstelling:	Universiteit Twente Drienerlolaan 5
Afstudeercommissie:	Dr. A.B.J.M. Wijnhoven Ir. H.G.A. Middel	Faculteit:	7522 NB Enschede Management & Bestuur
		Opleiding:	Technische Bedrijfskunde

Samenvatting

Dit onderzoek heeft tot doel in het kader van digitale duurzaamheid van archieven in het CDD+, het Expertise Centrum CDD+ als organisatorisch onderdeel van JustID en binnen de Strafrechtsketen in te richten en te voorzien van ondersteunende informatiesystemen, zodat zij voorbereid is en blijft op veranderingen in technologie, wet- en regelgeving.

De veranderende technologie, wet- en regelgeving refereren naar het vraagstuk rondom digitale duurzaamheid van informatie. Steeds meer informatie wordt opgeslagen in digitale computerbestanden (gecodeerd in digitale bestanden). Door vanuit deze digitale bestanden verschillende lagen te doorlopen, zoals applicaties, besturingssystemen en hardware, wordt er gekomen tot de presentatie van de informatie aan raadplegers. Ten gevolge van veranderingen in deze lagen, door bijvoorbeeld veranderende technologie, loopt digitale informatie het risico voor altijd te worden verminkt of zelfs geheel verloren te gaan.

De vraag “hoe” de digitaal gecodeerde informatie te bewaren speelt ook in het kader van de ontwikkelingen rondom het Centraal Digitaal Depot (CDD) van JustID. Zij wil digitale informatie, conform de in de archiefwet gestelde eisen, duurzaam opslaan om deze op een willekeurig later tijdstip te kunnen representeren aan daartoe gerechtigden.

Deze scriptie is het resultaat van onderzoek naar de volgende onderdelen:

- Twee overkoepelende bewaarstrategieën voor digitale informatie (migratie en emulatie);
- Criteria categorieën die van invloed zijn op het bewaren van en omgaan met digitaal gecodeerde informatie;
- Het selecteren van een bewaarstrategie aan de hand van een bewaarstrategie selectiemethode;
- Het ten uitvoer brengen van bewaaracties voortvloeiende uit een geselecteerde bewaarstrategie.

Daarbij is onderzocht wat de eigenschappen zijn van de twee overkoepelende bewaarstrategieën en hoe deze in de praktijk toegepast worden. Daarnaast is vanuit theorie en casestudies onderzocht welke criteria categorieën gehanteerd kunnen worden bij het evalueren van bewaarstrategieën. De bewaarstrategie selectiemethode voorziet daarbij in een stappenplan waarbinnen de concrete criteria opgesteld worden, deze vervolgens invulling krijgen vanuit experimenten. De resultaten hiervan worden gebruikt om met behulp van een aangepaste utiliteitsanalyse te berekenen wat de meest geschikte bewaarstrategie is voor de te bewaren digitale informatie.

Uit het onderzoek blijkt de noodzaak de bewaarstrategie-evaluatiemethode en de uitvoer van zowel migraties als emulaties als bewaarstrategie te faciliteren. Daartoe wordt de organisatie ingericht om te voorzien in de uitvoer van de eerder geïdentificeerde processen en wordt de architectuur ontworpen die deze organisatie ondersteunt.

Acties die JustID moet uitvoeren naar aanleiding van dit onderzoek is het inbedden van de bewaarstrategie-evaluatiemethode binnen de organisatie en haar technische infrastructuur. Daartoe beschrijft deze scriptie de benodigde processen, taken en architectuuronderdelen. De individuele functies die de taken binnen de processen uitvoeren moeten verder worden gespecificeerd en uitgediept.

Zodra de methode om de meest geschikte bewaarstrategie te kunnen selecteren is geïmplementeerd, dienen de daarmee te selecteren bewaarstrategieën te worden geoperationaliseerd. Daartoe moet JustID een omgeving opzetten waarbinnen migraties en emulaties uitgevoerd kunnen worden, zowel om experimenten met bewaarstrategieën uit te kunnen voeren, als om bewaaracties in productie te

kunnen brengen.

JustID moet zich gaan profileren als instantie die kennis heeft over het omgaan met digitale duurzaamheid en actief wil deelnemen aan onderzoek rondom het bieden van digitale duurzaamheid. Daartoe dient zij aansluiting te vinden bij en bij te dragen aan onderzoeksconsortia en dienstverleners. Zij hoeft namelijk niet alle geconcludeerde acties zelfstandig te ondernemen. In Europees verband beoogd het PLANETS onderzoeksconsortium gedistribueerd services te leveren die grotendeels aansluiten bij wat JustID moet realiseren. Door samen te werken kan zij daadwerkelijk de duurzaamheid van de in het CDD opgeslagen digitale informatie garanderen richting haar ketenpartners.

Voorwoord

Archiveren wordt vaak gezien als een black-box waar dossiers in verdwijnen onder een laag stof; een noodzakelijke kostenpost om het primaire proces af te ronden. Niets is echter minder waar! In het voortschrijdende digitale tijdperk wordt steeds vaker digitaal gewerkt en vindt er digitale neerslag van processen plaats in de vorm van bijvoorbeeld documenten. Het digitaal duurzaam bewaren van deze archieven en het digitaal weer beschikbaar stellen binnen de Strafrechtsketen, maakt het onderwerp en de context van mijn onderzoek een zeer uitdagende! Naast dit uitdagende onderwerp heb ik tijdens mijn onderzoek intensief kennis mogen maken met twee organisaties: De Justitiële Informatiedienst (JustID) en Deloitte Consulting.

Mijn onderzoek naar digitale duurzaamheid van archieven is uitgevoerd in het kader van mijn afstuderen voor de studie Technische Bedrijfskunde. De scriptie die voor u ligt is het resultaat van deze periode. Dit resultaat heb ik natuurlijk niet alleen bereikt. Met name wil ik de volgende mensen danken voor hun bijdrage aan mijn onderzoek.

Dany Venema – begeleider vanuit JustID – wil ik bedanken voor het mij aanmoedigen niet alleen initiatief naar hem te tonen, maar ook het initiatief eigenhandig te nemen. Dit resulteerde in meerdere informatieve bijeenkomsten in Den Haag, het bezoek van een congres in Frankfurt Am Main en veelvuldige mailconversaties met onderzoekers op het gebied van digitale duurzaamheid.

Anita, Esther, Maartje en Miranda – collega's binnen JustID – wil ik bedanken voor de gezellige tijd. Of er nou wel of niet kwarktaart, spekjes of andere kleurstofrijke producten waren: het was altijd gezellig.

Edwin Groen – begeleider vanuit Deloitte – las altijd kritisch mee, wat tot scherpe verbeteringen in de vorm van mijn scriptie heeft geleid.

Jeroen Dominicus – medewerker van Deloitte – ben ik dankbaar voor allerlei technische uitleg met betrekking tot het archiefsysteem waarop mijn onderzoek zich richt en het met mij delen van zijn ervaringen als consultant.

Fons Wijnhoven en Rick Middel – mijn afstudeerbegeleiders vanuit de Universiteit Twente – wil ik bedanken voor het keer op keer van commentaar voorzien van mijn onderzoeksopzet en scriptie, hun wijze van adviseren, voeren van discussies en altijd positieve instelling.

Ook mijn ouders, broer, vriendin en vrienden wil ik bedanken voor alle hulp en plezier tijdens mijn gehele studie en afstuderen.

Ik wens iedereen veel plezier met het lezen van deze scriptie,

Rob Berentsen

Almelo, 20-09-2007

Inhoudsopgave

Leeswijzer	1
Hoofdstuk 1 De context.....	2
1.1. De Strafrechtsketen	2
1.2. JustID - dienstverlener voor de Strafrechtsketen	3
Hoofdstuk 2 Digitale duurzaamheid en het CDD+	9
2.1. Digitale duurzaamheid	9
2.2. Digitale informatie	9
2.3. Records Continuüm en Archiefbeheer in het CDD+	14
2.4. Probleemstelling en onderzoeksvragen	20
2.5. Afbakening.....	22
2.6. Onderzoeksplan.....	22
Hoofdstuk 3 Bewaarstrategieën	25
3.1. Migratie	25
3.2. Emulatie	36
3.3. Conclusie.....	43
Hoofdstuk 4 Criteria	44
4.1. Archiefstukeigenschappen	45
4.2. (De)codeereigenschappen	45
4.3. Bewaaractieproces.....	47
4.4. Kosten	48
4.5. Conclusie.....	49
Hoofdstuk 5 Selectie van geschikte bewaarstrategie.....	50
5.1. Definieren van de basis en de criteria	50
5.2. Alternatieve bewaarstrategieën evalueren.....	52
5.3. Resultaten vergelijken en interpreteren.....	53
5.4. Conclusie.....	55
Hoofdstuk 6 Organisatie-inrichting.....	56
6.1. Services	56
6.2. Processen.....	57
6.3. Taken en afdelingen	61
6.4. Gegevensdragers	65
6.5. Conclusie.....	66
Hoofdstuk 7 Architectuur.....	67
7.1. Architectuuronderdelen	67
7.2. Architectuur.....	70
7.3. Conclusie.....	71
Hoofdstuk 8 Conclusies en vervolgonderzoek	72

8.1. Conclusies en aanbevelingen	72
8.2. Discussie	73
8.3. Vervolgonderzoek	75
Literatuur	77
Boeken/artikelen	77
Websites	80
Bijlage 1 Verklarende woordenlijst	1
Bijlage 2 Basisvoorzieningen Strafrechtsketen.....	4
Bijlage 3 Informatietypen	5
Bijlage 4 Entiteiten metadatamodel CDD+.....	10
Bijlage 5 Informatie pakket / Archival information package (AIP).....	12
Bijlage 6 Bestandsformaat herkenning en inhoudbepaling.....	13
Bijlage 7 Modulaire emulator	19
Bijlage 8 Voorbeelden object bomen	21
Bijlage 9 PLATO	23
Bijlage 10 Inrichting digitale duurzaamheid CDD+	24
Bijlage 11 Reflectie	52

Leeswijzer

Deze scriptie bestaat uit 8 hoofdstukken die het volgende beschrijven:

Hoofdstuk 1 gaat in op de context waarin het onderzoek is uitgevoerd. Hier worden de omgeving van de Justitiële Informatiedienst en haar Centraal Digitaal Depot (CDD) beschreven. Vervolgens begint hoofdstuk 2 met contextuele onderwerpen binnen het onderzoek en eindigt met de opzet van het onderzoek.

Hoofdstuk 3 beschrijft de overkoepelende bewaarstrategieën en hierbij wordt vanuit de praktijk beschreven wat er technisch gerealiseerd moet worden om deze bewaarstrategieën ten uitvoer te kunnen leggen. Hoofdstuk 4 gaat vervolgens in op criteria die ten grondslag liggen aan de duurzaamheid van digitale informatie. In hoofdstuk 5 wordt uitgelegd hoe methodisch te bepalen is welke bewaarstrategie voor digitale informatie de meest geschikte is. Daartoe wordt de bewaarstrategie-evaluatiemethode uitgewerkt.

Hoofdstuk 6 doet een voorstel voor de inrichting van de organisatie van het CDD+ om onder andere de in hoofdstuk 5 beschreven bewaarstrategie-evaluatiemethode te implementeren. Er wordt in dit hoofdstuk dan ook beschreven welke processen de organisatie moet gaan uitvoeren, welke taken hier in voorkomen en welke afdelingen de taken ten uitvoer moeten brengen in het kader van het omgaan met digitale duurzaamheid.

Hoofdstuk 7 gaat in op de architectuur die de in hoofdstuk 6 beschreven organisatie zal moeten gebruiken om haar taken te kunnen volbrengen: het garanderen van de digitale duurzaamheid van archieven in het CDD.

Hoofdstuk 8 bevat de uit het onderzoek te vormen conclusies en aanbevelingen. Daarnaast wordt er een discussie gevoerd over de relevantie van het investeren in digitale duurzaamheid en eindigt het hoofdstuk met enkele voorstellen voor vervolgonderzoek.

De uitleg van complexe termen en begrippen is te vinden in de verklarende woordenlijst (bijlage 1).

Hoofdstuk 1 De context

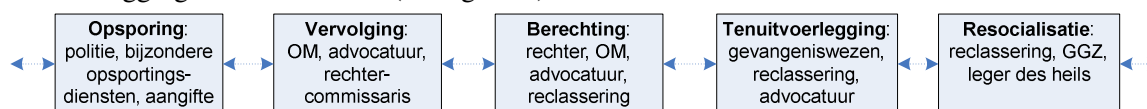
Dit onderzoek wordt uitgevoerd bij de Justitiële Informatiedienst (JustID) in het kader van digitale duurzaamheid van archieven. In dit inleidende hoofdstuk worden daarom de Strafrechtsketen, JustID en digitale duurzaamheid van archieven kort toegelicht om een beter beeld te krijgen van de context waarin dit onderzoek is uitgevoerd.

1.1. De Strafrechtsketen

Wie in ons land een overtreding begaat, weet dat hij of zij de kans loopt gestraft te worden. Voordat het zover is, komt een (justitie)apparaat in werking: de Strafrechtsketen. Ketens worden vanuit keteninformatisering [Grij06] beschreven als een netwerk van grote aantallen onafhankelijke organisaties en professionals die in constant wisselende combinaties samenwerken om een (maatschappelijk) product voort te brengen. Zo beschrijft Grijpink dat de Strafrechtsketen met meer dan duizend organisaties, beoogt bij te dragen aan een veiliger samenleving [Grij06].

1.1.1. De keten op hoofdlijnen

Op hoofdlijnen is de Strafrechtsketen onder te verdelen in opsporing, vervolging, berechting, tenuitvoerlegging en resocialisatie (zie figuur 1).



Figuur 1 - Overzicht strafrechtsketen

Opsporing

Allereerst moet geconstateerd worden dat er een overtreding is begaan, bijvoorbeeld door een opsporingsambtenaar (politieagent), of een aangifte door een burger. In het laatste geval wordt de politie ingeschakeld die op onderzoek uitgaat en probeert de dader op een later tijdstip alsnog op te pakken. Als de verdachte is opgepakt, is het eerste deel van de opsporing verricht. De politieagent stelt een procesverbaal op waarin hij een feitelijke weergave doet van wat zich heeft voorgedaan.

Vervolging

Het proces-verbaal gaat naar de Officier van Justitie die belast is met vervolging van de verdachte. Hij kan terugvallen op het opsporingsapparaat als hij aanvullende informatie nodig heeft. Vervolgens besluit de officier of hij de zaak voor de rechter brengt. Hij zal hierbij afwegen of er volgens hem voldoende bewijsmateriaal is verzameld om de verdachte veroordeeld te krijgen. Is dat niet het geval, dan zal hij besluiten de zaak te seponeren: er zal geen vervolging plaatsvinden.

Berechting

Wanneer de officier van mening is dat er voldoende bewijs is, wordt de zaak voor de rechter gebracht en volgt de berechting. De rechtbank organiseert dan een zitting waarbij de officier, de verdachte en een of meerdere advocaten die de verdachte bijstaan verschijnen. De rechter beoordeelt de informatie die de officier van Justitie aanlevert en kan hier vragen over stellen. Bovendien kan hij de verdachte(n) ondervragen en eventuele getuigen horen. Eveneens is het mogelijk om deskundigen te raadplegen om een goed beeld te krijgen van wat zich exact heeft afgespeeld (waarheidsvinding). De rechter bepaalt uiteindelijk ook de straf. Dat kan een boete zijn, een vrijheidsstraf of wellicht vrijspraak, als de rechter van mening is dat er onvoldoende bewijs is dat de verdachte daadwerkelijk de dader is.

Tenuitvoerlegging

Nadat er een straf is uitgesproken volgt de tenuitvoerlegging. Bij een geldboete is dat de inning, bijvoorbeeld door het sturen van een acceptgiro. Heeft de rechter geoordeeld dat een vrijheidsstraf moet worden opgelegd, dan kan de verdachte deze meteen of op een later tijdstip moeten ondergaan. Daarnaast is het ook mogelijk dat de rechter een taakstraf of behandeling oplegt.

Resocialisatie

Na het uitzitten van een eventuele vrijheidsstraf dient de veroordeelde weer goed in de samenleving te functioneren. De reclassering speelt hierin een rol en begeleidt de veroordeelde bij terugkomst in de maatschappij.

1.1.2. Ketensamenwerking

Een belangrijke randvoorwaarde voor ketensamenwerking is dat de informatie-uitwisseling van elk van de ketenpartners op elkaar is afgestemd. Op zich is ketensamenwerking in de Strafrechtsketen niet nieuw; al vele decennia vindt tussen de ketenpartners gestructureerd uitwisseling van informatie plaats. De steeds hoger wordende eisen die de samenleving stelt aan het presteren van de Strafrechtsketen leiden echter tot toenemende prestatiedruk. Omgang met deze druk wordt belemmerd door enkele eigenschappen van ketens. Zo stelt Grijpink [Grij06] dat overkoepelend gezag in ketens in algemene zin ontbreekt en de gemeenschappelijke belangen vaak beperkt en onduidelijk zijn: ketens gedragen zich irrationeel en onvoorspelbaar op ketenniveau. De ketensamenwerking groepeerde zich namelijk rondom een zogenaamd “dominant ketenprobleem”. In dit ketenprobleem is het probleem dat één enkele ketenpartner het niet alleen kan oplossen: er moet worden samengewerkt. Als het dominante ketenprobleem eenmaal is opgelost, hergroepeerde een nieuwe samenstelling van ketenpartners zich rondom een nieuw dominant ketenprobleem. Dominante ketenproblemen in de strafrechtsketen, zoals het “hoe een integer en integraal persoonbeeld te verkrijgen?”, hebben geleid tot diepgaande samenwerking tussen de verschillende ketenpartners.

Vanuit deze samenwerking zijn concrete koppelingen en informatiesystemen opgezet binnen de Strafrechtsketen, zoals het VerwijsIndex Personen (VIP) en het Justitieel Documentatie Systeem (JDS) [10] (zie bijlage 2). In VIP staan verwijzingen naar ketenpartners die met natuurlijke- en rechtspersonen bezig zijn (zgn. justitiabelen). Het systeem bevat geen zaakgegevens, maar alleen verwijzingen naar de systemen waar zaakgegevens kunnen worden gevonden. Het JDS bevat strafrechtelijke gegevens over justitiabelen, zoals de feiten die op een stafblad staan. JustID heeft onder andere deze systemen onder haar hoede en onderneemt nog meer ketenoverkoepelende projecten. Één daarvan is de ontwikkeling van een Centraal Digitaal Depot (CDD) dat digitale archivering voor de Strafrechtsketen mogelijk moet maken. JustID en haar dienstverlening wordt verder beschreven in de rest van dit hoofdstuk.

1.2. JustID - dienstverlener voor de Strafrechtsketen

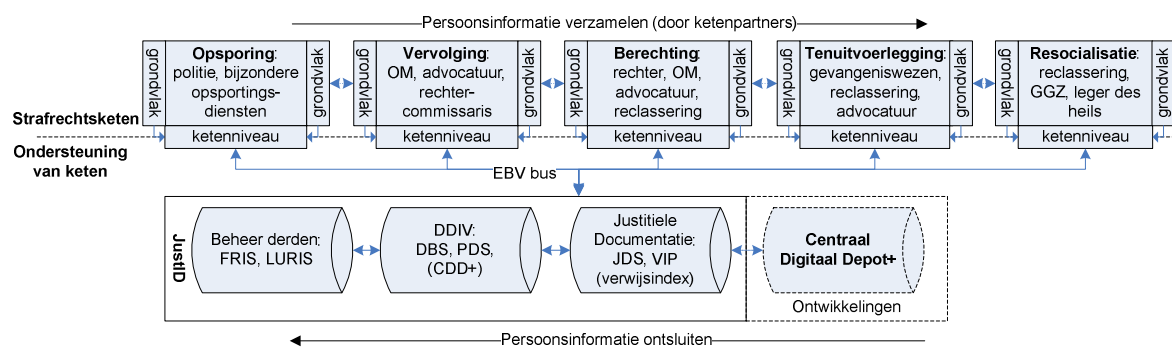
JustID bestaat sinds 1 januari 2006 en heeft als primaire taak het verstrekken van een integer en een integraal persoonsbeeld van justitiabelen aan daartoe gerechtigden. Gerechtigden zijn veelal ketenpartners in de Strafrechtsketen. Datgene wat integer en integraal kan zijn aan een persoonsbeeld wordt toegelicht in de volgende paragraaf. Om dit persoonsbeeld samen te stellen, worden in samenwerking met verschillende instanties binnen de Strafrechtsketen, zoals het Openbaar Ministerie, gegevens op een efficiënte en effectieve manier verzameld, bewerkt en beheerd.

JustID speelt hierin een faciliterende rol, die met name de samenwerking tussen de leveranciers en gebruikers van informatie bevordert. Een adequate informatievoorziening is hiervoor een essentiële randvoorwaarde [Delo06]. Meer specifiek betekent dit, dat de afzonderlijke partijen in de Strafrechtsketen er behoefte aan hebben, dat zij op een aantal momenten in hun bedrijfsproces snel over alle relevante informatie kunnen beschikken, die zij voor de uitoefening van hun taak nodig hebben. Bovendien moeten zij er van op aan kunnen dat die informatie betrouwbaar en volledig is, dus dat er een integer en integraal persoonsbeeld geboden wordt.

1.2.1. Bieden van een integer en integraal persoonsbeeld: Het ketendossier

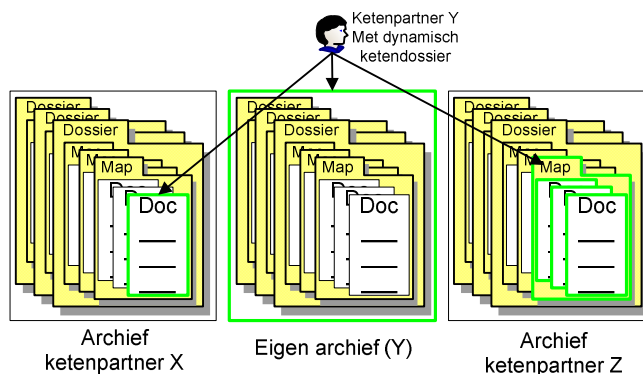
Informatie waarmee gewerkt wordt binnen de strafrechtsketen moet betrouwbaar zijn. JustID speelt hierin een centrale rol. Zo werkt zij allereerst aan het tot stand brengen van het integer persoonsbeeld. Dit betekent dat JustID al het mogelijke in het werk stelt om de juiste identiteit van de persoon vast te stellen. In de eerste plaats vergelijkt men de aangeleverde personalia met beschikbare gegevens in andere systemen, zoals de gemeentelijke basisadministratie (GBA). In de nabije toekomst wordt het mogelijk om de gegevens te vergelijken met registers waarin biometrische gegevens, zoals vingerafdrukken, iris scans of DNA, zijn opgeslagen. JustID doet er alles aan om betrouwbare informatie te verstrekken over de juiste persoon; het bieden van een integer persoonsbeeld.

Naast het leveren van betrouwbare informatie wil men dat de informatie die geleverd wordt aan de uitvoerende functionaris zo volledig mogelijk is. Deze functionaris heeft standaard de gegevens uit het locale bedrijfsproces tot zijn beschikking, maar krijgt door keteninformatiesystemen als het JDS, VIP en het CDD+ (zie paragraaf 1.1.2) en de communicatiestandaard tussen systemen in de Strafrechtsketen (Elektronische Berichtenuitwisseling - eBV) ook de beschikking over relevante informatie van ketenpartners. Via verwijzingen naar de achterliggende informatiebronnen wordt de informatie nog vollediger en wordt er een zo integraal mogelijk persoonsbeeld gevormd. Deze architectuur is gevisualiseerd in figuur 2, waarbij JustID is onderverdeeld naar een drietal doelstellingsclusters (paragraaf 1.2.2) en het CDD+ is weergegeven als zijnde in ontwikkeling. Een alternatieve weergave van de vijf basisvoorzieningen die JustID de Strafrechtsketen levert of wil leveren staat in bijlage 2.



Figuur 2 - Plaatsing CDD+ in de Strafrechtsketen en keteninformatisering theorie

JustID wil door middel van het CDD een betere bijdrage leveren aan het realiseren van een integraal persoonsbeeld van justitiabelen voor de Strafrechtsketen. Dit depot heeft namelijk tot doel om archiefstukken in de Strafrechtsketen éénmaal vast te leggen, zodat deze archiefstukken door de gehele keten heen benut kunnen worden. Samen met de coördinerende activiteiten die daarmee samenhangen, wordt het integer en integraal persoonsbeeld van justitiabelen verder vorm gegeven.



Figuur 3 - Dynamisch ketendossier

Met behulp van het CDD kan JustID een coördinerende en faciliterende rol spelen bij het bepalen van de inhoud en het realiseren van een zogenaamd digitaal “ketendossier”. Het digitale ketendossier zou die documenten bevatten die door meer dan één ketenpartner kunnen worden benut. Ketenpartners kunnen elkaar namelijk onderling toegang verschaffen op document-, dossier-, map- of archiefniveau (figuur 3). Zij blijven ook primair eigenaar van hun archiefstukken. Het digitaal beschikbaar hebben en onderling kunnen delen dient een ketenbelang: er wordt een vergaand integraal persoonsbeeld gevormd. Naast het persoonsbeeld kunnen ook andere dynamische beelden worden getoond, bijvoorbeeld vanuit het perspectief van een zaak of andere ketenpartner-overstijgende perspectieven. Voorwaarde voor een succesvol CDD+ is dat voldoende ketenpartners moeten archiveren in het CDD.

1.2.2. Organisatie

Ontstaan en ontwikkeling

De voorganger van JustID, de Centrale Justitiële Documentatie (CJD), was officieel sinds 1992 operationeel. De start van deze organisatie vloeide voort uit de wens om de diverse afdelingen van de rechtbanken te centraliseren en zaken als digitalisering en beheer van het Justitieel Documentatie Systeem (JDS) onder te brengen in één dienst. Het CJD kenmerkte zich door een grote mate van professionaliteit, ervaring en bundeling van specialistische kennis. Hierdoor werd de organisatie steeds vaker gevraagd bij de ontwikkeling en beheer van informatiesystemen. Daarnaast is een relevante kracht van de CJD dat men het overleg rond de diverse systemen coördineert en stroomlijnt.

Al die jaren is het streven geweest een onafhankelijke partner te zijn, die actief is tussen de verschillende organisatieprocessen in de strafrechtsketen, op het zogenaamde koppelvlak [Grij06]. Onder invloed van de wens een kenniscentrum voor de systemen te worden zijn allerlei organisatorische veranderingen nodig geweest. Op 1 januari 2006 is daarom de CJD officieel opgehouden te bestaan en zijn de taken ondergebracht en gebundeld, samen met de afdeling VerwijsIndex Personen (VIP) van het Centrale Justitiële Incasso Bureau (CJIB) Leeuwarden, in JustID. De primaire activiteiten van JustID bleven gehandhaafd, maar op het gebied van beheer en ontwikkeling van informatiesystemen vindt een duidelijke innovatie plaats. In 2005 werkten 55 medewerkers bij het toenmalige CJD en momenteel ruim 110 medewerkers bij JustID. JustID valt binnen justitie onder de Directeur Generaal Rechtspleging en Rechtshandhaving (DGRR).

Start Justitiële Informatiedienst

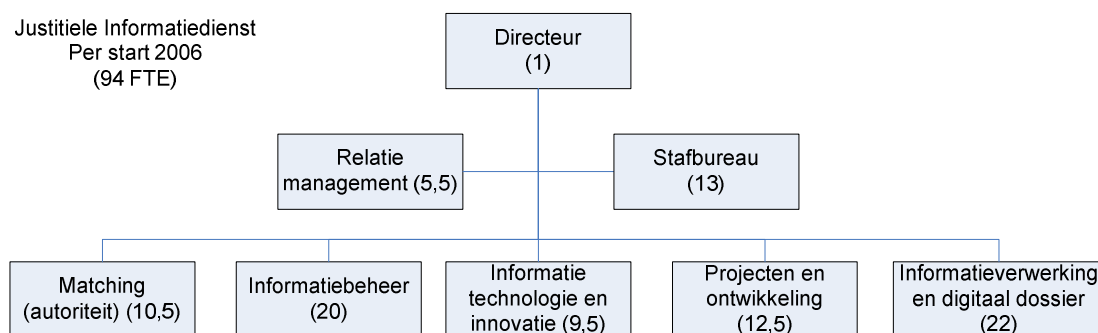
Eén van de belangrijkste voordelen van deze samenvoeging is het via één loket ontsluiten van de

informatiesystemen JDS en VIP. De dienst werkt met name aan het tot stand brengen van een integer en integraal persoonsbeeld en houdt zich vanuit deze werkzaamheden bezig met het vergelijken van beschikbare gegevens in authentieke bronnen met andere bronnen die deze gegevens bevatten, de zogenaamde bronregisters. Daarmee is de organisatie een matchingsautoriteit op het gebied van persoonsidentificatie.

JustID komt hiermee tegemoet aan de behoefte en eisen vanuit de ketenpartners voor een adequate informatievoorziening. Dit houdt in dat zij bij de uitvoering van de primaire bedrijfsprocessen vertrouwen op de aangeleverde informatie afkomstig van JustID.

Organogram

De reorganisatie van CJD naar JustID heeft tot veel organisatieveranderingen geleid. De nieuwe organisatie is weergegeven in figuur 4. Dit onderzoek wordt uitgevoerd vanuit het onderdeel Informatieverwerking en Digitaal dossier.



Figuur 4 - Organogram

Onderstaande tabel geeft de doelstellingen per afdeling weer.

Afdeling	Doelstelling
Stafbureau	Er voor te zorgen dat de processen die het primaire proces ondersteunen in kwantitatief en kwalitatief opzicht goed verlopen;
Relatiemanagement	Het initiëren, onderhouden en optimaliseren van de relatie met de klantorganisatie op strategisch, tactisch en operationeel niveau;
Informatiebeheer	Het op professionele wijze uitvoeren van de loketfunctie (frontoffice), het functioneel beheer en de kwaliteitszorg (gegevensbeheer) met betrekking tot de bij JustID ondergebrachte informatiesystemen;
Informatietechnologie en Innovatie	Het op een marktconforme, moderne en professionele wijze uitvoeren van het operationele beheer van de infrastructuur, het tactische technisch beheer van de informatie-voorziening en de informatieplanning;
Informatieverwerking en Digitaal dossier	Het op een professionele wijze uitvoeren van alle (handmatige) administratieve en op archivering gerichte werkprocessen met betrekking tot de bij JustID ondergebrachte informatiesystemen;
Matching	Het komen tot een uniciteit in de vastlegging van persoonsgegevens, zodat juistheid, volledigheid en tijdigheid van het gebruik van deze gegevens wordt gegarandeerd;
Projecten en Ontwikkeling	Het optimaal ondersteunen van de primaire processen van JustID door middel van productverbetering en of organisatieverbetering, met behulp van een vaste kern van experts en een projectmatige werkwijze;

Tabel 1 - Doelstellingen per afdeling. Bron: AO JustID

Visie en missie

JustID hanteert de volgende visie bij de uitvoering van haar taken:

JustID streeft naar een optimale informatievoorziening binnen de justitiële bedrijfsketen. Zij is dienstverlenend en gaat klantgericht te werk. Bij de kwaliteitsverbetering van de gegevensbestanden neemt JustID het initiatief en voert tevens de regie. De meerwaarde van JustID wordt in dit verband gerealiseerd door een bundeling van kennis op het gebied van technisch beheer, functioneel beheer en informatieverstrekking.

Uit deze visie vloeit de volgende missie voort:

JustID fungeert als centraal punt bij de informatievoorziening in de strafrechtssketen en ontsluit een integer en een integraal persoonsbeeld van justitiabelen aan daartoe gerechtigden. Om dit persoonsbeeld samen te stellen worden in samenwerking met de ketenpartners gegevens op een efficiënte en effectieve manier verzameld, bewerkt en beheerd. Vanuit de uitvoeringspraktijk acteert JustID pro-actief op de beleidsmatige en uitvoeringsontwikkelingen uit de strafrechtssketen en doet aan de hand daarvan voorstellen tot aanpassing en verbetering van de informatievoorziening t.b.v. de strafrechtssketen.

Doelstellingen

Om invulling te kunnen geven aan de missie en visie van de dienst worden er jaarlijks doelstellingen

geformuleerd. Deze doelen houden als vanzelfsprekend verband met de diverse werkzaamheden en systemen binnen JustID. Sinds 1 april 2004 is de Wet Justitiële Gegevens (WJG) richtinggevend binnen een groot gedeelte van de werkzaamheden. De diverse genoemde doelstellingen zijn ontleend aan het jaarplan 2006. In dit plan worden de doelstellingen geclusterd weergegeven, op een drietal gebieden:

- Justitiële Documentatie (JD)
- Duurzame Documentaire Informatie Voorziening (DDIV)
- Beheer derden

Binnen JustID is een belangrijke doelstelling het verstrekken van inlichtingen uit de Justitiële Documentatie. Dit wordt vormgegeven door het registreren van de in het bedrijfsproces benodigde gegevens. Kwaliteitseisen aan de processen verschuiven steeds meer van registratie naar verstrekking. De registratieprocessen zijn in de opbouwfase (1990-1995) van de voorloper van JustID geoptimaliseerd en in 2005 ten gevolge van de Wet JSG herzien. Kwaliteitswinst in de totale dienstverlening moet derhalve meer in het verstrekkingproces worden gezocht

Naast de doelstellingen die in relatie tot de Justitiële Documentatie zijn geformuleerd, zijn er ook doelen voor de Duurzame Documentaire Informatie Voorziening (DDIV) opgesteld. De dienst functioneert als beheerder van de centrale bewaarplaats voor de Dubbelen van de Burgerlijke Stand (DBS). Zo worden de akten van de burgerlijke stand altijd in tweevoud opgesteld. Vervanging van de Dubbelen (door vermissing of verminking) wordt op die manier gewaarborgd.

Daarnaast is de oprichting van het CDD+ een belangrijk doel. Dit systeem brengt kwalitatief hoogwaardige zaak- en persoonsgegevens samen zodat de kloof tussen papieren en digitale werkstromen zal komen te vervallen, evenals de noodzaak tot het bestaan van papieren archieven.

De laatste doelstelling van JustID is het uitvoeren en coördineren van beheertaken rondom de informatiesystemen van derden. Afspraken over de betreffende dienstverlening worden afgestemd op de wensen van individuele organisaties.

Informatiesystemen

JustID heeft per doelstellingclustering informatiesystemen toegekend die bijdragen aan de realisatie van de doelstellingen. In onderstaande tabel zijn deze weergegeven, eveneens als in figuur 2.

Doelstellingclustering	Systemen		
Justitiële Documentatie	JDS	OBJD	
Duurzame Documentaire Informatie Voorziening	DBS	PDS	(CDD+)
Beheer derden	FRIS	LURIS	

Tabel 2 - Systemen

Hoofdstuk 2 Digitale duurzaamheid en het CDD+

Voortbouwend op het vorige hoofdstuk waarin de context van het uitgevoerde onderzoek is beschreven, wordt in dit hoofdstuk enkele van de achterliggende concepten uitgelegd en wordt aangegeven hoe deze in het CDD voorkomen. Zo komt in dit hoofdstuk aan bod wat digitale informatie en duurzaamheid daarvan eigenlijk is, welke soorten digitale informatie er te onderscheiden zijn en welke kenmerken deze hebben. Daarna wordt toegelicht welk modellen ten grondslag liggen aan het ontwerp van het CDD, het systeem dat de digitale informatie opslaat. De eerste paragrafen (2.1, 2.2 en 2.3) van dit hoofdstuk vormen de aanleiding voor dit onderzoek. De overige paragrafen gaan in op de probleemstelling, onderzoeksvragen en aanpak.

2.1. Digitale duurzaamheid

Bij het bewaren van informatie met de bedoeling deze later weer te kunnen raadplegen, is het van belang dat deze ook daadwerkelijk later weer te representeren is. Bewaring moet bijvoorbeeld eeuwig zijn in het kader van ons cultureel erfgoed: toekomstige generaties moeten kunnen terugkijken op vroegere tijden, zoals wij nu ook terugkijken op kleitabletten, hiërogliefen en grottekeningen. Naast de doorsnede van informatiesoorten die culturele relevantie hebben, wordt er ook veel informatie gegenereerd in en tussen organisaties. Deze werken steeds vaker digitaal, maar het archiveren gaat in de Strafrechtsketen voornamelijk nog op papier.

Van papier zijn de duurzaamheideigenschappen bekend; onder goede condities kan hoogwaardig papier zeer lang worden bewaard. Voor een digitaal archiefobject, bestaande uit bits op een drager (harde schijf, optische media, etc.) waarbij het object door specifieke hard- en software wordt gerepresenteerd aan de gebruiker, is dat problematisch. De hard- en software zijn immers in de tijd aan veel veranderingen onderhevig, wat veroorzaakt dat een digitaal bestand dat nu wordt opgeslagen over bijvoorbeeld 50 jaar niet meer te representeren is. Daarnaast is veranderende wet- en regelgeving van invloed op de manier van bewaren. Dat digitale informatie onderhevig is aan veranderingen door de tijd heen is een wereldwijd probleem en wordt ook wel het vraagstuk van digitale duurzaamheid genoemd [Task96], het onderwerp dat centraal staat in deze scriptie.

2.2. Digitale informatie

Deze paragraaf beschrijft, grotendeels aan de hand van informatie van het testbed digitale bewaring [Test05] en het expertise centrum DAVID [Eynd01], wat digitale informatie is, welke eigenschappen deze hebben en welke soorten er te onderscheiden zijn.

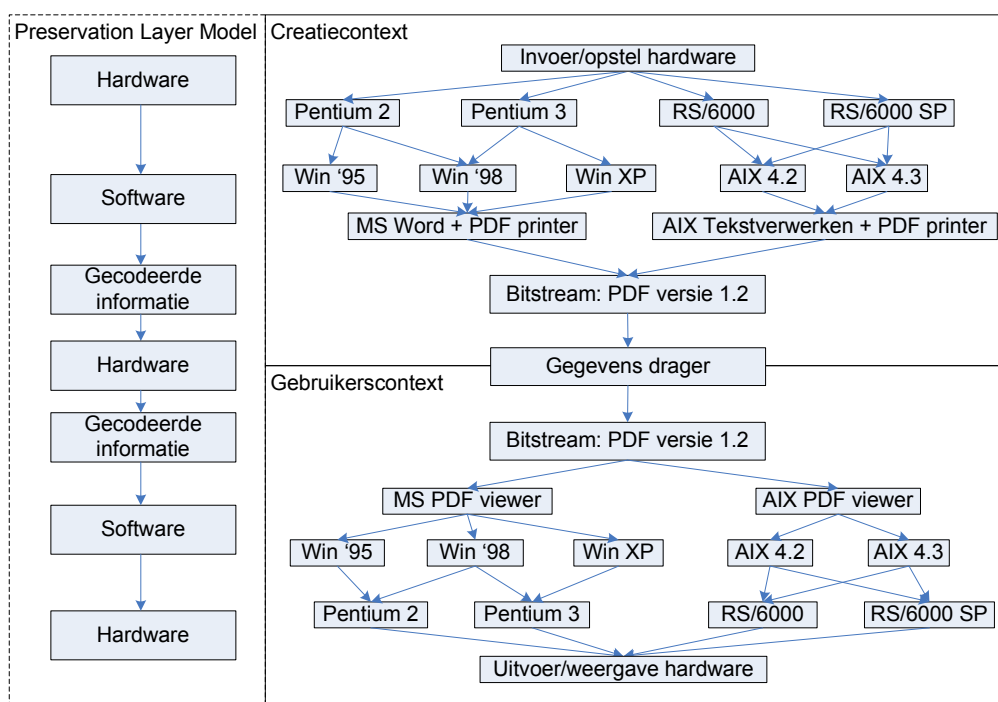
2.2.1. Definitie

Tijdens werken op papier was er weinig tot geen verwarring rondom het begrip informatie. Een document was veelal een bewijs van (informatie)transactie en was fysiek, in de vorm van bijvoorbeeld een A4, bon, brief, memo of foto. Bij digitaal werken verschuiven deze eigenschappen. Informatie wordt technologisch geproduceerd en vastgelegd en biedt daarmee ruimte aan meer dan alleen een statisch visuele weergave. Ook audio en dynamisch visuele informatie wordt digitaal vastgelegd.

Voor het opstellen en de representatie van deze informatie is technologie nodig, tenzij de informatie wordt overgezet naar een fysiek medium, zoals papier. Zo zijn er verschillende componenten te onderscheiden het opstellen en de representatie van digitale informatie aan de gebruiker mogelijk

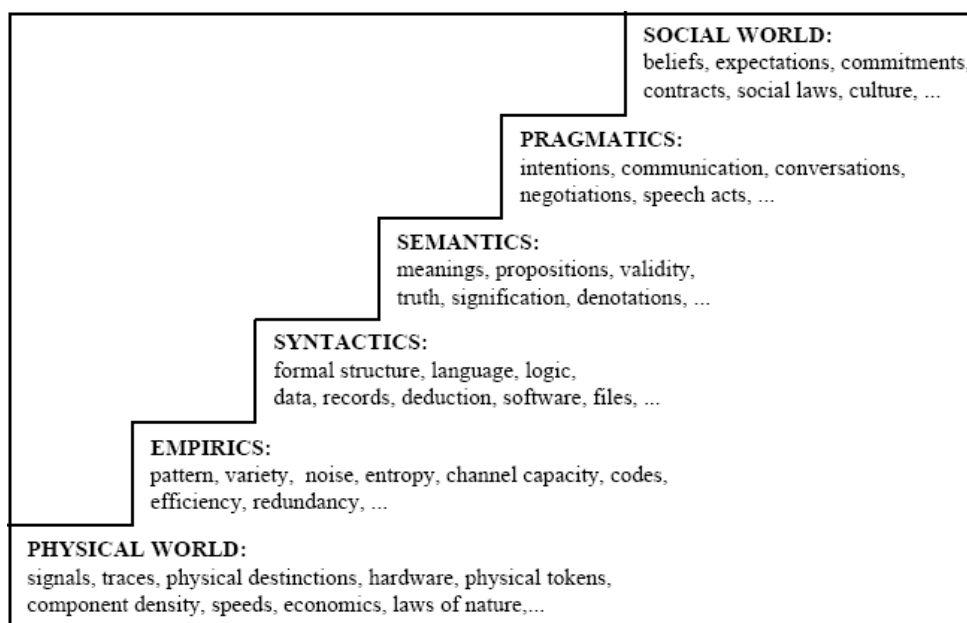
maken (linker kolom figuur 5). De originele informatie wordt digitaal opgesteld in een creatiecontext waarin zij, gecodeerd met behulp van een bepaalde combinatie van soft- en hardware, wordt opgeslagen in een digitaal computer bestand. Dit bestandsformaat (of combinatie hiervan) bestaan uit bits, de enen en nullen. Binnen de gebruikerscontext worden de bits gelezen en geïnterpreteerd door software. De interpretatieslag door de software wordt vervolgens voortgezet door de interpretatie door hardware, die informatie weergeeft op bijvoorbeeld een beeldscherm of printerpapier. Daarna is het aan de raadpleger om datgene dat wat wordt weergegeven te interpreteren.

Een digitaal bestand (bestaande uit bitstreams) kan alleen correct geïnterpreteerd worden door een beperkte combinatie van hardware en software. Een niet dusdanig geschikte combinatie kan leiden tot een weergave die anders is dan het origineel of dat er zelfs niets gerepresenteerd kan worden. Er kan dus worden gesteld dat alle digitale informatie een bepaalde set van afhankelijkheden (op elkaar voortbouwende lagen) nodig heeft om gerepresenteerd te kunnen worden. Deze set van lagen wordt in het Preservation Layer Model [Dies02] ook wel viewpaths genoemd. Een concreet voorbeeld van een viewpath is weergegeven in figuur 5. Deze toont een PDF v1.2 bitstream die gevormd wordt uit audio en/of visuele informatie vanuit de creatiecontext, die via invoer- en opstel-hardware wordt verwerkt op een computer. Op deze computer draait een besturingsstelsel en een tekstverwerker, zoals MS Word. Daarnaast biedt deze applicatie de mogelijkheid om te printen naar een PDF bestand. Dit bestand bevat de digitaal gecodeerde informatie die vervolgens in een gebruikerscontext al dan niet geraadpleegd kan worden. In het geval van het PDF bestand zal er geen audio of bewegende visuele delen als zodanig worden gecodeerd, omdat dit bestandsformaat daar niet de mogelijkheid voor biedt.



Figuur 5 - Voorbeeld viewpath voor PDF v 1.2. Deels afkomstig uit [8]

Daar waar het Preservation Layer Model [Dies02] met viewpaths zich beperkt tot het IT platform, gaat het semiotisch raamwerk model van Stamper [Stam96] verder. Dit model identificeert zes op elkaar voortbouwende lagen van waaruit gekeken kan worden naar informatiesystemen (figuur 6).

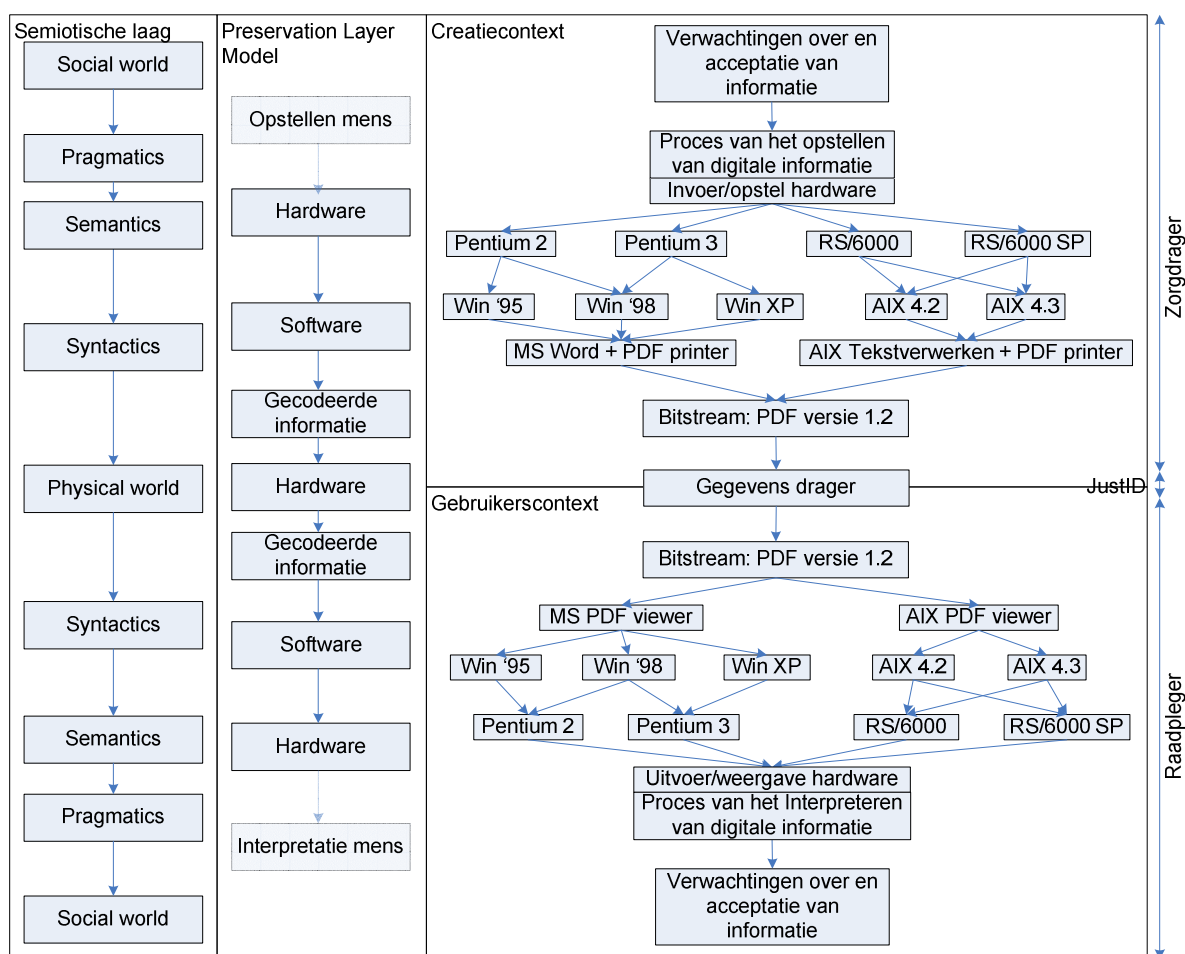


Figuur 6 - Semiotisch raamwerk [Stam96]

De onderste drie lagen (physical world, empirics en syntactics) komen overeen met die van het Preservation Layer Model. Informatie is opgeslagen in de wereld op een fysieke drager (papier, hardware). Het is daar naartoe geschreven (empirisch), zoals met inkt op papier of als lading op een geheugen gezet (opslag van de bitstream op een harde schijf). Er kan logisch gekeken worden naar de opgeslagen informatie en er kunnen structuren en taal worden geïdentificeerd, zoals goed vervoegde Nederlandse zinnen op papier, in een applicatie of vanuit spraak, structuren die een afbeelding logisch vormgeven.

De overige drie lagen die in het semiotisch raamwerk genoemd worden (semantics, pragmatics en social world), worden niet binnen het Preservation Layer Model beschreven. Deze lagen gaan in op de betekenis die mensen toewijzen aan datgene dat op syntactisch niveau gepresenteerd is (semantiek), de achterliggende bedoeling die met de informatie gemoeid is (pragmatiek) en de houding die men naar de informatie heeft (sociaal).

Wanneer de viewpath gedachtegang in figuur 5 wordt uitgebreid met de door Stamper besproken semiotische lagen, wordt onderstaande figuur gevormd. Hierin beschrijven de pijlen in de creatiecontext het empirische, ook wel het transport van de codering van informatie.



Figuur 7 - Uitbreiding viewpaths vanuit ontbrekende semiotische lagen

De drie lagen buiten het Preservation Layer Model (semantics, pragmatics en social world) zijn belangrijk voor het over langere tijd toegankelijk bewaren van informatie. Deze lagen gaan namelijk in op de context waarin de informatie ooit is opgesteld, zoals waarden en normen en de (bedrijfs)processen. Zij geven betekenis aan de informatie.

2.2.2. Informatie eigenschappen

Papieren archiefstukken vormen fysiek een eenheid. De combinatie van context, inhoud, structuur en uiterlijk maken een papieren document authentiek; de aanpassing van een van de eigenschappen heeft invloed op de andere eigenschappen. Bij digitale informatie is dit anders. Zij heeft nog wel de eigenschappen context, inhoud, structuur en uiterlijk, maar biedt ook nog de ruimte aan 'gedrag' [Roth99]. De combinatie van eigenschappen kan authentiek worden gesteld, maar digitale informatie is niet fysiek of onontkoombaar verbonden qua eigenschappen: alles is afhankelijk van de interpretatie die geschiedt per laag in het eerder beschreven model. De eigenschappen waarin het ene digitale archiefstuk van het andere kan verschillen, worden hieronder kort toegelicht.

Context

Context gaat in op de oorspronkelijke omgeving, de creatiecontext, waarin de digitale informatie gecreëerd is. Het gaat hierbij niet over de technische omgeving noodzakelijk voor de representatie.

Voorbeelden van creatiecontext is informatie over de functie waardoor de digitale informatie gecreëerd is en binnen welk bedrijfsproces dit gedaan is. Maar ook referenties naar andere informatie, zoals documenten, zaken en personen vallen onder de context eigenschap van informatie [Hofm00].

Inhoud

Digitale informatie kan de meest uiteenlopende informatie bevatten, zowel tekstuele als audio en/of visuele informatie.

Structuur

De structuur van digitale informatie wordt gevormd door een logische hiërarchie van, en de relaties tussen, de delen van de digitale informatie. Zo kan bijvoorbeeld de inhoud van een website bestaan uit tekst, foto's, links en video, die in een HTML structuur staan die de hiërarchische structuur beschrijft. Een database bestand kan daarentegen bestaan uit verschillende tabellen, views en functies die in relatie tot elkaar de structuur zijn van de database. Zonder deze structuur zal database representatiesoftware de informatie niet op eenzelfde manier kunnen tonen aan de gebruiker.

Uiterlijk

De eigenschap uiterlijk gaat in op de weergave die gedaan wordt van de inhoud. Bij bijvoorbeeld een website wordt uiterlijk toegevoegd door gebruik te maken van CSS. Het gaat bij uiterlijk dan ook om bijvoorbeeld aspecten als lettertype, lettergrootte en kleur. In het geval van een database bestand gaat het om hoe de velden van de database worden weergegeven in de gebruikersapplicatie (Graphical User Interface, GUI).

Gedrag

Het gedrag dat digitale informatie kan bevatten is het moeilijkste te bewaren. Het gaat in op de interactieve kenmerken van de informatie. Zo weet bijvoorbeeld alleen de spreadsheetapplicatie dat de formule die in een spreadsheet staat moet worden uitgevoerd en dat de uitkomst moet worden weergegeven. In het geval van een website met een Java-applet, moet de webbrowser beslissen om de Java-applet te laten interpreteren door Java software. Deze software zorgt dan vanuit een bepaalde aanvangtoestand voor een interactie met de gebruiker.

2.2.3. Informatietypen

Informatietypen gaan in op verschillende manieren van het coderen van digitale informatie in de bitstream. De eigenschappen van digitale informatie verschillen per digitaal informatietype [Test03]. Zo zijn tekstuele bestanden of documenten anders dan databases, spreadsheets of video en zijn er veel verschillende bestandsformaten die inhoudelijk dezelfde informatie kunnen bevatten. Welke combinatie van digitale informatietypen en bestandsformaten gebruikt wordt is afhankelijk van de eisen die gebruikers/werkprocessen aan de informatie stellen. Bij de keuze voor informatietypen en daar onder te schalen bestandsformaten, hechten zij een bepaald belang aan de informatie eigenschappen waarin zij voorzien. Zo mag men ervan uitgaan dat uiterlijk minder belangrijk is voor e-mailberichten, aangezien e-mail berichten op verschillende PC's met verschillende e-mail programma's en persoonlijke instellingen verschillend worden weergegeven. Voor tekstdocumenten kan uiterlijk weer wel van wezenlijk belang zijn [test03].

Er zijn verschillende digitale informatietypen te onderscheiden (zie bijlage 3), met nog veel meer manieren van informatie coderen in digitale bestanden, wat zich uit in bestandsformaten (zie tabel 3).

Er wordt hierbij onderscheid gemaakt naar enkelvoudige en samengestelde informatietype. Deze laatste gebruikt binnen digitale bestanden meerdere enkelvoudige informatietypen en werkt met referenties om onderlinge relaties te leggen.

	Informatietype	bestandsformaten
Enkelvoudig	Tekstueel	ASCII, Unicode, txt
	Audio	Wav, mp3, m4a, wma, ogg, acc
	Visueel (stil)	Tiff, jpeg, png, gif, eps, wmf, svg
	Visueel (bewegend)	Avi, mpeg, mov, gif, asf, swf, ogm
Samengesteld	Inpak	Zip, rar, gz, ace
	Database	db2, sql, oracle, mdb
	Document	doc, docx, wpd, odf, abw
	Script, bron-, objectcode, executables & codebibliotheek	dll, exe, com, src, o, c, java, bat, class, jar, php, cgi, asp
	Pagina beschrijvend	pdf, xps, ps, css, xslt, xsl, dvi
	Presentatie	pps, ppt, pdf, sxi, otp
	Spreadsheet	123, xls, ods, wks, xlsx
	Lettertype	otf, fon, afm, pfm, ttf, tfm
	Webpagina	html, asp, cgi, php, mhtml
	Expertiseformaat	Cad, step, stl, ccc, spss

Tabel 3 - Informatietypen met bestandsformaten

De overheid stelt eisen aan de opslag van archiefstukken, zowel wanneer het daar papier betreft als digitaal. De regeling “geordende en toegankelijke staat archiefbescheiden” [Ploe02] gaat in op het opslaan van informatie in digitale bestanden. Deze regeling zorgt voor een versimpeling van wat er dient te worden bewaard. Dit doet zij door aan te geven wat (context, inhoud, structuur en vorm) op welke manier (informatietypen/bestandsformaten) dient te worden opgeslagen. In deze regeling wordt echter niets genoemd over de informatie eigenschap ‘gedrag’. Het is dan ook de vraag in hoeverre wet- en regelgeving de veranderingen in de technologie kan bijbenen en kan blijven aangeven op welke manier informatie digitaal moet worden opgeslagen.

2.3. Records Continuüm en Archiefbeheer in het CDD+

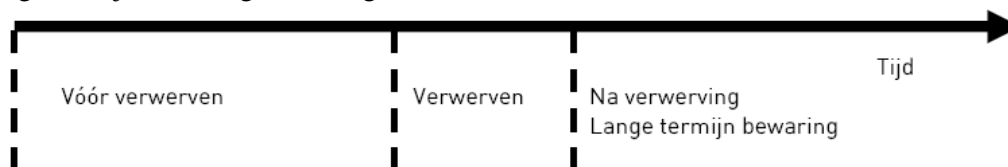
In voorgaande paragrafen is uitgewerkt wat digitale informatie is en waar deze uit kan bestaan. Deze classificatie is echter niet voldoende voor het goed en duurzaam beheren van de informatie. Hiervoor zijn in relatie tot het doel van het CDD, zijnde duurzame digitale archivering, verschillende archiefbeheermodellen beschikbaar. Het traditionele model voor archiefbeheer ('records management') is het levensloop model ('life cycle') [Test03-2]. Dit model kent enkele duidelijke breekpunten binnen de levensloop van archiefstukken. In Nederland zijn dat de dynamische, semi-statische en statische fase. In dit model betekent de overgang van de ene fase naar de andere een overgang naar een andere

beheervorm. Bij de overgang tussen semi-statisch en statisch is dat het meest duidelijk, waarbij - voor de overheid - sprake is van een wettelijk verplichte, formele overdracht van de archiefvormende organisatie naar een openbare archiefbewaarplaats, zoals beschreven in de Archiefwet 1995. Deze wet ziet er, samen met daaruit voortvloeiende regelgeving/archiefwettelijke kaders, op toe, dat gedigitaliseerde archiefstukken dezelfde bewijskracht, bewaartermijn, authenticiteit, betrouwbaarheid en representeerbaarheid moeten hebben als het originele papieren exemplaar. Digitale archieven moeten dus duurzaam worden bewaard.

JustID wil archiefstukken van Strafrechtsketenpartners digitaal archiveren met behulp van het CDD. Met de realisatie van dit keteninformatiesysteem [Grij06] streeft zij naar een drietal zaken: (1) het bieden van een integraal persoonsbeeld door middel van een dynamisch digitaal ketendossier, (2) substitutie (vernietiging) van het papieren origineel door het digitale exemplaar als origineel ter verantwoording te stellen en (3) het verkorten van de doorlooptijd bij het lichten van archiefstukken. De bouw van het CDD is vergevorderd en wordt als keteninformatiesysteem, samen met aanvullende dienstverlening en beheer vanuit JustID (CDD-plus), gezien als een kansrijk keteninformatiesysteem project [Grij06]. De kans bestaat dus dat het CDD+ een basisvoorziening voor de Strafrechtsketen zal worden.

Ten grondslag aan het CDD+ ligt een relatief nieuwe benadering van archiefbeheer, namelijk het 'records continuüm model' [Upwa94][Jonk05]. Deze visie steunt op een integrale benadering, waarbij er al vanaf het ontstaan van de informatie nadrukkelijk rekening wordt gehouden met alle belangen. Het model doorbreekt de indeling van de traditionele fasen en bouwt consequent aan een zichtbare, ononderbroken lijn in beheer, dat zich uit in fasen, namelijk (figuur 8):

- Vóór overbrenging naar een archiefbewaarplaats;
- Het verwerven/de overbrenging;
- Lange termijn bewaring in een digitaal archief.



Figuur 8 - Drie fasen van het records continuüm [Test05]

Het onderzoek in deze scriptie richt zich met name op de derde fase: lange termijn bewaring in een digitaal archief. De voorgaande fasen zijn echter van groot belang en buiten dit onderzoek al goed uitgewerkt in het kader van de ontwikkelingen rondom het CDD. Het belang van deze fasen wordt onderstreept door het Open Archive Information System referentie model [Cons02] (paragraaf 2.3.3), waaruit blijkt dat om effectief aan de slag te kunnen tijdens de laatste fase van het records continuüm model, er al veel gedaan moet zijn in de eerdere fasen. Bij het ontstaan van een breuk in een van de drie fasen, komt de duurzaamheid in het gedrang. Alle fasen worden dan ook hieronder besproken en toegespitst naar de situatie van het CDD+.

2.3.1. Voor verwerving: Archiveringssysteem voor informatieobjecten

De eerste fase in de records continuüm model is die van *vóór verwerving*. In deze fase dient een veelvoud van gegevens over de ooit te archiveren informatie te worden opgeslagen, namelijk de zogenaamde metadata (gegevens over gegevens). Deze metadata moet worden opgeslagen in een

archiveringssysteem, zoals het CDD, wat in wezen een bijzonder soort documentair informatiesysteem is. Het heeft als taak informatie die ontstaan is binnen bedrijfsprocessen op te slaan, te beheren en weer te kunnen aanbieden in een te representeren vorm.

Metadata

Metadata zijn altijd al aanwezig geweest in het archiefbeheer. Papieren documenten dragen voor een groot deel zelf de metadata, zoals de naam van de auteur, afzender, datum en afzender. Later kan het nog een handtekening of stempel krijgen. Verder kan een functionaris een blad toevoegen met aanvullende metadata, zoals een registratienummer, classificatiecode en naam van de behandelaar. Daarnaast liggen de documenten bijvoorbeeld op een plank, in een doos, in een archief, wat op zijn beurt al een onderscheid kan maken naar het type archiefstuk. Bij de omgang met digitale informatie is het proces grotendeels hetzelfde, maar geschiedt dan digitaal via procedures en applicaties.

Metadata beschrijven de eigenschappen en kenmerken van de digitale informatie. Ze zijn een hulpmiddel om informatie later weer terug te vinden, ze kunnen nodig zijn voor een juiste interpretatie en zijn noodzakelijk voor adequaat beheer en onderhoud. Metadata kunnen bijvoorbeeld gaan over een proces-verbaal dat in het kader van een zaak over een justitiabele, door een politiefunctionaris wordt opgesteld op een bepaalde datum, opgeslagen in een .doc bestandsformaat, gecreëerd door MS Word 2000 onder Windows XP te gebruiken, dat draait op 32-bits x86-instructieset hardware. Met metadata kunnen dus veel verschillende eigenschappen en kenmerken van de digitale informatie worden beschreven. Metadata zijn onder te verdelen naar enkele soorten [Hors05]:

- Beschrijvende metadata

De metadata die de digitale informatie betekenis geeft en identificeert, zoals gegevens over wie de informatie gecreëerd heeft, over wie of wat de informatie gaat, de organisatie, de functie, het mandaat en het bedrijfsproces. Het gaat om het intellectuele beheer van de informatie.

- Administratieve of beheersmetadata

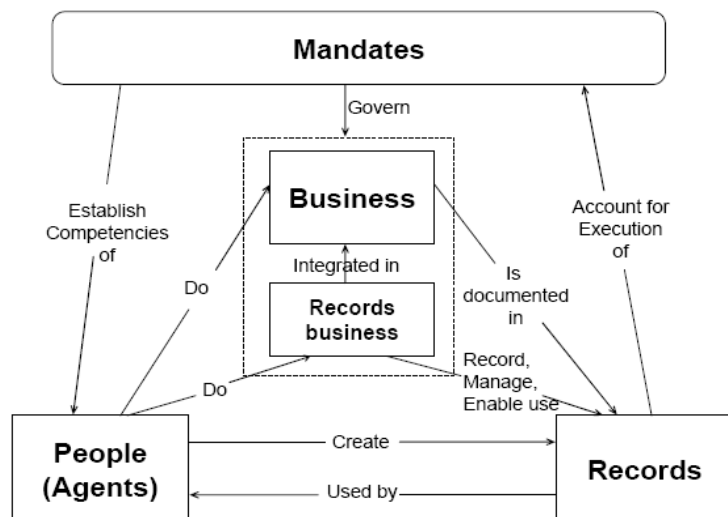
Deze metadata gaat over informatie die iets zegt over autorisaties, logistieke gegevens, eigendom, formele herkomst, raadpleeg frequentie, wijze van ordenen en bijvoorbeeld of het ter verantwoording is van beheersactiviteiten (overheid). Deze groep bestaat uit gegevens ten behoeve van het archiefsysteem en wordt gebruikt door de archiefbeheerders.

- Technische metadata

Metadata over herkomst en huidig opgeslagen bestandsformaat, gebruikte soft-en hardware. Hierbij gaat het dus ook om bijvoorbeeld de uitgevoerde handelingen op de informatie in het kader van een bewaarstrategie. De metadata dienen functionarissen die belast zijn met het “materieel” beheer.

Het verkrijgen en opslaan van metadata

De te verkrijgen metadata dienen te worden opgeslagen in het archiveringssysteem. Deze maakt gebruik van een metadatamodel, een model dat dezelfde functie heeft als dat een datamodel heeft in gangbare ontwerpmethoden van informatiesystemen [Hors05]. Er zijn op dit gebied veel standaarden die beschrijven welke entiteiten er met welke metadata beschreven kunnen worden. Bijvoorbeeld ISO 23081-1:2006 (principes) en ISO 23081-2:2007 (technisch document). Deze beschrijven de overkoepelende entiteiten en onderlinge relaties (zie figuur 9) waaruit typen metadata zijn opgesteld.



Figuur 9 - Overkoepelende entiteiten en onderlinge relaties - ISO 23081-2

document, de gegevens over de verschijningsvormen (technisch), de rechten tot het document, de uitgevoerde beheershandelingen, etc. Zie voor een overzicht van de entiteiten van het metadatamodel bijlage 4. Met de opslag van de contextuele gegevens in het metadatamodel, ligt de focus van dit onderzoek op het behoud van de overige eigenschappen van digitale informatie: inhoud, structuur, uiterlijk en gedrag.

De metadata begeleiden de digitale informatie gedurende de gehele levenscyclus. Op basis van de aanwezige metadata kunnen de juiste bewaarmaatregelen worden genomen en kan worden getoetst of de digitale informatie na bijvoorbeeld een conversie van het bestandsformaat nog wel dezelfde informatie is. Metadata zijn hierdoor dus van cruciaal belang.

Niet altijd worden alle metadata direct bij ontstaan toegevoegd aan of in relatie gebracht met de digitale informatie. Technische metadata kunnen bijvoorbeeld ook tijdens de verwervingsfase worden toegevoegd door de ontvangende archiefinstantie. De archiefinstantie zal echter vaak eisen stellen aan de digitale informatie die zij zal archiveren. Zo kan er besloten worden om slechts een klein aantal digitale bestandsformaten per informatietype toe te staan en deze zelf vóór of na overdracht te converteren naar een duurzamer bestandsformaat.

2.3.2. Verwerven: Overdracht

De verwervingsfase omvat de daadwerkelijke overdracht van de digitale informatie en alle beschrijvende metadata (samen het informatieobject) vanuit de creërende organisatie naar de archiverende organisatie (JustID). De overdracht naar het CDD kan op verschillende manieren

plaatsvinden. Zo kunnen fysieke documenten middels de combinatie van een geautoriseerde scanstraat en barcodevellen worden gedigitaliseerd naar het PDF/A-1b bestandsformaat en digitaal in het CDD worden geplaatst. Dit kan desgewenst intern bij JustID of op locatie bij de ketenpartner. Wanneer het op locatie bij de ketenpartner gescand wordt of ‘digital born’ informatie is (digitaal ontstaan), kan via een op XML gebaseerd berichtenprotocol (het EBV-protocol) de informatie worden overdragen naar het CDD. De originele digitale vorm moet dan wel aan enkele eisen voldoen, wil JustID digitale duurzaamheid kunnen garanderen. Daarvoor wordt bij JustID bijvoorbeeld het digitale bestand geconverteerd naar PDF/A-1b, wanneer dit nog niet het geval was. Voor de overdracht van digitale archiefstukken tussen archiveringssystemen heeft het CDD+ een in- en exportfunctie.

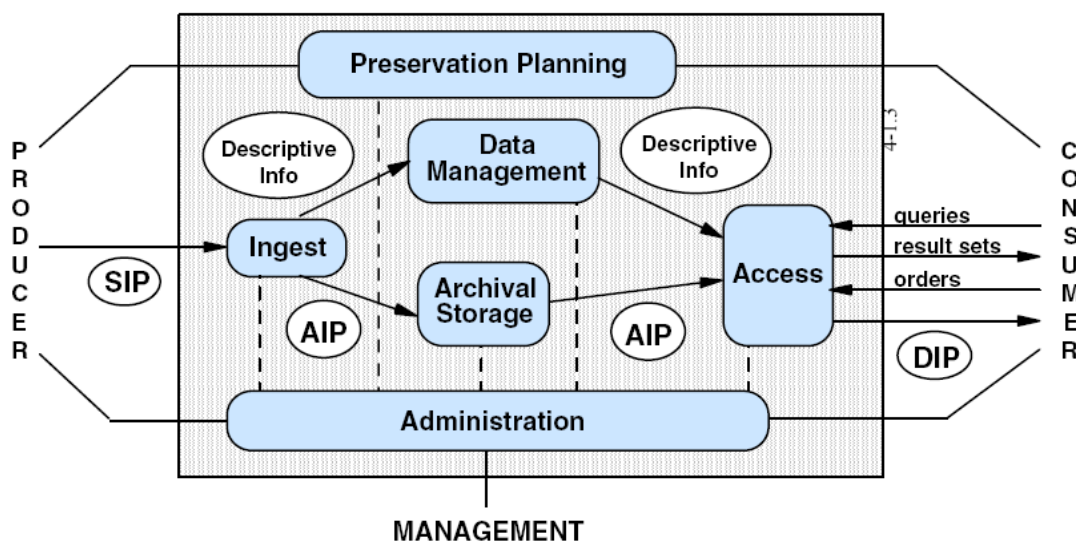
2.3.3. Verworven: Open Archival Information System

De derde en afsluitende fase van het records continuüm model is de “lange termijn bewaring in een digitaal archief”. Eigenlijk is dit niet een afsluitende fase omdat de activiteiten rondom digitaal archiveren gewoon door blijven gaan. Functies en eigenschappen waarover een archiefsysteem moet beschikken om aan lange termijn bewaring te kunnen doen, worden beschreven in het referentiemodel voor het “Open Archival Information System” (OAIS-model). Dit model wordt over het algemeen beschouwd als de standaard bij het definiëren van kaders en procedures ten behoeve van het bewaren van informatieobjecten. Het is in eerste instantie ontwikkeld door NASA’s Consultative Committee for Space Data Systems (CCSDS) in 2001 [Cons02] en is in 2003 tot ISO verwerkt: ISO 14721:2003.

Onderdelen OAIS

Het OAIS kan als gids worden beschouwd bij de ontwikkeling van een archiveringssysteem, zoals het CDD. Het zorgt er voor dat wanneer men archieven in toegankelijke staat langdurig wil bewaren, er rekening gehouden wordt met de ingrijpendheid van deze veranderingen in deelprocessen. Het OAIS is niet een model waar vanuit direct kan worden geïmplementeerd: het is een conceptueel referentiemodel. Het beschrijft de primaire functiegroepen (functies, activiteiten en workflow) die terug te vinden moeten zijn in elk archiveringssysteem en gestalte geven aan de archieffunctie van een archiefdienst of -instelling. De functiegroep entiteiten zijn weergegeven in figuur 10 en als onderstaand te benoemen [Cons02][Boud04]. De zogenaamde informatiepakketten, aangegeven met afkortingen in figuur 10, worden verder behandeld in de volgende paragraaf.

- Inname / *ingest*: kwaliteitscontrole, registratie, beschrijving, extractie metadata, transformatie archief informatie, etc.;
- Archiefopslag (fysieke opslag en beheer) / *archival storage*: voorzien in goede materiële omstandigheden, vervangen van informatiedragers, foutopsporing (checksums), rampenplan, backups maken, leesbaar houden, etc.;
- Toegankelijk maken (intellectueel beheer) / *data management*: beheren en bijwerken van beschrijvingen en metadata, voorzien in zoekmogelijkheden;
- Beheer / *administration*: formuleren van beleid, overleggen met archiefvormers, standaarden vastleggen, beheer digitaal depot, bijhouden documentatie, opvolgen technologische veranderingen, etc.;
- Ter beschikking stelling / *access*: archieftoegangen en archiefdocumenten beschikbaar stellen;
- Planning m.b.t. digitale bewaring / *preservation planning*: services en functies rondom het monitoren van de omgeving van het archiveringssysteem en het zekerstellen dat de informatie over de lange termijn toegankelijk blijft voor de gebruikers.

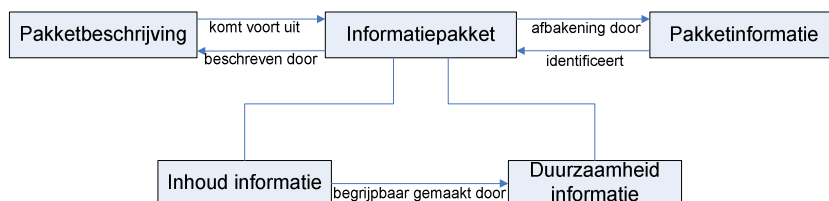


Figuur 10 - OAIS onderdelen [Cons02]

Informatiepakketten

Bij de OAIS beschrijving van functies, processen en informatiestromen staan informatiepakketten, die in het digitale archief worden opgenomen, beheerd en geraadpleegd, centraal. Informatiepakketten zijn conceptuele containers met twee soorten informatie (zie figuur 11). Enerzijds bevat het de inhoud informatie (met bijbehorende voorstellingsinformatie) en anderzijds duurzaamheidsinformatie (ter interpretatie). Elk van deze soorten informatie bestaat uit data (bits) en daarbij behorende voorstellingsinformatie.

De verschillende informatiepakketten die een rol spelen in het OAIS zijn de Submission Information Package (SIP), Archival Information Package (AIP) en Dissemination Information Package (DIP) (figuur 10). Deze specifieke informatiepakketten



Figuur 11 - Schematische weergave Informatiepakket [Cons02]

verwijzen naar de wijze waarop de informatie is georganiseerd tijdens de inname bij het archief (SIP), de opslag van het archiefstuk in het digitaal depot (AIP) en de verspreiding naar de eindgebruikers (DIP) [Test05]. De SIP en DIP pakketten zijn noodzakelijk tijdens communicatie tussen het archiveringssysteem en de producten/gebruiker van het archiefstuk.

Er zijn daarnaast pakketbeschrijvingen om informatie snel te kunnen terugvinden. Deze bevat alle informatie om een ketenpartner de benodigde gearcheveerde informatie te laten vinden, te analyseren en op te vragen. De bron van de pakketbeschrijvingsinformatie is de inhoud informatie en de duurzaamheidsinformatie. De archivaris leidt per type archiefstuk hieruit de pakketbeschrijvingsinformatie af, die toe te passen is in bijvoorbeeld zoekfuncties binnen het CDD. Dit zijn bijvoorbeeld velden als identificatienummer, zaaknummer, datum, archiefvormer en titel van het archiefstuk.

Naast de pakketbeschrijvingsinformatie is er zogenaamde pakketinformatie. Deze is nodig om de verschillende onderdelen van een informatiepakket te kunnen afbakenen op de informatiedrager en deze delen als samenhangend aan te kunnen duiden. Pakketinformatie kan bijvoorbeeld bestaan uit de

beschrijving van gemaakte afspraken over het toekennen van bestandsnamen of het beschrijven van de structuur van directories. Bij migratie van het informatiepakket moet meestal de pakketinformatie ververst worden.

De exacte inhoud van de informatiepakketten is afhankelijk van waar de informatie zich in het archiveringsproces bevindt. Dit kan bijvoorbeeld in het eigen archiefsysteem zijn of dat van een ketenpartner. Daarnaast kan het informatiepakket worden overgedragen naar een ketenpartner, een nieuw eigen archiefsysteem of naar een andere archiefinstelling. Alle moeten echter informatiepakketten hebben met de structuur en relaties zoals weergegeven in figuur 11. In bijlage 5 staat een gedetailleerdere weergave van informatiepakketten.

Informatiepakketten en het CDD

In termen van het CDD vormt alle informatie afkomstig van een ketenpartner (metadata, digitale bestanden, etc.) SIP pakketten (creatiecontext). Deze pakketten vormen samen de informatie die de archiverende instantie (CDD) identificeert een AIP pakket en worden opgeslagen in het archiefsysteem. Dergelijke informatie beschrijft bijvoorbeeld waar het archiefstuk vandaan komt, op welk tijdstip het verstuurd is en met welke software de informatie gecodeerd is. Deze AIP pakketten zijn ter archivering en worden zonodig geüpdate aan de hand van handelingen (bewaar- en beheeractiviteiten) die ingrijpen op het AIP, zoals systeem migratie, bestandformaat conversie en het toevoegen van andere metadata.

Het AIP pakket dient als basis voor de vorming van een DIP pakket, het pakket dat kan worden verstrekt aan archiefraadplegers. Het DIP wordt gevormd aan de hand van informatie die de archiverende instelling heeft over de omgeving van de eindgebruiker (gebruikscontext). Zo kunnen (delen van) AIPs worden geconverteerd naar andere bestandsformaten ter representatie bij de eindgebruiker (ketenpartner). Een informatiepakket beschrijft dus niet noodzakelijk één archiefdocument; een AIP kan overeenkomen met één archiefdocument, maar kan ook overeenkomen met één component van een archiefdocument of met meerdere archiefdocumenten. Belangrijk om op te merken is dat in het CDD één AIP niet individuele componenten van een enkele representatievorm kan opslaan. Hiervoor mist een object in het metadatamodel dat de onderlinge relaties beheerd tussen AIPs. Een AIP kan echter wel meerdere manifestaties [Thib02] van bestandsformaten bevatten, dus of dit niet 'atomair' kunnen opslaan in het CDD direct een probleem vormt is nog maar de vraag. In termen van het OAIS beschrijft Boudrez enkele voorbeelden van de theoretisch mogelijke inhoud die een AIP samen met zijn relevante metadata kan bevatten [Boud05]:

- Één element van een archiefdocument: bv. een afbeelding ingevoerd in een tekstdocument;
- Alle componenten van één representatie van een archiefdocument: bv. een tekstdocument (inclusief de afbeeldingen en tekst);
- Alle representaties van een archiefdocument: bv. een tekstdocument in zijn originele vorm en zijn archiveringsformaat;
- Alle onderdelen van een archiefdeel: bv. alle archiefdocumenten die deel uit maken van een dossier- of onderwerpsmap.

2.4. Probleemstelling en onderzoeksvragen

De in dit hoofdstuk beschreven context leidt tot de aanleiding van dit onderzoek. Deze paragraaf beschrijft de probleemstelling en is onderverdeeld in wat het probleem is, welke doelstelling daaruit

wordt afgeleid en welke vraagstelling gehanteerd wordt om de doelstelling te bereiken.

JustID wil archiefstukken van Strafrechtsketenpartners digitaal archiveren in het CDD. Archieven zijn echter onderhevig aan de archiefwet 1995 en de daaruit vloeiende regelgeving: archiefwettelijke kaders. Deze zijn ook aan de orde bij digitaal archiveren en dit heeft tot gevolg dat gedigitaliseerde archiefstukken dezelfde bewijskracht, bewaartermijn, authenticiteit, betrouwbaarheid en representeerbaarheid moeten hebben als het originele papieren exemplaar. Digitale archieven moeten dus duurzaam bewaard worden.

Van papier zijn de duurzaamheidseigenschappen bekend; onder goede condities kan hoogwaardig papier zeer lang bewaard worden. Voor een digitaal archiefobject, bestaande uit bits op een drager (harde schijf, optische media, etc.) waarbij het object door specifieke hard- en software wordt gerepresenteerd aan de gebruiker, is dat problematisch. De hard- en software zijn immers in de tijd aan veel verandering onderhevig wat veroorzaakt dat een digitaal bestand dat nu wordt opgeslagen over bijvoorbeeld 50 jaar niet meer te representeren is. Daarnaast is veranderende wet- en regelgeving van invloed op de manier van bewaren. Dat digitale informatie onderhevig is aan veranderingen door de tijd heen is een wereldwijd probleem en wordt ook wel het vraagstuk van digitale duurzaamheid genoemd [Task96].

Omdat er binnen JustID geen organisatiedeel is dat zich bezig houdt met de constant veranderende technologie en wet- en regelgeving, voorziet het CDD+ momenteel nog niet in de duurzame opslag van digitale archiefstukken. Tijdens het vooronderzoek naar archieven in de Strafrechtsketen en de kansrijkheid van het CDD+ [Delo06] is naar voren gekomen dat het garanderen van digitale duurzaamheid van archieven niet te automatiseren is. Een van de conclusies van het vooronderzoek is dat er een organisatie moet worden ingericht die, naast eventuele andere taken, de veranderingen in de wet- en regelgeving evenals de technologische ontwikkelingen op de voet volgt en hierop reageert: de vorming van het Expertise Centrum CDD+. Vanuit dit expertise centrum moet JustID de participerende ketenpartners de daadwerkelijke duurzaamheid van hun digitale archiefstukken in het CDD+ garanderen.

Vanuit de aanleiding en de probleemomschrijving wordt de volgende doelstelling geformuleerd:

Dit onderzoek heeft tot doel: in het kader van digitale duurzaamheid van archieven in het CDD+ het Expertise Centrum CDD+ als organisatorisch onderdeel van JustID en binnen de Strafrechtsketen in te richten en te voorzien van ondersteunende informatiesystemen, zodat zij voorbereid is en blijft op veranderingen in technologie, wet- en regelgeving.

De centrale onderzoeksvraag, die afgeleid is uit de doelstelling, is als volgt:

Hoe moet het Expertise Centrum CDD+ van JustID, dat zich binnen de Strafrechtsketen richt op digitale duurzaamheid van archieven in het CDD, organisatorisch en met ondersteunende informatiesystemen worden ingericht?

Om de centrale onderzoeksvraag te kunnen beantwoorden, dient een antwoord te worden gevonden op onderstaande onderzoeksvragen. Per hoofdstuk wordt er een onderzoeksvraag beantwoord.

1. *Welke bewaarstrategieën beschrijven hoe om te gaan met de duurzaamheid van digitale informatie en hoe worden deze in de praktijk gebracht?*
2. *Welke criteria liggen ten grondslag aan de duurzaamheid van digitale informatie?*
3. *Hoe voorziet een bewaarstrategie selectiemethode in de selectie van de meest geschikte*

bewaarstrategie?

4. *Welke organisatie-inrichting is geschikt om deze bewaarstrategie te kunnen volgen?*
5. *Met welke architectuur is deze organisatie te ondersteunen?*

2.5. Afbakening

Alvorens te beschrijven op welke manier de onderzoeksvragen beantwoord worden moet de scope van het onderzoek worden afgebakend. Op onderstaande gebieden is afbakening nodig.

Bewaarstrategie CDD+

De ontwikkeling van het CDD vanuit richtlijnen als NEN-ISO 15489-1/2 en ISO 14721:2003 (Open Archival Information System / OAIS) leveren technische afbakening voor de te volgen bewaarstrategie. De keuze voor deze richtlijnen volgt uit overwegingen die al zijn gemaakt ter attentie van het ontwerp en de bouw van het CDD. Op het gebied van bewaarstrategieën ligt migratie van en naar open standaarden voor de hand, conversie naar latere versies van bestanden zal noodzakelijk zijn, maar emulatie is ook mogelijk. JustID wil echter geen hard copy's maken (printen op papier of microfilm) of zelf een computermuseum huisvesten. Combinaties of allianties van bewaarmogelijkheden zijn mogelijk.

Metadatamodel

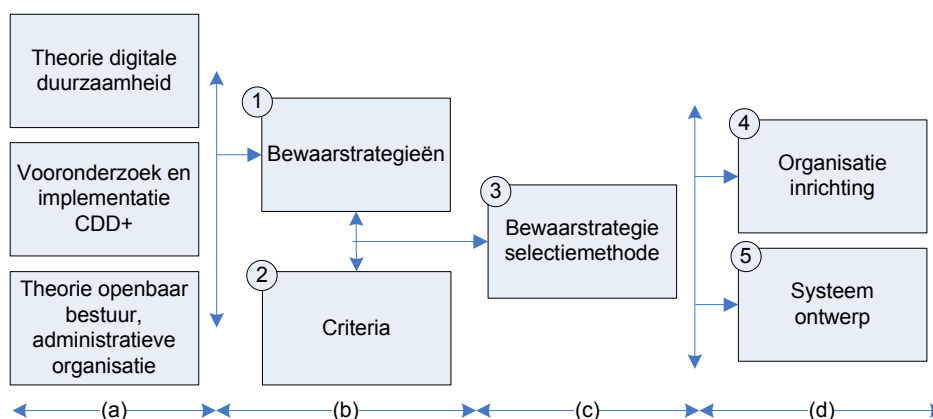
Het metadatamodel van het CDD, besproken in paragraaf 2.3.1, is ontstaan na uitvoerig onderzoek door Peter Horsman van de Stichting Archiefschool en heeft plaatsgevonden onder tal van andere metadatasets [Hors05]. ISO 23081-1:2006 (principes) en ISO 23081-2:2007 (technisch document) gaan in op metadata voor archiefbeheersystemen en zijn mede opgesteld door diegenen die het metadatamodel van het CDD hebben opgesteld/gereviewd. Het metadatamodel van het CDD zal dan ook zo volledig mogelijk voorzien in de opslag van metadata over de context van digitale informatie. Daarmee is in de duurzame bewaring van de eigenschap context, één van de vijf eigenschappen van digitale informatie, voorzien in het CDD. Daarom is besloten deze eigenschap verder buiten het onderzoek te houden. Er ligt dus een focus op de overige eigenschappen van digitale informatie, namelijk inhoud, structuur, uiterlijk en gedrag.

Tijd

Vanwege de beperkte tijd voor het verrichten van onderzoek (totaal circa 6 maanden) zal in eerste instantie worden gericht op het bepalen van de te volgen overall bewaarstrategie. Daarna komen de organisatie-inrichting en architectuurontwerp aan de beurt. De nadruk ligt hierbij op het architectuurontwerp. Dit omdat deze de basisvoorziening moet bieden voor de organisatie. Er is echter wel een wisselwerking tussen het inrichtings- en ontwerpvragestuk waarmee goed rekening dient te worden gehouden. Vanwege de beperkte tijd voor het verrichten van onderzoek, wordt er geen projectplan opgesteld dat het implementatietraject beschrijft. Wel worden vervolgacties benoemd.

2.6. Onderzoeksplan

De in paragraaf 2.4 opgestelde onderzoeksvragen zijn gevormd uit kernbegrippen van de centrale onderzoeksvraag en zijn gericht op het oplossen van kennis- en handelingsproblemen. Deze vragen worden in volgorde beantwoord en er worden verbanden tussen de antwoorden gelegd. Dit is in figuur 12 gevisualiseerd, waarbij de nummers overeenkomen met die van de onderzoeksvragen.



Figuur 12 - Onderzoeksmodel

Het onderzoeksplan bevat de volgende componenten:

(a) Het door Deloitte afgeronde vooronderzoek en in uitvoering zijnde implementatie en theorie over digitale duurzaamheid en openbaarbestuur/administratieve organisatie leveren (b) de bewaarstrategieën en op hoofdlijnen de criteria die bij digitale bewaring van belang zijn. (c) Deze resulteren in de noodzaak voor een bewaarstrategie selectiemethode om de meest geschikte bewaarstrategieën te kunnen selecteren. Het CDD+ moet kunnen voorzien in de uitvoer en selectie van deze bewaarstrategieën en er met daarom worden gekeken naar hoe het CDD+ deze moet operationaliseren. Dit wordt gedaan aan de hand van hoe andere organisaties dit doen. Dit proces resulteert in een voorstel voor organisatie-inrichting en architectuurontwerp (d).

Om aan de nodige informatie te komen voor het onderzoek worden onderzoeksmethoden toegepast. Deze worden hieronder per onderzoeksvraag behandeld.

1. Welke bewaarstrategieën beschrijven hoe om te gaan met de duurzaamheid van digitale informatie en hoe worden deze in de praktijk gebracht?

Het is van belang te bekijken welke bewaarstrategieën er zijn rondom de duurzaamheid van digitale informatie en hoe deze bewaarstrategieën worden geoperationaliseerd. Vanuit bureauonderzoek worden in de literatuur bewaarstrategieën en casestudies gevonden die beschrijven hoe digitale informatie duurzaam toegankelijk te houden. Hierbij wordt veel gebruik gemaakt van de resultaten van de Testbed Digitale Bewaring [Test03-05]. Verder wordt het concept van migratiepaden toegelicht en hoe deze te toetsen zijn [Ferr05-06]. De praktijk cases worden besproken aan de hand van publicaties van (onderdelen van) projecten en concepten. Daarnaast wordt aanvullende informatie gevraagd bij instanties die zich bezig houden met deze projecten en concepten. De theorie en casestudy informatie is verkregen na bureauonderzoek naar literatuur en aanvullend door expertgesprekken.

2. Welke criteria liggen ten grondslag aan de duurzaamheid van digitale informatie?

Er zijn verschillende eigenschappen te onderscheiden die het lastig maken om digitale informatie voor langere tijd te bewaren. Het antwoord op deze onderzoeksvraag moet resulteren in een overzicht van de criteria die van invloed zijn. Dit overzicht zal echter met het oog op de veranderende techniek en wet-/regelgeving niet limitatief zijn. Om te komen tot het overzicht van criteria wordt gebruik gemaakt van onderzoek vanuit het DELOS project [Rauc05][Stro06] en hierop voortbouwende theorie van Strodl ea. [Stro07-1/-2].

Met de bewaarstrategieën uitgewerkt, in beeld hoe andere organisaties de bewaarstrategieën

operationaliseren en de criteria geclassificeerd, is het van belang uit te werken hoe de overall bewaarstrategie bepaald wordt. De vraag hierbij is:

3. Hoe voorziet een bewaarstrategie selectiemethode in de selectie van de meest geschikte bewaarstrategie?

Het antwoord op deze vraag resulteert in een methodiek die vanuit een uiteenzetting van gegroepeerde criteria en bewaarstrategie-alternatieven een bewaarstrategie voorstelt door een aangepaste utiliteitsanalyse uit te voeren [Rauch05]. Deze theorie is uiteengezet na bureauonderzoek naar literatuur die ten grondslag ligt aan het PLANETS project (inleiding hoofdstuk 5).

De twee kernbegrippen uit de centrale onderzoeksvraag die nog over blijven zijn “organisatorisch” en “ondersteunende informatiesystemen”. Deze wijzen op een organisatie inrichtingsvraagstuk en een architectuur ontwerp-vraagstuk. Het inrichten van een organisatie is echter niet hetzelfde als het ontwerpen van een architectuur. Er zullen dan ook verschillende manieren van onderzoek nodig zijn om deze tweeledigheid te beantwoorden vanuit onderzoeksvragen.

Hiervoor beschrijft Starreveld ea. [Star97] vanuit de bestuurlijke informatievoorziening een interessante overkoepelende aanpak voor beide vraagstukken. Zo wordt methodisch omschreven hoe vanuit de informatiebehoefte en informatiebeleid een organisatie is te ontwerpen met ondersteunende informatiesystemen en techniek. In processen worden organisatiedelen gekoppeld aan te ondersteunen informatiesystemen. Op deze manier wordt aangegeven op welke wijze de informatieverzorging van een organisatie zo effectief en efficiënt mogelijk kan worden ingericht, zodat de juiste gegevens op het juiste moment ter beschikking staan bij het uitvoeren van taken.

Door bovenstaande theorie te gebruiken wordt een onderzoek gedaan naar processen, taken, functies, informatiesystemen en gegevensdragers die in de te vormen organisatie aanwezig dienen te zijn. Deze elementen richten de organisatie in en zijn eventueel overzichtelijk in verband te brengen via een intersectiemodel [Star97][Bush87]. Operationeel kan de procesmodelleersoftware van Deloitte goed gebruikt worden (IndustryPrint Process Modeler) om de gewenste onderdelen in beeld te brengen en aan elkaar te relateren.

Deze aanpak resulteert hiermee in de volgende onderzoeksvragen:

4. Welke organisatie-inrichting is geschikt om deze bewaarstrategie te kunnen volgen?

5. Met welke architectuur is deze organisatie te ondersteunen?

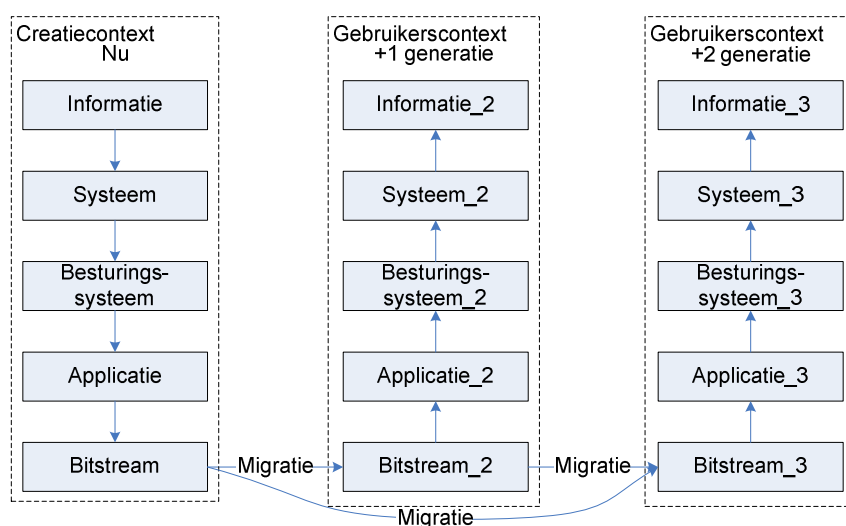
Bij de twee bovenstaande vragen moet in het achterhoofd gehouden worden dat de methode van Starreveld ea. werkt vanuit beleid richting systeemondersteuning. Er dient echter niet alleen te worden gekeken naar bestaande systemen, maar ook naar eventueel nieuw te ontwerpen systemen. Zo is er interactie tussen organisatie-inrichting en systeemontwerp (Henderson ea. [Hend93]). Nieuwe informatietechnologie kan bijvoorbeeld via te vormen informatiebeleid de infrastructuur en processen van de organisatie in grote mate beïnvloeden: de techniek maakt steeds meer mogelijk. Met deze samenhang die door Henderson ea. worden beschreven moet rekening worden gehouden. Door deze wederzijdse invloeden moet er na enig organisatie-inrichting ook begonnen worden met architectuurontwerp: het gaat grotendeels gelijk op. Architectuuronderdelen die gebruikt worden binnen de context van digitale duurzaamheid worden echter al besproken in de casestudies per bewaarstrategie (vraag 1). Er wordt op deze manier voorzien in de invulling van de architectuur die noodzakelijk is voor het operationaliseren van bewaarstrategieën.

Hoofdstuk 3 Bewaarstrategieën

Vanuit het voorgaande hoofdstuk is in gegaan op wat digitale informatie en duurzaamheid hiervan is. Om te onderzoeken welke bewaarstrategie JustID moet gaan volgen, moet er eerst uitgewerkt worden welke bewaarstrategieën er zijn, welke eigenschappen deze hebben en hoe deze in de praktijk gebracht worden. Dit hoofdstuk gaat daar op in door de twee overkoepelende bewaarstrategieën uit een te zetten: migratie en emulatie.

3.1. Migratie

De meest bekende en veelvuldig gebruikte bewaarstrategie is migratie. Bij migratie wordt de manier van digitale codering van informatie aangepast aan nieuwe hard- en/of software of aan een nieuwe generatie computer technologie [Task96]. De digitale bestanden waarin de informatie gecodeerd is wordt dus aangepast aan de omgeving. De relatie tussen de componenten die bijdragen in de representatie van digitaal gecodeerde informatie is bij toepassing van migratie (bestandsformaat conversie) als volgt weer te geven:



Figuur 13 - Relaties bij migratie, aangepaste [Test03]

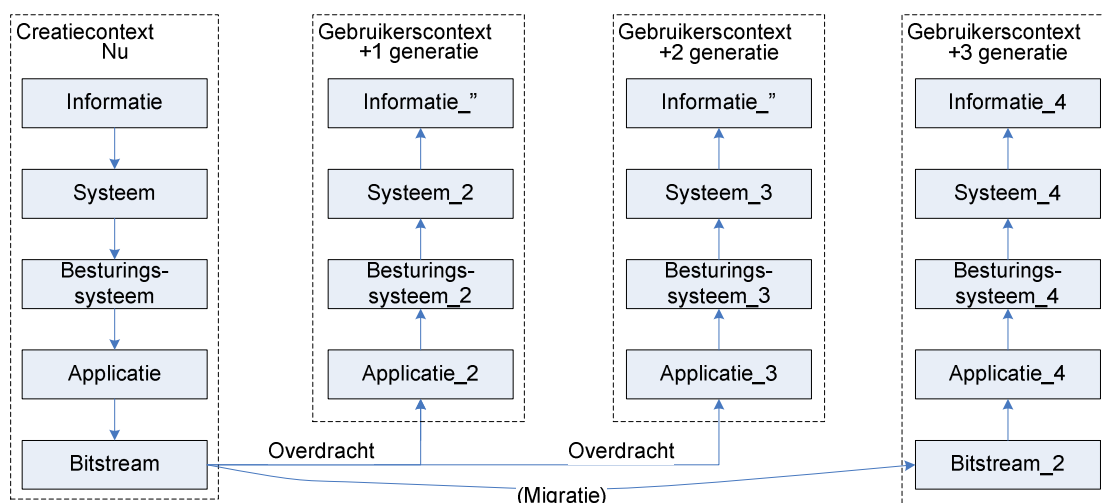
3.1.1. Soorten migratie

Er zijn verschillende soorten migratie te onderscheiden, onder te verdelen naar hardwarematig (HW) en softwarematig (SW), namelijk:

- Wijzigen van het media (HW). Hierbij wordt de bitstream verplaatst of gekopieerd naar andere media. Bestanden worden bijvoorbeeld vanaf een CD-ROM gekopieerd naar een harde schijf. Dit kan echter ook het migreren van digitale informatie in het CDD naar een fysiek ander archief of een in de toekomst te ontwikkelen nieuwe versie van het CDD omvatten. Het wijzigen van het media is alleen ter behoud van de bitstream en voorziet niet in de representatie van deze bits;
- 'Backward' compatibiliteit (SW). Hierbij is een toekomstige applicatie van eenzelfde leverancier zo gemaakt dat het over weg kan met de door een oudere applicatie opgestelde digitale bestanden. MS Word 2007 kan bijvoorbeeld ook MS Word '95 bestandsformaten lezen door

gebruik te maken van geïntegreerde conversiesoftware;

- Interoperabiliteit (SW). Hierbij kunnen meerdere applicaties overweg met dezelfde digitale bestanden. Zo kan OpenOffice 2 grotendeels ook MS Word 2005 bestandsformaten representeren. Een ander voorbeeld zijn plain-text bestandsformaten die te openen zijn in software onder Windows, maar ook in software onder Linux. Interoperabiliteit wordt vergemakkelijkt door het gebruik van open standaarden;
- Conversie (SW). De oorspronkelijke bestandsformaten worden omgezet naar andere bestandsformaten. Meestal wordt gekozen voor een goed gedocumenteerd open bestandsformaat dat stabiel is en niet afhankelijk van één leverancier ofwel: open standaarden (figuur 14). De levensverwachting van dergelijke bestandsformaten wordt op die manier verlengd en dit vermindert het aantal formaat conversies (migratie), en daarmee het risico van informatieverlies. In figuur 14 wordt met “Informatie_” aangegeven, dat er een deels andere representatie van de informatie plaatst vindt.



Figuur 14 - Gebruik open standaarden vermindert benodigde bewaaracties [Test03]

3.1.2. Migraties en migratieacties

Migratieacties kunnen enerzijds worden uitgevoerd bij de aanbieder partij, door op die locatie gebruik te maken van bijvoorbeeld bestandsformaat conversiesoftware, en anderzijds in het archief tijdens de inname, opslag of uitgifte, door lokaal gebruik te maken van bestandsformaat conversiesoftware. Deze verschillende soorten migraties kunnen daarnaast op verschillende momenten worden toegepast, namelijk:

- Tijdens inname. Er kan voor gekozen worden om digitale informatie bij inname in het archief meteen te migreren naar een of meerdere alternatieve bestandsformaten, ook wel manifestaties, zoals naar een PDF/A-bestandsformaat. Meerdere manifestaties van digitale informatie kunnen volgens Thibodeau [Thib02] dezelfde informatie weergeven. Zo kan informatie in een MS Word-bestand ook worden gerepresenteerd in een PDF-bestand en/of OpenOffice-bestand.
- Gedurende de opslag. Er kan besloten worden om gedurende de opslag de informatie te migreren. Dit kan bijvoorbeeld worden besloten nadat bekend is geworden dat een bepaald bestandsformaat aangemerkt kan worden als duurzamer. Er kan dan besloten worden om de digitaal gecodeerde

informatie te converteren naar het duurzamere alternatief.

- Tijdens uitgifte. De informatie kan ook gemigreerd worden tijdens de uitgifte. Dit betekent dat zodra een eindgebruiker bepaalde informatie wil inzien, de manifestatie direct gecreëerd wordt ('on-the-fly') via conversiesoftware vanuit het origineel of een andere manifestatie, en wordt verstrekt aan de eindgebruiker.

Om gebruik te kunnen maken van de verschillende migratiemogelijkheden zijn er tijdig acties nodig, de zogenaamde bewaaracties. Per migratie bewaarstrategie zijn er de volgende te onderscheiden:

- Wijzigen van het media. De eigenschappen en specificaties van het doel media (hardware) moeten goed beschreven en getest zijn. Er moet getoetst worden welk doel media het beste is;
- 'Backward' compatibiliteit. Softwareontwikkelaars moeten worden aangemoedigd de backwards compatibiliteit met oudere bestandsformaten in nieuwe applicaties in te bouwen;
- Interoperabiliteit. Softwareontwikkelaars moeten worden aangemoedigd om ook bestandsformaten van concurrerende software te ondersteunen. Dit wordt bemoedigd wanneer het gaat om bestandsformaten die als open standaard zijn aan te merken;
- Conversie. De softwarematige eigenschappen en specificaties van originele en doel bestandsformaten moet uitgekristalliseerd zijn. Hiervoor moeten applicatieontwikkelaars duidelijk specificaties documenteren. Vervolgens moet er een conversietool worden geschreven en moeten bestaande conversietools te kunnen worden getest.

Alleen op het wijzigen van het media en het uitvoeren van conversies kan JustID directe invloed uitoefenen. Voor de overige twee bewaarstrategieën kan JustID alleen lobbyen bij software- en hardwareontwikkelaars.

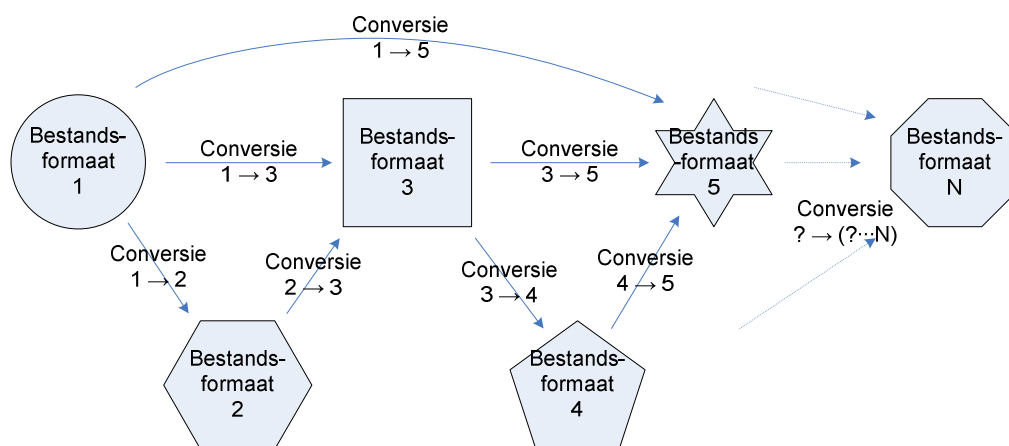
Aangezien het wijzigen van het media alleen voorziet in bitstream bewaring en niet direct in de bewaring van de representatiemogelijkheid, gaat de rest van dit hoofdstuk niet over dit type migratie.

3.1.3. Gevolgen van migratie

Toepassing van de migratiebewaarstrategie wordt als nuttig beschouwd wanneer de inhoud van informatie centraal staat [Test05]. Het zeker stellen van de authenticiteit (de oorspronkelijke 'look and feel') is lastiger. Dit komt doordat de omgeving rondom de informatie iedere keer bij een migratie iets zal veranderen, en daarmee ook de informatie zelf (zie ook figuur 13). Met omgeving wordt hierbij het bestandsformaat bedoeld waarin de informatie is gecodeerd en de interactie van de gebruiker met de software en hardware. Dit enige verlies van informatie is ook meteen een van de bezwaren tegen het gebruik van de migratie bewaarstrategie. Er wordt gesteld dat het migratie resultaat (X_{tn}) nooit gelijk kan zijn aan het origineel (X_{t0}) en dat er altijd in zekere mate een verandering (ΔX):

$$X_{tn} = X_{t0} + \Delta X \quad [\text{Thib02}]$$

Hoe goed een migratieresultaat is, is moeilijk te voorspellen. Er zijn door een verscheidenheid aan conversiesoftware en bestandsformaten meerdere manieren om van een bestandsformaat de informatie te migreren naar mogelijk duurzamere bestandsformaten. Deze verscheidenheid van mogelijkheden en alternatieven wordt uitgedrukt in migratiepaden. Ferreira visualiseert migratiepaden als weergegeven in figuur 15 [Ferr05]. Deze figuur beschrijft enkele bestandsformaatconversies, waarbij iedere pijl een conversie voorstelt. De wisselende vorm van de bestandsformaten 1 tot en met N geeft weer dat ieder informatieobject in kleine mate verschilt van de ander. Hoe met deze kleine verschillen om te gaan wordt toegelicht in paragraaf 3.1.7.



Figuur 15 - Migratiepaden [Ferr05]

In verhouding tot emulatie (paragraaf 3.2) wordt een bestandsformaat conversie als goedkoop en eenvoudig aangemerkt. Het is echter ook de meest bekritiseerde bewaarstrategie. Naast dat de resultaten enigszins onvoorspelbaar zijn en moeilijk getest kunnen worden, zijn er nog een tweetal bezwaren tegen migratie:

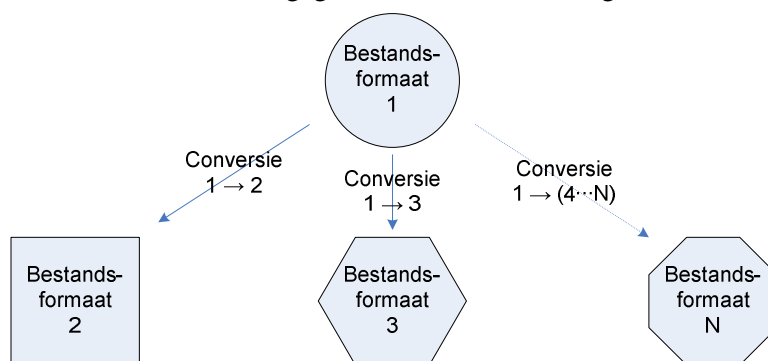
- Migratie kan van invloed zijn op de authenticiteit van digitale informatie. Bij het ‘authentiek’ duiden van een informatieobject, moeten de betekenis en de geldigheid worden gewaarborgd. Verlies van authenticiteit kan zowel juridische als archivistische implicaties hebben.
- Migratie moet met enige regelmaat worden herhaald.

Ondanks deze bezwaren is migratie een goede bewaarstrategie voor organisaties die bepaalde informatietypen, zoals databases, spreadsheets of documenten voor de korte termijn (maximaal tien jaar) willen bewaren [Test05]. De verwachte levensverwachting is echter verschillend per bestandsformaat en kan worden gerekt door informatie op te slaan in open standaarden.

Daarnaast voorziet het testen van migratiepaden (uitvoeren van experimenten) en het scoren hiervan via de bewaarstrategie-evaluatiemethode er in dat er onderbouwd een keuze kan worden gemaakt voor het meeste geschikte migratiepad.

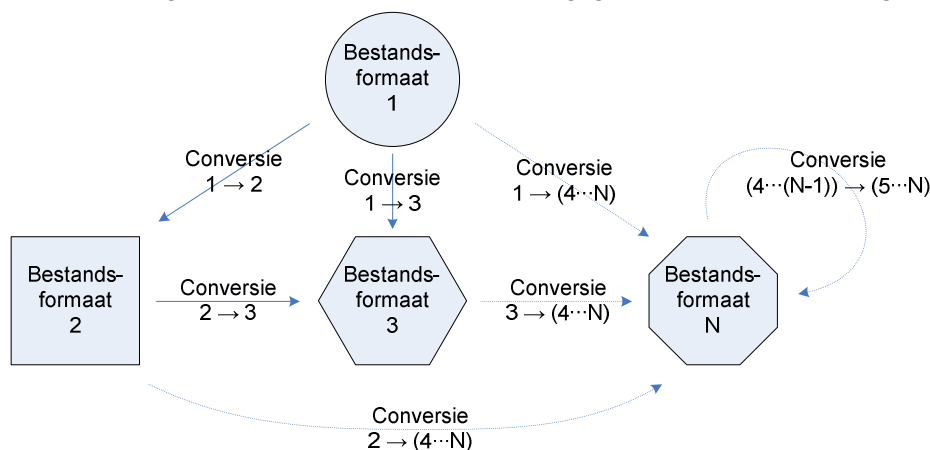
3.1.4. Aantallen migratiepaden en experimenten

Bij N aantal bestandsformaten die mogelijk duurzamer zijn dan het originele bestandsformaat, zijn minimaal N migratiepaden te overwegen. Hierbij wordt uitgegaan van de migratie van het origineel naar alternatieve manifestaties, zoals weergegeven in onderstaande figuur.



Figuur 16 - Minimaal aantal migratiepaden

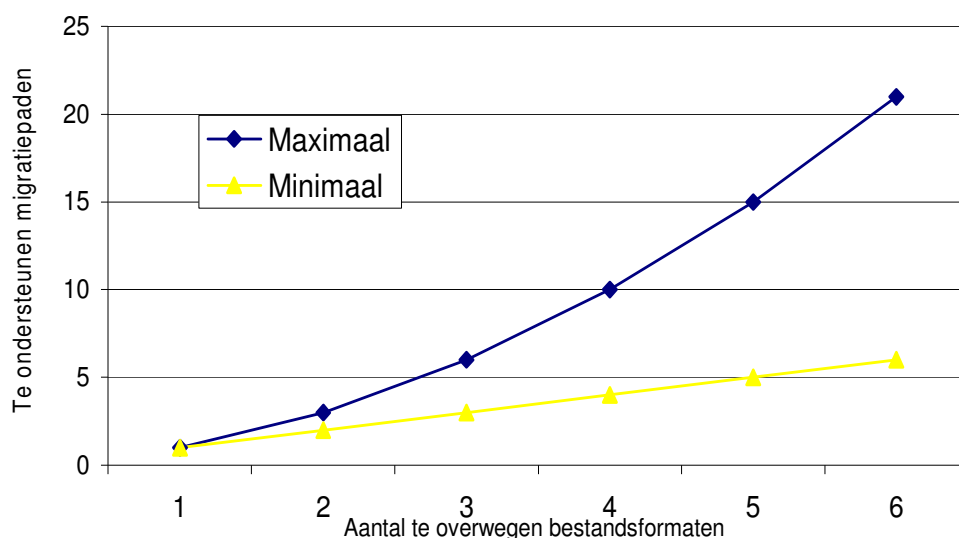
Wanneer er echter sprake is van meerdere ‘digital born’ bestanden van het zelfde informatietype, zoals documenten van het MS word formaat en OpenOffice formaat, moet er vanuit beide originele bestandsformaten gemigreerd worden naar mogelijke duurzamere. Om al deze migraties mogelijk te maken vanuit oude bestandsformaten naar alle te overwegen nieuwere, waarbij alleen directe migratiepaden gebruikt worden, geeft $(N+1)*N/2$ het aantal migratiepaden dat overwogen moet worden. Logisch beredenerend stelt dit het maximaal te ondersteunen migratiepaden vast bij een bepaald aantal te overwegen bestandsformaten. Dit is weergegeven in onderstaande figuur:



Figuur 17 - Aantal migratiepaden bij alleen de overweging van oude naar nieuwe bestandsformaten

Het kan voorkomen dat opeenvolgende migratiepaden overwogen moet kunnen worden alsof zij een enkel migratiepad zijn, bijvoorbeeld wanneer de implementatie voor een direct migratiepad mist. Er zijn bijvoorbeeld geen migratietools beschikbaar. Via andere migratiepaden kan dan informatie toch gecodeerd worden in het doel bestandsformaat. Het gebruik van opeenvolgende migratiepaden binnen één experiment draagt niet bij aan het maximale aantal te ontwikkelen migratietools. Alleen de directe migratiepaden hebben migratietools nodig; de indirecte paden zijn samen te stellen door de migratietools achtereenvolgens toe te passen.

Het aantal migratiepaden dat onderzocht moet worden tijdens bewaarexperimenten, kan uitgezet worden tegen de te overwegen bestandsformaten. Deze uiteenzetting is weergegeven in grafiek 1. Omdat migratieresultaten per migratietool kunnen verschillen, zijn er altijd evenveel of meer migratietools nodig dan dat er minimaal aan migratiepaden ondersteund moet worden. De migratietools kunnen intern worden ontwikkeld of extern worden aangeschaft.



Grafiek 1 - Migratiepaden per alternatieve bestandsformaten (migratiestrategie)

Het is overigens niet logisch om alle mogelijke migratiepaden te ondersteunen met migratietools. Migratie richt zich op de conversie van vele bestandsformaten naar open standaarden, bijvoorbeeld PDF/A-1a, OpenDocumentFormat/OpenXML of een ander XML formaat. Dit proces kan worden aangeduid als normalisatie en vereist een viewer om de XML te kunnen representeren, zoals Adobe PDF Reader, de normaliseertool XENA (XML Electronic Normalising of Archives) van de National Archives of Australia [24] of de Metadata Extraction tool van de National Library of New Zealand [17].

Puur voor de representatie bij een ketenpartner kan het zijn dat men het archiefstuk wil ontvangen in een alternatief bestandsformaat. Het origineel of de archivistische manifestatie kan dan gemigreerd worden naar het gevraagde bestandsformaat.

3.1.5. Project(en)

Veel van de projecten die digitale duurzaamheid onderkennen claimen migratie als strategie te voeren. Er zijn echter weinig projectbeschrijvingen die concreter worden dan slechts dit gegeven. Binnen het LOCKSS project, wat staat voor “Lots Of Copies Keeps Stuff Safe” [Mani05], wil men migratie bij toegang toepassen [Rose05]. Hiervoor heeft men een proof-of-concept gemaakt voor de conversie van GIF-naar-PNG, maar men beschrijft nog niet hoe dit proces geautomatiseerd uitgevoerd kan worden. Er is dan ook gekozen om de toepassing van migratie verder toe te lichten aan de hand van het project van Ferreira, genaamd CRiB [Ferr05]. Alleen daar waar relevant worden andere projecten, zoals LOCKSS, gebruikt. CRiB is een Service Oriented Architecture (SOA) en ontworpen om instanties die digitaal erfgoed willen bewaren te ondersteunen met een migratieservice. Het project is nog volop in ontwikkeling. Ferreira beschrijft de elementen van het systeem, waarbij ook toetsingscriteria aan bod komen om tot automatische toetsing te komen (zie volgende paragraaf) [Ferr05]. Het project maakt hiervoor gebruik van andere applicaties en koppelingen naar andere informatiesystemen en registers.

3.1.6. Services

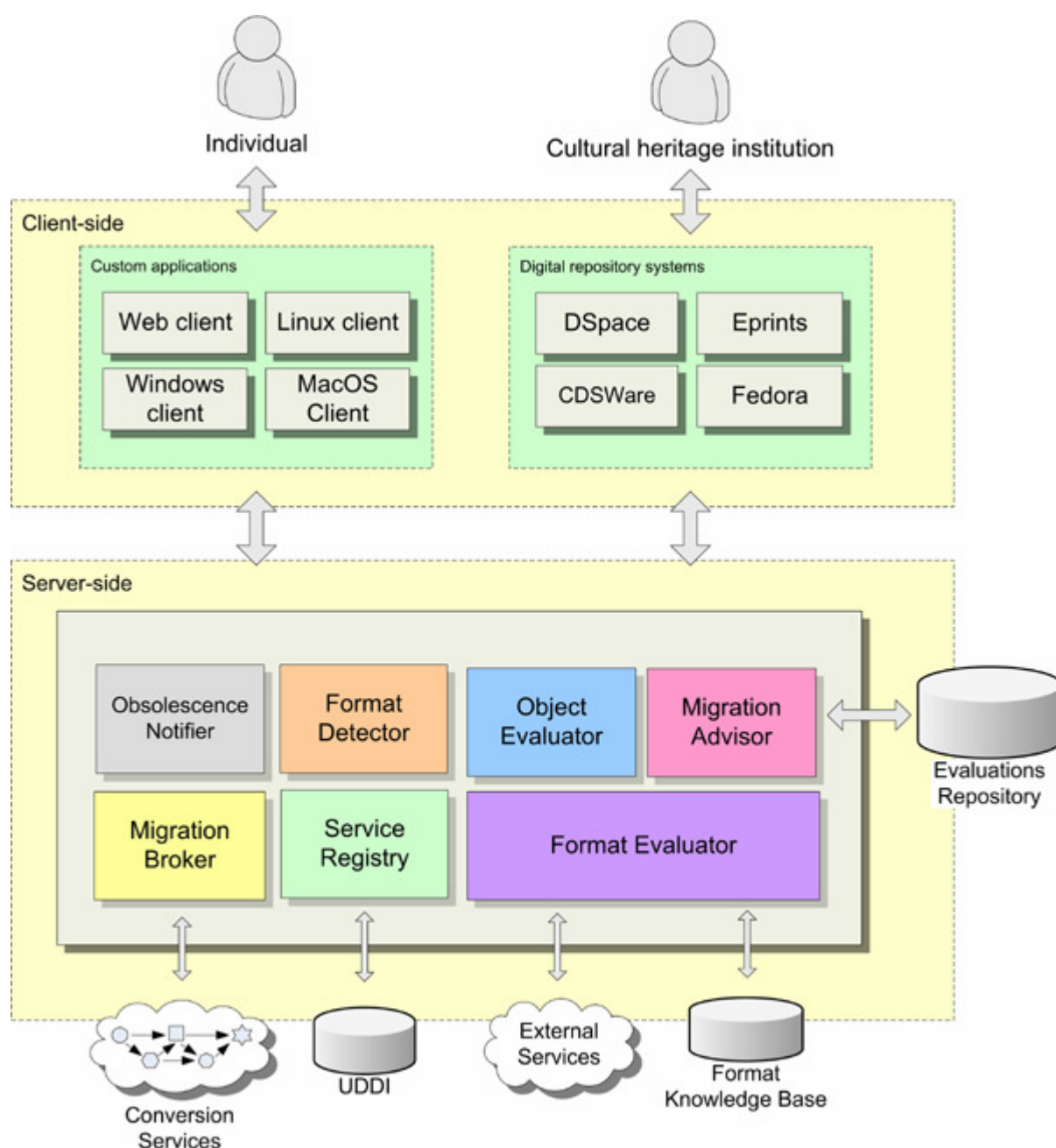
CRiB is gericht op het geven van advies rondom digitale duurzaamheid en het doen van migraties. Ferreira ea. [Ferr06] identificeren hiervoor de volgende services:

- Bestandsformaat herkenningsservice. Deze service moet de digitale bestanden herkennen (informatietype, software versie), maar ook de inhoud van de digitale bestanden bepalen en valideren;
- Aanbevelingsservice. Een service die tot een onderbouwde aanbeveling komt voor de te kiezen migratietool en het doel bestandsformaat;
- Migratie-uitvoerservice. De service die daadwerkelijke de originele digitale bestanden converteert met een migratietool naar de doel bestandsformaten;
- Migratie-evaluatieservice. Een service die het informatieverlies bepaalt dat gemoeid is met de migratie.
- Informatieservice. Een service die informatie geeft over welke bestandsformaten het risico lopen te verouderen;

Het PLANETS project identificeert ook deze services, maar richt zich naast migratieservices ook op emulatieservices.

3.1.7. Toepassing

Om de beschreven services te realiseren heeft Ferreira ea. [Ferr06] een architectuur uitgedacht die onderdelen verbindt die spelen in de context van digitale duurzaamheid. Onderstaande afbeelding geeft deze architectuur weer.



Figuur 18 - CRiB SOA architectuur [Ferr06]

De client-side kan hierbij worden gezien als het CDD en de server-side als het systeem dat de migratie bewaarstrategie in praktijk brengt. De elementen van de server-side worden hieronder toegelicht aan de hand van het artikel van Ferreira ea. [Ferr06] en eigen onderzoek.

Informatieservice (Obsolescence notifier)

De informatieservice moet ketenpartners en de interne organisatie informatie verstrekken over de verwachte levensduur van digitale bestanden en mogelijke belangrijke gebeurtenissen in het kader van digitale duurzaamheid. Ferreira ea. identificeren een drietal bronnen die hiervoor gebruikt kunnen worden:

- Lawrence ea. [Lawr00] hebben een onderzoek gedaan naar de invloed van migratie op de

integriteit van bestandsformaten. Het biedt hiermee een mogelijke opzet voor een inschatting van het risico;

- Stanescu [Stan04] beschrijft de INFORM methodologie. Een benadering van de mate van duurzaamheid van digitale bestanden;
- Experts. Een groep experts kan zich richten op het volgen van trends en nieuwe technologieën. Enkele instanties die dit al doen zijn DigiCULT met haar Technology Watch Reports [13] en de Digital Preservation Coalition met haar maandelijkse blad [14].

Bestandsformaat- en inhoudbepaling (Format Detector)

Dit gedeelte van de architectuur moet voorzien in de herkenning van bestandsformaten en de bepaling van de interne onderdelen van de digitale bestanden: de bitstreams. Deze worden gebruikt tijdens de inname van bestanden en bij het evalueren van migratiestrategieën. Er zijn enkele applicaties geëvalueerd die in enige mate voorzien in bestandsformaat herkenning en inhoudelijke bepaling (zie bijlage 6). De volgende zijn aan te raden om in te zetten op dit onderdeel van de architectuur:

- DROID [15]. DROID staat voor Digital Record Object Identification en is een initiatief van The National Archives (UK). DROID werkt samen met bestandsformaat database PRONOM (zie onderstaande “Service registers” onderdeel) om bestanden te kunnen herkennen aan de hand van enkele verschillende technieken.
- JHOVE [Jhov06][16]. JHOVE staat voor “JSTOR/Harvard Object Validation Environment” en is een initiatief van JSTORE en Harvard University. Deze tool richt zich op het bepalen van de interne onderdelen waaruit bestanden opgebouwd zijn. Zo worden lettertypen, images, URL’s, filters, metadata-info, enz. per pagina/object herkend. Een TIFF afbeelding kan bijvoorbeeld worden herkend in een MS Word document. Deze TIFF afbeelding kan op zijn beurt worden onderverdeeld naar een TIFF bitstream, ICC kleur profiel en XMP metadata. Deze uiteenrafeling kan worden gebruikt bij het invullen van criteria en vandaar uit het scoren van bewaarstrategieën. JHOVE2 beoogt gebruikt te maken van DROID voor de bestandsformaat herkenning;
- Metadata extraction tool [17]. Metadata extration tool is een initiatief van de National Library of New Zealand en is vergelijkbaar aan JHOVE: het bepaald de interne objecten.

Service registers (Service Registry)

Het onderdeel service registers is verantwoordelijk voor het beschikbaar hebben van informatie over bestandsformaten, softwareleveranciers, migratietools en andere metadata die mogelijk van belang zijn in het kader van de representatie van digitale gecodeerde informatie [Day03]. Ferreira ea. identificeren vier service registers die overwogen kunnen worden:

- PRONOM [Darl03][Brow06][18]. Een initiatief van The National Archives (UK) dat zich richt op het vergaren van informatie over alle bestandsformaten, software leveranciers, besturingssystemen, hardware componenten en andere technische onderdelen die spelen rond digitale duurzaamheid. Momenteel bevat PRONOM informatie rondom ongeveer 500 bestandsformaten en werkt samen met DROID (zie bovenstaand “bestandsformaat herkenning” onderdeel);
- Digitale Formats Web [19]. Een initiatief van de Library of Congress (US) dat zich richt op informatie over digitale bestanden;
- GDFR registry [Abra05][20]. Een initiatief van Harvard University Library dat als doel heeft het

ontdekken en leveren van representatie informatie over digitale bestandsformaten. Eind 2008 zou dit register gerealiseerd moeten zijn, waarbij PRONOM mogelijk een onderdeel van dit register gaat vormen;

- Representation Information Registry/Repository [Rusb05][Giar05][21]. Een initiatief van het Digital Curation Centre (UK) om te voorzien in representatie informatie op de manier van het OAIS.

Migratie-uitvoering (Migration Broker)

Dit deel van de architectuur is verantwoordelijk voor het uitvoeren van migraties. Dit kan extern (door derden) of intern worden uitgevoerd. In het kader van de privacy gevoeligheid van de informatie in het CDD, zal het absoluut de voorkeur hebben om de bestandsmigraties intern uit te voeren.

Ferreira ea. zijn bezig met een prototype dat de verschillende migratietools en migratietool aanbieders kan vergelijken. Dit zijn de zogenaamde bewaaractieprocescriteria die ook worden meegenomen in de methode bewaarstrategie evaluatie methode van Strodl ea. [Stro07]. Het prototype van Ferreira ea. toetst de volgende criteria: beschikbaarheid, stabiliteit, doorvoer/tijd, kosten, resulterende bestandsgrootte, resulterend aantal bestanden.

Bestandsformaat evaluatie (Format Evaluator)

Dit onderdeel van de architectuur moet voorzien in feiten over de huidige status van bestandsformaten. Deze informatie stelt de migratie adviesverstrekker in staat te beoordelen welk bestandsformaat mogelijk beter is. De PRONOM registry voorziet hier in.

Het prototype van Ferreira ea. is in staat de utiliteit te berekenen, net als via bewaarstrategie evaluatie methode van Strodl ea. [Stro07]. Dit is in het prototype mogelijk op basis van de volgende criteria: marktaandeel, ondersteuningsniveau, standaard, open specificaties, compressie, lossy compressie, transparantie, ingebedde metadata, royalty vrij, open source, backwards compatible, documentatie niveau, soortgelijke bestandsformaten, DRM, update frequentie, eigen bestandsextensies, leeftijd, transparant leesbaar, aantal leveranciers reader, open source reader, interoperabiliteit.

Uitbreiding van de criteria is mogelijk, dus de gedachtegang van het werken met een criteriaboom (paragraaf 5.1) is mogelijk. De criteria die binnen CRiB ondersteund worden kunnen zich ook bevinden in de criteria categorieën van de criteriaboom.

Migratieresultaat evaluatie (Object Evaluator)

Migratieresultaten moeten geëvalueerd worden. Dit wordt gedaan door te definiëren wat de digitaal gecodeerde informatie tot een unieke entiteit maakt. Het migratieresultaat moet hieraan voldoen. Voorbeelden zijn het aantal karakters waar uit een document bestaat, de volgorde van deze karakters, het aantal pagina's en de grafische weergave. Bestandsformaat- en inhoudbepaling tools worden gebruikt om de uniekheden te bepalen voor en na de migratie.

Migratieadviseur (Migration Advisor)

De migratieadviseur is verantwoordelijk voor het bepalen van het beste migratiealternatief. Het gebruikt hiervoor de bewaareisen die de klant heeft gesteld (criteria en het belang van deze via weegfactoren) en de resultaten van de geteste migratiebewaarstrategieën. Deze resultaten worden verworven uit de verschillende sets van criteria die samen de criteriaboom vormen en door het ondernemen van experimenten (zie ook onderstaande onderdeel "evaluatie register").

De migratieadviseur gebruikt de bewaareisen en gegevens uit het evaluatieregister om via een aangepaste utiliteitsanalyse te komen tot een score per bewaarstrategie. Dit is dezelfde methode die Rauch ea. [Rauc05] en Strodl ea. [Stro07-1/-2] beschrijven en bevat de volgende stappen (overeenkomstig met de stappen in figuur 19):

1. Voor ieder migratiepad (Aangegeven met “MP1 en MP2 in Figuur 19) wordt een standaard of gemiddelde score op de evaluatie criteria vastgesteld;
2. De gemiddelden van migratiepaden per criteria worden getransformeerd naar een schaal van 0 tot 1 om deze onderling te kunnen vergelijken;
3. De klant geeft weegfactoren aan per criteria;
4. De getransformeerde gemiddelden van migratiepaden per criteria worden vermenigvuldigd met de weegfactoren die de klant heeft gesteld. Het migratiepad met de hoogste score, die tevens op alle criteria een acceptabele scoren heeft, is de beste migratiebewaarstrategie.

Evaluation criteria		MP1	MP2			Weights	MP1	MP2
Migration Broker	Availability	0.95	0.9	1.0	0.0	0.2	0.2	0.0
	Stability	0.8	0.9	0.0	1.0	0.2	0.0	0.2
	Throughput	1 Kb/s	2 Kb/s	0.0	1.0	0.2	0.0	0.2
	Cost	0.01 €	0.05 €	1.0	0.0	0.3	0.3	0.0
	Outcome size	1.2	0.9	0.0	1.0	0.07	0.0	0.07
Format Evaluator	File Count	1.0	1.0	1.0	1.0	0.03	0.03	0.03
	Market share	0.5	0.9	0.0	1.0	0.2	0.0	0.2
	Support level	1.0	0.2	1.0	0.0	0.2	0.2	0.0
	Is standard	1.0	0.0	1.0	0.0	0.2	0.2	0.0
	Open specification	1.0	1.0	1.0	1.0	0.3	0.3	0.3
Format Evaluator	Supports compression	0.0	1.0	0.0	1.0	0.1	0.0	0.1
	...	...	...	...	...	...	...	...
	Textual content	0.9	0.95	0.0	1.0	0.4	0.0	0.4
	Page size	0.6	0.9	0.0	1.0	0.2	0.0	0.2
	Number of pages	1.0	1.0	1.0	1.0	0.2	0.2	0.2
Format Evaluator	Graphical layout	0.8	0.6	1.0	0.0	0.1	0.1	0.0
	Font type & size	0.5	0.8	0.0	1.0	0.1	0.0	0.1
	...	...	...	...	...	...	...	...
							1,53	2.0

(1) Standard behaviour of each migration path
(2) Normalised behaviours [0-1]
(3) User assigned weights of importance
(4) Overall rating

Figuur 19 - Stappen tijdens het scoren van migratiealternatieven

Evaluatieregister (Evaluations Repository)

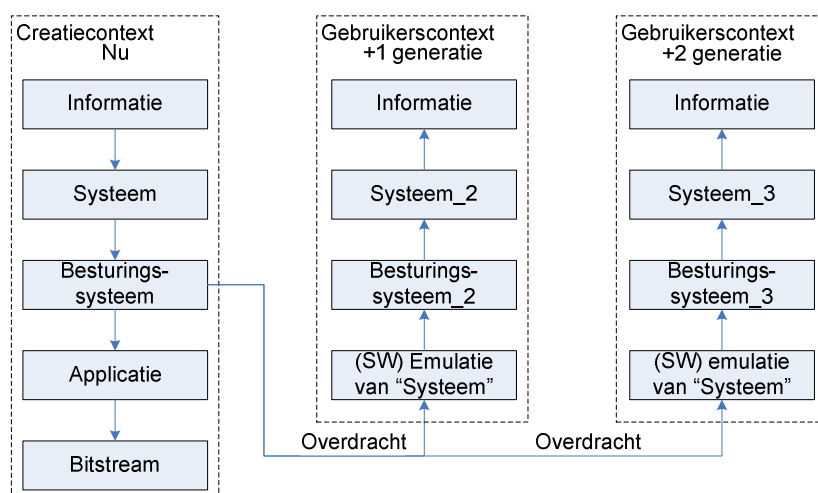
Om verantwoord tot het beste migratiealternatief te komen, gebruikt de migratieadviseur het evaluatieregister. Dit register bevat alle resultaten van experimenten die uitgevoerd zijn door de onderdelen die verantwoordelijk zijn voor de migratieresultaat evaluatie, bestandsformaat evaluatie en migratie uitvoering. De inhoud van het evaluatieregister is erg belangrijk en wordt verkregen door training. Deze training omvat het zelf opzetten van enkele migratiebewaarstrategie bewaarscenario's. Hierbij verwacht men bepaalde resultaten, en aan de hand van deze verwachting kan men het systeem trainen.

3.1.8. Samengevat

Migratie is de meest gebruikte bewaarstrategie die projecten claimen na te leven. Een voordeel van deze bewaarstrategie is dat het relatief goedkoop is om migratietools te maken en te gebruiken. Een belangrijk nadeel is echter dat na iedere migratie iets van de informatie zal veranderen of verdwijnen. Er is echter in dit gedeelte over migratie geschetst dat er systemen (in ontwikkeling) zijn die het mogelijk maken resultaten geautomatiseerd te beoordelen. De migratiestrategie met het beste resultaat kan dus gekozen worden. Dit is echter te reduceren tot een absoluut minimum door gebruik te maken van standaarden (langere life cycle) en de 'atomaire' delen waaruit bestandsformaten bestaan individueel te migreren (geen conversie van delen van bestandsformaten die niet geconverteerd dienen te worden).

3.2. Emulatie

Emulatie is een bekende maar erg weinig gebruikte bewaarstrategie. Bij deze bewaarstrategie worden nieuwe hardware en software zo ingericht zich te gedragen als de oude hard-/software, zodat deze de oorspronkelijke informatie kunnen weergeven. Omdat de interpretatie/decoding van de digitale informatie hetzelfde geschiedt als dat dit door de originele software werd gedaan, wordt met hogere zekerheid dan bij migratie de oorspronkelijke 'look and feel' van de digitale informatie gewaarborgd en daarmee de authenticiteit. Via emulatie kunnen dus de oorspronkelijke weergaveprogramma's of de methode hiervan op toekomstige computers worden uitgevoerd. De relatie tussen de componenten die bijdragen in de representatie van digitaal gecodeerde informatie is bij toepassing van (hardware)emulatie als volgt weer te geven:



Figuur 20 - Relaties bij emulatie, aangepaste [Test03]

De emulatie bewaarstrategie wordt veelal geassocieerd met duur en complex. De Koninklijke Bibliotheek (Den Haag) en IBM hebben echter ervaring met de emulatiebewaarstrategie en stellen dat de kosten van emulatie op langer termijn lager zijn dan die voor migratie [Oltm05]. Ze concluderen dat emulatie in tegenstelling tot migratie een veel hogere initiële investering vereist, maar dat dit door het delen van emulatoren en het niet constant noodzakelijke ontwikkelen van migratietools, zich terug betaald.

3.2.1. Soorten

Er zijn enkele verschillende soorten emulatie te onderscheiden [Test03][Dies02], namelijk:

- **Hardware-emulatie.** Door het softwarematig nabootsen van de hardwarematige omgeving, wordt voorkomen dat er in de toekomst nieuwe weergaveprogrammatuur moet worden geschreven die de verouderde bestandsformaten weergeven. Juist deze weergaveprogrammatuur en de gedetailleerde manier van interpretatie van de bestandsformaten is complex en vergt onderzoek en/of reverse engineering. Bij toepassing van hardware-emulatie is deze kennis niet of nauwelijks nodig. Er moet dan ‘slechts’ één (samengestelde) emulator worden geschreven per computersysteem (inclusief randapparatuur): ook een complexe onderneming. Hardware-emulatie wordt echter wel succesvol gebruikt bij het ontwikkelen van software en om oude computerspelletjes te draaien op nieuwe hardware. Een van de eerste emulatoren, ontwikkeld in 1997, die nog altijd succesvol is, is de Multiple Arcade Machine Emulator (MAME). De meeste recente ontwikkelingen rondom hardware-emulatie is de vormgeving van een modulaire emulator. Deze emulator wordt samengesteld uit deel-emulatoren, waarbij deze deel-emulatoren zorgen voor de emulatie van hardwareonderdelen. Op deze manier blijft men bij de basis van de Von Neumann architectuur en ontstaat een zeer flexibele emulator (zie bijlage 7);
- **Besturingssysteem-emulatie.** Een software-emulator kan er voor zorgen dat een compleet (oud) besturingssysteem kan draaien onder een nieuw besturingssysteem. Het biedt echter slechts de mogelijkheid oude applicatie te draaien via de oude manier van opstarten en op vergelijkbare hardware. Het schrijven een besturingssysteem-emulator vergt allereerst reverse engineering en daarnaast moeten de gebruikers weten hoe het oude besturingssysteem werkt om applicaties te kunnen gebruiken.
- **Toepassingsprogrammatuur-emulatie.** Een software-emulator kan in de taken voorzien die anders het originele toepassingsprogramma uitvoert. De toepassing uit zich bijvoorbeeld in een nieuwe applicatie die de oude bestandsformaten van (te) oude applicaties kunnen weergeven op soortgelijke manier. Omdat veel bestandsformaten gesloten zijn, vergt deze emulatie methode reverse engineering.
- **Universal Virtual Computer (UVC).** De UVC combineert de hardware emulatie- en migratiebewaarstrategie. Er wordt nog altijd een emulator geschreven, maar dan voor een virtuele versimpelde computer [Lori01-02]. Deze versimpelde computer bestaat uit geëmuleerde vereiste basis hardware (gesegmenteerd geheugen, registers en een set van low-level instructies). Deze UVC draagt via een voor ieder bestandsformaat te schrijven applicatie (de UVC-dataformaatdecoder) zorg voor het conversie van de oorspronkelijke digitale informatie naar een Logische Data Beschrijving (LDB). De LDB is een taal die de inhoud beschrijft aan de hand van labels, zoals XML. De LDB's kunnen in de toekomst via een viewer worden geïnterpreteerd zonder extra hulpmiddelen.

De migratiebewaarstrategie wordt toegepast bij het opvragen. Er vindt dan een conversie plaats van het originele bestand naar een gedefinieerde standaard (de LDB), die geïnterpreteerd kan worden door een viewer. De UVC is zelf de enige schakel die per hardware omgeving moet worden ontwikkeld. Het decodeerprogramma moet worden ontwikkel per bestandsformaat. In de toekomst kunnen relatief eenvoudig viewers worden gemaakt die de LDB's interpreteren. Desgewenst kan de weergave van de informatie via de viewers mee veranderen door de tijd heen om te voldoen aan een zo goed mogelijke menselijke interpretatie.

Per emulatiesoort is aan te geven wat er ontwikkeld zou moeten worden en in welke mate hergebruik mogelijk zou zijn. Zo moet nieuwe hardware bij hardware emulatie iedere keer weer opnieuw in de emulator worden verwerkt. Als de emulator op andere hardware gaat draaien, moet de emulator op zijn beurt worden omgeschreven naar deze nieuwe hardwareomgeving. Dit zelfde geldt voor besturingssysteememulatie. Deze is afhankelijk van de hardware die fysiek of virtueel beschikbaar is. Toepassingsprogrammatuur emulatie gaat een stap verder. De emulator moet per besturingssysteem worden herschreven of in een platformonafhankelijke taal geschreven zijn. Daarnaast kan deze afhankelijk zijn van hardware of besturingssysteem elementen die fysiek of virtueel aanwezig zouden moeten zijn.

Bij de UVC is de hardware altijd al virtueel. Er moet per hardware architectuur een UVC emulator worden geschreven, die vervolgens de voor de UVC geschreven programmatuur kan draaien. Daarnaast moeten viewers geschreven worden voor minimaal ieder informatietype [Test03-2].

In onderstaande tabel (tabel 4) is weergegeven voor welke representatielaag er herhaaldelijk ontwikkeling nodig is en voor welke lagen de emulator vervolgens ondersteuning biedt.

Emulatiesoort	Hardware	Besturingssysteem	Toepassings-programmatuur	UVC
Representatielaag				
Hardware	Ontwikkelen	-	-	Ontwikkelen
Besturingssysteem	Ondersteuning	Ontwikkelen	-	Ontwikkelen/ Ondersteuning
Applicatie	Ondersteund via	Ondersteuning	Ontwikkelen	
Bestandsformaten	Ondersteund via	Ondersteund via	Ondersteuning	

Tabel 4 - Ontwikkeling en ondersteuning per emulatiesoort

3.2.2. Emulaties en emulatieacties

De daadwerkelijke emulatie vindt bij alle soorten emulatie plaats tijdens het opvragen van digitale informatie. Hoewel het wellicht niet voor de hand ligt, kan emulatie zowel in het archief plaatsvinden (server-side) als bij de gebruiker (client-side). In het laatste geval levert het systeem een digitaal bestand aan de gebruiker, die lokaal een viewer heeft die deze interpreteert. De noodzakelijke acties die moeten worden ondernomen om emulatie te gebruiken en de momenten waarop deze moeten worden ondernomen verschillen van elkaar:

- **Hardware-emulatie:** De eigenschappen en specificaties van de originele hardware moet uitgekristalliseerd zijn. Daarnaast moet er kennis beschikbaar zijn over welke besturingssystemen en applicaties op de geëmuleerde hardware kunnen draaien en hoe dat in zijn werk gaat. Hiervoor moet alles tijdig (wanneer alles nog in gebruik is) worden gedocumenteerd en beschikbaar zijn. Er moet vervolgens een emulator worden geprogrammeerd of een bestaande emulator moet worden getest;
- **Besturingssysteem-emulatie:** De eigenschappen en specificaties van het originele besturingssysteem moeten uitgekristalliseerd zijn. Daarnaast moet er vergelijkbare hardware als dat er origineel gebruikt werd worden gebruikt om de besturingssysteem-emulator op te kunnen draaien. Hiervoor moet het besturingssysteem en de onderliggende hardware tijdig (wanneer het besturingssysteem nog gebruikelijk wordt) worden gedocumenteerd en beschikbaar zijn. Er moet vervolgens een emulator worden geprogrammeerd of een bestaande emulator moet worden

getest. Er moet bij de aanschaf van nieuwe hardware in dezelfde lijn als waarvoor de emulator is geprogrammeerd hardware worden aangeschaft. Deze bewaarstrategie raakt aan het aanhouden van een computermuseum en wordt daarmee verder buiten dit onderzoek gehouden;

- Toepassingsprogrammatuur-emulatie: De eigenschappen en specificaties van de originele toepassing moeten uitgekristalliseerd zijn. Hiervoor moeten de toepassingen tijdig (wanneer de toepassingen nog gebruikelijk zijn) worden gedocumenteerd en beschikbaar zijn. Er moet vervolgens een emulator worden geprogrammeerd of een bestaande emulator moet worden getest;
- UVC: De eigenschappen en specificaties van de originele bestandsformaten moet uitgekristalliseerd zijn, evenals dat van de applicatie die met het originele bestand overweg kan. Hiervoor moeten applicatie en bestand tijdig (wanneer de applicaties en bestanden nog maar net in gebruik zijn) worden gedocumenteerd en beschikbaar zijn. Een UVC moet worden geschreven voor ieder hardware platform in de tijd waar het betreffende hardware platform nog maar net beschikbaar is. Dit verzekert de werking van de UVC op de nieuwe hardware. Daarnaast moeten UVC-dataformaatdecoder applicaties worden geprogrammeerd voor ieder bestandsformaat. Deze applicaties moeten kunnen draaien in de UVC en moeten worden gemaakt zodra de bestandsformaten in gebruik raken. Er moet per hardware platform viewers worden geprogrammeerd per informatietype die de LDB's kan interpreteren.

3.2.3. Gevolgen van emulatie

Het nut van emulatie wordt toegewezen aan de situatie van het behouden van complexe samengestelde bestanden of applicaties. Hierbij staat vaak de authenticiteit, gedrag en functionele eigenschappen centraal, dus de oorspronkelijke 'look and feel' [Test05]. Dit komt doordat de bestandsformaten en de omgeving rondom de informatie iedere keer hetzelfde blijven.

Het bewaren van de informatie in zijn authentieke vorm en omgeving kan ook als nadeel worden gezien. Na verloop van tijd zal namelijk de kennis van de geëmuleerde omgeving afnemen (bv. het gebruik van de applicatie) en zal de raadpleger geschoold moeten worden in het gebruik van de oude software. Als laatste zal het noodzakelijk zijn om met enkele regelmaat nieuwe emulatoren te maken of de originele emulatoren te emuleren.

Ondanks deze bezwaren is emulatie een goede bewaarstrategie voor organisaties die complexe informatietypen willen bewaren in hun originele 'look and feel', zoals databases, samengestelde documenten, applicaties en websites.

3.2.4. Aantallen emulatiepaden en experimenten

Net als bij de migratiebewaarstrategie kan er bij emulatie gesproken worden over paden om te komen tot representatie: emulatiepaden. Bij N aantal bestandsformaten die geëmuleerd moeten worden, is minimaal één emulator van de hardware (en apparaten) nodig, met ondersteuning voor wellicht meerdere besturingsystemen. Hierop moeten vervolgens meerder applicaties kunnen draaien, die de bestanden decoderen en weergeven. Om N bestandsformaten kunnen representeren, zijn R applicaties nodig, waarbij $R \leq N$. Daarnaast zijn S besturingsystemen noodzakelijk, waarbij $S \leq R$. Deze besturingsystemen zijn uit te voeren op verschillende soorten hardware (aantal H), waarbij $S \leq H$, met echter een beperkt aantal architecturen (A), waarbij $A \leq H$.

Wanneer emulatie wordt vergeleken met migratie, waarbij gekeken wordt naar het aantal paden dat

noodzakelijk is voor de representatie (representatiepaden), zal het aantal minimal emulatiepaden lager liggen dan dat noodzakelijk is voor de migratiebewaarstrategie. Dit omdat veel applicaties veel meerdere bestandsformaten kunnen interpreteren. Zo kan de videoweergave applicatie VideoLAN [22] meer dan 25 bestandsformaten van het informatietype ‘visueel’ weergeven. Emulatie van deze applicatie zou dan vanuit één emulatiepad voorzien in de weergave waar anders meer dan 25 migratiepaden voor nodig zouden zijn geweest. Per applicatie en besturingssysteem zal echter de gebruikerservaring en de geboden functionaliteit verschillen, wat resulteert in een hoger aantal emulatiepaden die de mogelijkheid bieden de digitale informatie te representeren. Dit betekent dat het maximaal te overwegen aantal emulatie experimenten per bestandsformaat bepaald wordt door het aantal applicaties dat overweg kan met het bestandsformaat, de voor deze applicaties noodzakelijke besturingssystemen en de hardware. Daarnaast kunnen de experimenten hergebruikt worden om als emulatieexperiment te dienen voor de overige bestandsformaten die de applicatie ondersteunt.

3.2.5. Projecten

Wijngaarden ea. [Wijn04] stellen dat emulatie nog niet eerder is geoperationaliseerd in de context van digitale duurzaamheid, anders dan de ervaringen die de Koninklijke Bibliotheek (KB) heeft opgedaan met de emulatiebewaarstrategie (UVC) vanuit haar e-Depot. Dit e-Depot is gebaseerd op een product van IBM, genaamd DIAS (Digital Information and Archiving System) [Test03-2]. Zij stellen dat, hoewel het wellicht moeilijk klinkt, emuleren van hardware eigenlijk eenvoudiger moet zijn dan het emuleren van de oorspronkelijke toepassingsprogramma's die de digitale formaten weergaven (dat wil zeggen opnieuw schrijven van deze toepassingsprogramma's). Voor het ene hoeft er slechts computerhardware op logisch niveau te worden geëmuleerd, en niet op het niveau van de schakelingen of de implementatie. De interne details van digitale hardware mogen dan complex zijn, het externe logische gedrag is relatief eenvoudig. Dit is de reden waarom hardware kan worden gebruikt om software te draaien. Op dit logische niveau is de meeste hardware in feite veel eenvoudiger en beter gespecificeerd dan software. Vanwege deze reden is het relatief recht-toe-recht-aan om programma's te schrijven die computerhardware emuleren en dergelijke programma's te valideren met het logische gedrag van de betreffende hardware.

Wat zij echter ook concluderen is dat voordat er ook maar één digitaal informatieobject kan worden geëmuleerd, er veel moet worden ontwikkeld. Zodra dit is gedaan, zijn andere digitale informatieobjecten die op de hardware te draaien zijn, met veel minder moeite te emuleren. Bij migratie ligt de investering per digitaal informatieobject veel lager, maar blijft deze per digitaal informatietype constant.

Rondom emulatie participeren de Koninklijke Bibliotheek en het Nationaal Archief in het al eerder beschreven PLANETS project. Onder dit overkoepelend Europees project werkt men samen aan Dioscuri, een modulaire emulator [23]. De PLANETS architectuur moet naast de migratiebewaarstrategie ook de emulatiebewaarstrategie als mogelijke service bieden.

3.2.6. Services

De services rondom de emulatiebewaarstrategie zijn nog niet zoals bij CRiB (zie paragraaf 3.1.6) duidelijk in een architectuur geplaatst. De ontwikkelingen van emulatoren blijven nog rond het uitvoeren van experimenten in testomgevingen [Hoev06]. Wel is er binnen het PLANETS project ruimte aangegeven voor emulatie experimenten. De invulling daarvan wordt in latere stadia van het

project vorm gegeven [Beck07]. Deze worden op soortgelijke manier behandeld als de migratie experimenten. Op deze manier kunnen dus ook emulatie experimenten worden gescoord op de criteria en op gelijk niveau worden vergeleken met de migratie experimenten.

Functioneel gezien zou de architectuur van CRiB onder vervanging van enkele services kunnen voldoen. Dit kan resulteren in de volgende services:

- Bestandsformaat herkenningsservice. Deze service moet de digitale bestanden herkennen (informatietype, software versie), maar ook de inhoud van de digitale bestanden bepalen en controleren;
- Aanbevelingsservice. Een service die tot een onderbouwde aanbeveling komt voor de te kiezen emulator en emulator element invulling (Welke applicatie, besturingssysteem, hardware?);
- Emulatie-uitvoerservice. De service die daadwerkelijke de emulatie uitvoert op het systeem van de klant of op afstand over een interactieve verbinding;
- Emulatie-evaluatieservice. Een service die het informatieverlies bepaalt dat gemoeid is met de optimale emulatie;
- Informatieservice. Een service die informatie geeft over welke bestandsformaten en software het risico lopen te verouderen;

3.2.7. Toepassing

Om de beschreven services te realiseren worden de verschillen ten opzichte van de migratiebewaarstrategie architectuur beschreven (paragraaf 3.1.7).

Bestandsformaat- en inhoudbepaling (Format Detector)

Binnen de migratiebewaarstrategie wordt de herkenning van bestandsformaten en het bepalen van de interne onderdelen van de digitale bestanden voor en na de migratieacties gebruikt. De bij dit onderdeel besproken tools kunnen gebruikt worden om te valideren of datgene wat ontstaan is na de migratie nog wel gelijk is aan het origineel (validatie). Daarnaast kunnen ze worden gebruikt te bepalen in welke mate er toch verandering of verlies is opgetreden.

Bij de emulatiebewaarstrategie zullen deze tools deze rol minder tot niet kunnen vervullen. In principe worden de digitale bestanden namelijk niet gewijzigd. Zij dienen nog wel herkend te worden bij inname en er moet wel bepaald worden waar de digitale bestanden uit bestaan, maar het zal niet het geëmuleerde resultaat kunnen bepalen. Er bevindt zich namelijk een laag tussen het huidige besturingssysteem en het originele besturingssysteem/de applicatie: de emulator. Deze laag bestaat uit een (interactieve) applicatie en er is nog geen software om de daarin gecodeerde informatie geautomatiseerd te bepalen, noch te vergelijken met de in de originele bestandsformaten gecodeerde informatie.

Service registers (Service Registry)

Dit onderdeel dat verantwoordelijk is voor het beschikbaar hebben van informatie over bestandsformaten, softwareleveranciers, migratietools en andere metadata. Om de emulatiebewaarstrategie toe te passen zal nog meer moeten worden opgeslagen. Zo moet er ook letterlijk software of softwarecomponenten worden opgeslagen die gebruikt kunnen worden om (onderdelen van) emulatoren te ontwikkelen. Daarnaast moeten ook de (onderdelen van) emulatoren zelf een plek krijgen. De al eerder besproken vier service registers moeten ook voor de

emulatiebewaarstrategie worden overwogen. Het GDFR register [Abra05][20] in dezen beoogt ook (onderdelen van) software die noodzakelijk zijn voor de representatie van de digitale informatie te gaan opnemen. Daarnaast heeft de Representation Information Registry/Repository [Rusb05][Giar05][21] de insteek om alle representatie informatie die noodzakelijke is om de digitale informatie te kunnen representeren in registers op te nemen. Dit zal ook de informatie omvatten die noodzakelijk is voor de emulatiebewaarstrategie.

Emulatie-uitvoering

Dit deel van de architectuur is verantwoordelijk voor het uitvoeren van de emulatie. De emulator onderdelen of emulatoren kunnen door derden worden geleverd of intern worden ontwikkeld. Daarnaast zouden deze emulatoren kunnen worden uitgevoerd op de PC van de klant of vanaf een centrale server. In het kader van de privacy gevoeligheid van de informatie in het CDD, zal het absoluut de voorkeur hebben om de emulaties intern uit te voeren.

De verschillende emulatietools en emulatietools aanbieders kunnen op vergelijkbare manier worden vergeleken als genoemd bij de migratiearchitectuur. Dit kan door zogenaamde procescriteria op te stellen en deze mee te wegen in de methode bewaarstrategie evaluatie methode van Strodl ea. [Stro07].

Software evaluatie

Dit onderdeel van de architectuur moet voorzien in feiten over de huidige status van software. Deze informatie stelt de emulatie adviesverstrekker in staat te beoordelen welke software mogelijk beter is om te ondersteunen met emulatie.

Het prototype van Ferreira ea. gebruikt een veelvoud van criteria waarop bestandsformaten kunnen worden gescoord. Dit zelfde zou kunnen worden gedaan voor software.

Emulatieresultaat evaluatie

De emulatieresultaten moeten geëvalueerd worden. Dit wordt gedaan te definiëren wat de digitaal gecodeerde informatie tot een unieke entiteit maakt. Het emulatieresultaat moet hieraan voldoen. Voorbeelden zijn het aantal karakters waar uit een document bestaat, de volgorde van deze karakters, het aantal pagina's en de grafische weergave. Bij de migratiebewaarstrategie worden tools gebruikt om bestandsformaten en de daarin opgeslagen bitstream te bepalen. Binnen de emulatiebewaarstrategie is er echter nog geen software om emulatieresultaten (ná de emulatie) op een soortgelijke manier te bepalen.

Emulatieadviseur

De emulatieadviseur is verantwoordelijk voor het bepalen van het beste emulatiealternatief. Het gebruikt hiervoor de bewaareisen die de klant heeft gesteld (criteria en het belang van deze via weegfactoren) en de resultaten van de geteste emulatiebewaarstrategieën. Deze resultaten worden verworven uit de verschillende sets van criteria die samen de criteriaboorn vormen en door het ondernemen van experimenten (zie ook onderstaande onderdeel "evaluatie register"). De emulatieadviseur werkt verder op dezelfde manier als bij de migratiebewaarstrategie.

Evaluatieregister (Evaluations Repository)

Om verantwoord tot het beste emulatiealternatief te komen, gebruikt de emulatieadviseur het evaluatieregister. Dit register bevat alle resultaten van experimenten die uitgevoerd zijn door de onderdelen die verantwoordelijk zijn voor de emulatieresultaat evaluatie, software evaluatie en de

emulatie uitvoering. De inhoud van het evaluatieregister is erg belangrijk en wordt verkregen door training. Dit kan worden gezien als het zelf opzetten van enkele emulatie bewaarscenario's met te verwachten resultaten en deze uit te voeren met het te trainen systeem. Dit trainingsonderdeel is voor emulatie nog niet vormgegeven. Tests worden momenteel nog handmatig uitgevoerd en geëvalueerd [Hoev06].

3.2.8. Samengevat

Emulatie is nog de minst toegepaste bewaarstrategie. Het wordt gezien als een complexe en dure bezigheid, maar mogelijk wel de bewaarstrategie die het beste de 'look and feel' bewaard. Het onderdeel om op een geautomatiseerde manier evaluaties van emulatiebewaarstrategieën uit te kunnen voeren mist nog en zal moeten worden ontwikkeld. Verder kunnen handmatig geteste emulatiebewaarstrategieën wel worden vergeleken op dezelfde manier als dat dit gedaan kan worden bij migratiebewaarstrategieën. Een emulatiebewaarstrategie kan dus worden verkozen als de bewaarstrategie die het beste past bij de te bewaren digitale informatie. Ontwikkelingen rondom de Virtuele Universele Computer en modulaire emulator dragen bij aan een universeler toe te passen emulatieconcept.

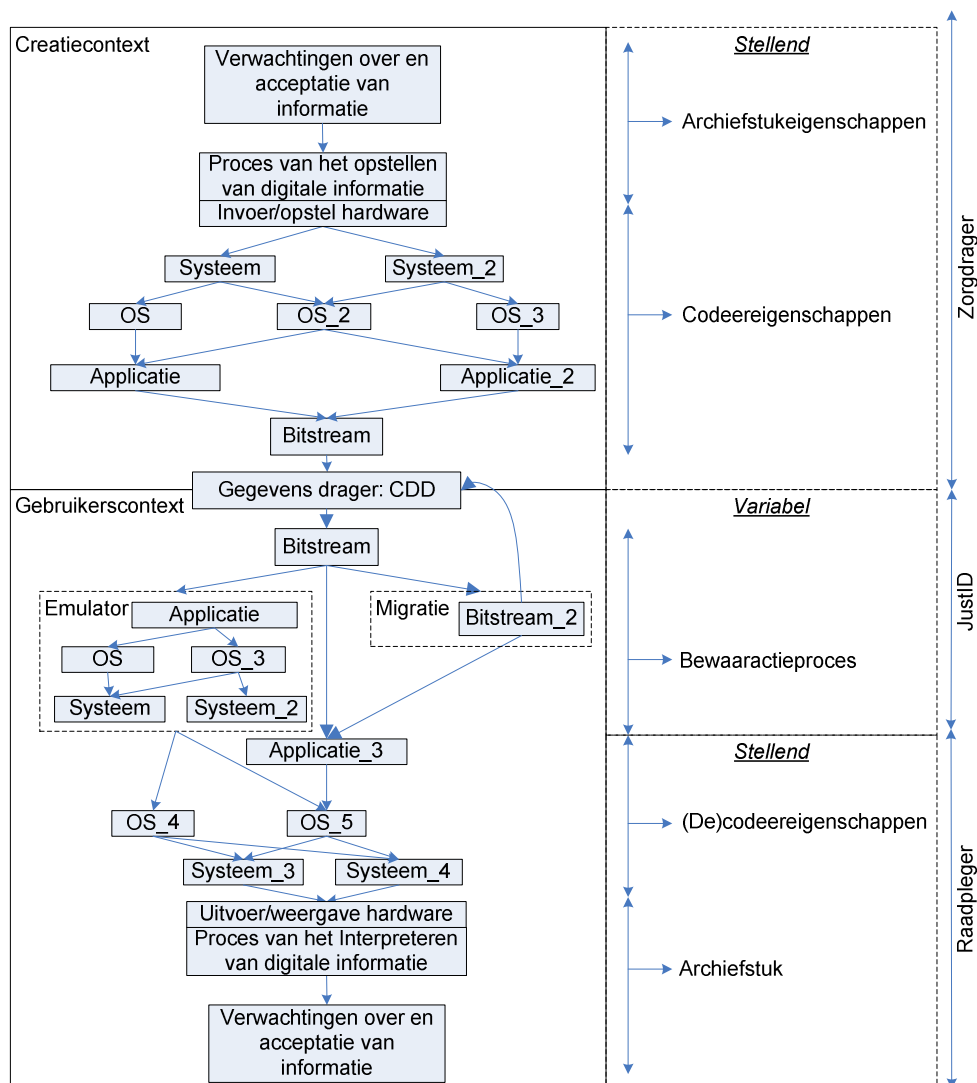
3.3. Conclusie

In dit hoofdstuk is beschreven welke overkoepelende bewaarstrategieën er zijn, wat de voor- en nadelen van beiden zijn en hoe deze in de praktijk worden gebracht. Gekenmerkt wordt dat de migratiebewaarstrategie vooral gebruikt wordt om informatie te bewaren die niet langer bewaard hoeft te worden dan 10 jaar, daarna moet de voorkeur gegeven worden aan de emulatiebewaarstrategie. De migratiebewaarstrategie is goedkoop, eenvoudig en wordt door veel instanties gebruikt. Het is daarmee toegankelijker dan de emulatiebewaarstrategie. Het omvat veel werk om een emulator te creëren die de juiste omgeving biedt voor de representatie van opgeslagen digitale informatie. Daarnaast wordt het heel weinig gebruikt in relatie tot digitale duurzaamheid.

De verschillende bewaarstrategieën omvatten een representatieproces. Deze processen verschillen in uitvoering en in resultaat. Deze eigenschappen kunnen vergeleken worden, er zullen eisen aan gesteld kunnen worden vanuit de ketenpartners die digitale informatie willen bewaren (bewaareisen). Daarnaast zullen raadplegende ketenpartners eisen stellen vanuit de hard- en software die bij hen aanwezig zijn. Er kan dan ook een bepaalde voorkeur zijn voor een bewaarstrategie. In het volgende hoofdstuk worden daarom op hoofdlijnen criteria opgesteld waaraan de bewaring en hun representatie aan moeten voldoen.

Hoofdstuk 4 Criteria

In voorgaand hoofdstuk is toegelicht welke bewaaracties er te onderscheiden zijn en hoe deze in praktijk worden gebracht. Daarnaast is geconcludeerd dat bewaarstrategieën onderling verschillen in eventueel uit te voeren bewaaracties en in het resultaat dat dit proces levert. Ook het bewaaractieproces en zijn resultaat is te plaatsen in het lagen model dat in hoofdstuk twee was opgesteld (figuur 7), net als de onderdelen van waaruit bepaalde eisen aan de bewaring en representatie worden gesteld. In dit hoofdstuk worden op hoofdlijnen de criteria uitgewerkt aan de hand van de in het model weergegeven onderdelen (figuur 21). Met ‘eisen’ wordt in dit onderzoek eigenlijk meer dan alleen de eisen bedoeld, zo geven Roozenburg ea. [Rooz92] aan dat ‘eisen’ ook wel de ‘absolute criteria’ zijn, criteria waaraan voldaan moet worden. Bij de selectie van een geschikte bewaarstrategie kunnen ook wensen worden meegenomen. Aan wensen hoeft niet of slechts ten dele te worden voldaan. Daarnaast zijn er normen of standaarden te verwerken tot eisen of wensen. Normen kunnen door wet- of regelgeving van kracht worden voorzien.



Figuur 21 – Bewaaronderdelen

4.1. Archiefstukeigenschappen

Dit gedeelte van de bewaareigenschappen gaat in op de context waarin informatie wordt opgesteld. Het gaat hierbij over de beschrijvende metadata (zie ook paragraaf 2.3.1), dus de metadata die de digitale informatie betekenis geeft en identificeert, zoals gegevens over wie de informatie gecreëerd heeft, over wie of wat de informatie gaat, de organisatie, de functie, het mandaat, het bedrijfsproces en de op het archiefstuk van toepassing zijnde wettelijke kaders. Het gaat om het intellectuele beheer van de informatie. Het is de technische basis van het archiefstuk en hoe deze te relateren is aan andere informatie en metadata. Daarnaast geven Strodl ea. aan dat ook de eigenschappen van het opslagmedia onder de archiefstukeigenschappen vallen [Stro07-2].

De archiefstukeigenschappen worden door de zorgdrager opgeslagen in metadata en vervolgens samen met alle digitale bestanden overgedragen naar het CDD. Daar worden alle metadata in het datamodel geplaatst en de digitale bestanden opgeslagen in storage. Eventueel worden bewaaracties uitgevoerd.

Welke archiefstukeigenschappen (metadata) moeten worden opgeslagen ligt buiten de scope van dit onderzoek, zoals ook beschreven staat in de afbakening, paragraaf 2.5. Het is wel belangrijk te realiseren dat de hoeveelheid en correctheid van opgeslagen archiefstukeigenschappen leidend zijn voor het kunnen vinden van archiefstukken door de raadpleger. Daarnaast zal zij de inhoud zo goed mogelijk aan de creatiecontext moeten kunnen relateren.

4.2. (De)codeereigenschappen

Codeereigenschappen komen zowel in de creatie- als de gebruikerscontext terug. Vanuit de creatiecontext zijn er codeereigenschappen. Hieraan kunnen eisen worden gesteld, de zogenaamde bewaareisen. Vanuit de gebruikerscontext leiden op hun beurt de decodeereigenschappen tot representatie-eisen.

Bewaareisen

Dit gedeelte van de codeereigenschappen benoemt de visuele en contextuele ervaringen van de opsteller van het archiefstuk. Belangrijke delen van deze visuele en contextuele ervaringen moet de raadpleger ook ervaren bij het raadplegen van het archiefstuk. De opsteller van de digitale informatie moet dus eisen stellen aan wat er van belang is om later in ieder geval weer te geven aan de raadpleger: de bewaareisen.

Als subcategorieën van de bewaareisen zijn de eigenschappen van informatie te gebruiken, dus context, inhoud, structuur, uiterlijk en gedrag (paragraaf 2.2.2). Hierbij moet worden opgemerkt dat de eigenschap context ook terugkomt bij de eerder genoemde archiefstukeigenschappen. De eigenschap context binnen de codeereigenschappen gaat echter alleen over de context die binnen het bestand of de software liggen, zoals URL's of andere referenties binnen het digitale bestand naar externe bronnen. Hieronder worden enkele subcategorieën tot meetbare criteria voor een deel uitgewerkt vanuit het onderzoek van Racuh ea. [Rauc05]. De opsteller zal moeten aangeven welke bewaareisen zij stellen en hoe belangrijk het hieraan voldoen is.

Niveau 1	Niveau 2	Niveau 3	Meetwaarde
Uiterlijk	Letters/tekens	Style (lettertype)	Categorie
		Grootte	Punten
		Kleur	Categorie
	Pagina	Nummering	Wel/niet
		Marges	Millimeter
Structuur	Pagina	Eindemarkeringen	Wel/niet
		Koptekst	Wel/niet
		Voettekst	Wel/niet
		Paginanummering	Wel/niet
		Index	Wel/niet
Inhoud	Letters/tekens	Codering	Categorie
	Afbeelding	Breedte	Punten
		Hoogte	Punten
		Resolutie	DPI
		Codering	Categorie
		Bitdiepte	Bits

Tabel 5 – Criteria aan de bewaring

Representatie-eisen

Naast dat ketenpartners bewaareisen stellen aan de te bewaren informatie, stellen zij ook eisen aan het representatieproces: de representatie-eisen. De raadpleger zal namelijk eisen dat de in het CDD opgeslagen visuele en contextuele ervaringen weer te geven (te representeren) zijn op de bij de ketenpartner in gebruik zijnde combinatie van hard- en software. Dit houdt dus in dat er via migratie of emulatie een representatie geboden moet worden die aansluit op de hard- en software van een ketenpartner. Hiervoor moet goed in beeld worden gebracht welke representaties gedaan kunnen worden binnen de technische omgeving van de ketenpartners. Momenteel wordt bijvoorbeeld Adobe PDF gebruikt om documenten in op te slaan en weer te geven bij ketenpartners. In het geval dat een ketenpartner geen Adobe PDF heeft, kan hij een archiefstuk niet weergeven. Dit kan een probleem worden wanneer er naast Adobe PDF nog meer software noodzakelijk wordt om archiefstukken volledig te kunnen representeren. De ketenpartner kan allerlei redenen hebben om een bepaalde combinatie van hard- en software te gebruiken die beperkend is voor de mogelijke representatie.

De subcategorieën van representatie-eisen zijn minimaal dezelfde als die van de bewaareisen, dus een indeling naar de eigenschappen van informatie. Hierin bieden de bij de ketenpartner te representeren bestandsformaten of emulatie methoden een afbakening van weer te geven informatie (beschikbare viewpaths). Per applicatie bij de ketenpartner moet dus worden bekeken met welke bestandsformaten

deze overweg kan en of deze mogelijke manieren van representeren wel voldoen aan de gestelde bewaareisen. Idealiter moet een representatie voldoen aan alle bewaareisen. JustID moet aangeven aan welke bewaareisen niet voldaan wordt en wat in dat geval het 'informatieverlies' is. JustID moet dan een adviserende rol vervullen en een te gebruiken combinatie van hard- en software (viewpath) aanraden, los van of deze combinatie al dan niet aanwezig is bij de ketenpartner. Voor het overwegen van welk viewpath het beste is zijn onder andere de volgende criteria voor bestandsformaten en applicatie op te stellen [Rauch05][Stro07]:

Niveau 1	Niveau 2	Meetwaarde
Vertegenwoordiging/marktaandeel		Alom, wijd, beperkt, verouderd
Ondersteuning weergave		0/1-5/6-10/10+ applicaties
Documentatie	Kwaliteit	Primair/secundair
	Beschrijvendheid	Volledig/deels/niet
	Openheid	Standaard/open/gesloten
	Beschikbaarheid	Publiek/beperkt/niet
Stabiliteit	Veranderlijkheid	< 1/1-2/3-5/> 5 jaar
	Achterwaarts comp.	Niet/vorige versie/enkele vorige/allen
Identificatie		Automatisch/manueel/niet
Validatie		Automatisch/manueel/niet
Informatievastheid		Los/vast
Intellectueel eigendom		Ja/nee
Samengesteldheid		Hoog/middel/laag
Fouttolerantie		Geen/detectie/reparatie
Grootte		Groter/zelfde/kleiner
Minimale kosten applicatie		Geen/minimaal/normaal/duur/maatwerk

Tabel 6 - Criteria aan bestandsformaten en software (onderdeel van de representatie-criteria)

4.3. Bewaaractieproces

JustID dient door middel van een door haar uit te voeren proces een representatie te leveren aan de raadplegende ketenpartner. Dit omvat het versturen aan de raadpleger van originele of gemigreerde digitale bestanden of het leggen van een interactieve verbinding tussen raadpleger en een bij JustID draaiende emulator. Deze worden in figuur 21 weergegeven door drie pijlen van de bitstream naar een proces binnen JustID en applicatie bij de raadpleger. JustID legt dus de brug door de tijd tussen de opgeslagen digitale informatie in een bitstream (verleden) en een veelvoud van haar raadplegers (toekomst).

Hiervoor zijn de volgende criteria opgesteld [Stro07]:

Niveau 1	Niveau 2	Meetwaarde
Automatisering		Geconfigureerd/API/scripts/niet
Prestaties	Te streamen	Ja/nee
	Capaciteit	Mb/seconde
	Schaalbaarheid	Aantal ondersteunde bestandsformaten
Documentatie	Kwaliteit	Primair/secundair
	Beschrijvendheid	Volledig/deels/niet
	Openheid	Standaard/open/gesloten
	Beschikbaarheid	Publiek/beperkt/niet
Toepasbaarheid		Goed/acceptabel/slecht
Validatie	Syntactisch	Volledig/deels/geen
	Semantisch	Volledig/deels/geen

Tabel 7 - Criteria aan het bewaaractieproces

4.4. Kosten

Met het in kaart brengen van bewaar- en representatie-eisen en het uitvoeren en testen van bewaaracties zijn kosten gemoeid. Kosten worden niet weergegeven in het lagenmodel als een overkoepelende eigenschap van digitale bewaring, omdat zij namelijk toe te rekenen zijn aan de specifieke onderdelen. Rauch ea. en Strodl ea. kiezen echter ervoor om deze op gelijke hoogte te plaatsen als de bewaar- en representatie-eisen en de bewaaracties [Rauch05][Stro07-2]. Hiervoor voeren zij als redenen aan dat kosten vaak van directe invloed zijn op bewaring en dat deze dreigen onder te sneeuwen wanneer deze op een lager niveau geplaatst worden. Daarnaast zijn lage kosten vaak beter dan hoge en is deze manier van redeneren in tegenspraak met merendeel van de andere criteria, waarbij hoge waarden de voorkeur genieten. Men verdeelt kosten onder naar kosten voor techniek (hard- en software) en personeel, zoals in onderstaande tabel wordt weergegeven.

Niveau 1	Niveau 2	Meetwaarde
Hardware	Aanschaf	Euro/jaar
	Operationeel	Euro/jaar
Software	Aanpassingen/maatwerk	Euro/jaar
	Aanschaf	Euro/jaar
	Operationeel	Euro/jaar
Personeel	Training	Beschikbaarheid Euro/jaar
		Kosten
	Onderzoek	Euro/jaar
	Beheer	Euro/jaar

Tabel 8 - Criteria aan kosten

4.5. Conclusie

In dit hoofdstuk zijn op hoofdlijnen de criteria vanuit de creatie- en gebruikerscontext opgesteld waarop een bewaarstrategie geselecteerd kan worden en beantwoordt daarmee de tweede onderzoeksvraag. Deze zijn onder te verdelen naar de criteria gesteld aan de bewaring, representatie en het tussenliggende proces. Daarnaast zijn er kosten verbonden aan het duurzaam beschikbaar houden van digitale informatie.

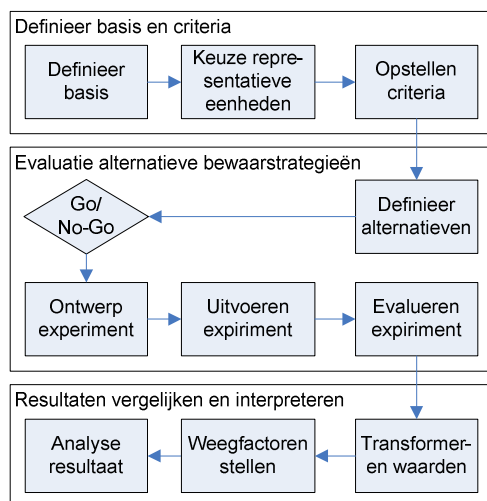
Het proces dat tussen de bewaring en representatie ligt wordt uitgevoerd door JustID, de eventueel noodzakelijke bewaaracties: migratie of emulatie. Omdat de meeste geschikte bewaarstrategie afhankelijk is van de informatietypen (verschillende coderingen) waar de digitale informatie uit bestaat en de eisen die de bewarende en de raadplegende ketenpartners stellen, moeten beide bewaarstrategieën uitgevoerd kunnen worden door JustID.

Met de criteria op hoofdlijnen uitgewerkt, moet er een methode worden uitgewerkt die per creatiecontext en de verschillende gebruikerscontexten uitgevoerd wordt om de exacte criteria op te stellen. Vervolgens moeten de criteria gemeten worden en bewaarstrategieën gescoord. In het volgende hoofdstuk wordt een methode uitgewerkt die voorziet in deze onderdelen.

Hoofdstuk 5 Selectie van geschikte bewaarstrategie

In hoofdstuk 3 is uitgewerkt welke bewaarstrategieën er zijn, wat de eigenschappen hiervan zijn en hoe deze in de praktijk worden toegepast. In het voorgaande hoofdstuk zijn op hoofdlijnen de criteria opgesteld waarop een bewaarstrategie geselecteerd kan worden. Met de criteria en bewaarstrategieën beschreven, is het noodzaak een selectiemethode uit te werken. Dit hoofdstuk beschrijft een bewaarstrategie selectiemethode die ontwikkeld is door Rauch ea. [Rauc05] en Strodl [Stro07-1/-2] ea. De ontwikkelingen rondom deze methode worden momenteel voortgezet binnen het PLANETS project van de Europese Unie, waarbinnen een consortia van onderzoekers en het bedrijfsleven rondom digitale duurzaamheid samenwerken. PLANETS staat voor Preservation and Long-term Access through Networked Services. De kern van het onderzoek binnen PLANETS is het ontwikkelen van praktisch toepasbare diensten en tools die helpen bij het duurzaam houden van digitale informatie en deze in te bedden in een gedistribueerde architectuur [12]. Een van de onderwerpen onder het gebundelde onderzoek vanuit PLANETS is “preservation planning”. Dit onderdeel gaat in op hoe uit verschillende bewaarstrategieën een keuze te maken.

De bewaarstrategie selectiemethode beschrijft hoe tot een keuze voor een bewaarstrategie te komen door criteria te definiëren en via bewaarexperimenten en een aangepaste utiliteitsanalyse [Rauc04] te berekenen wat de meest geschikte bewaarstrategie voor de digitale informatie is. Het doorloopt daarvoor de in figuur 22 weergegeven stappen. Deze worden hieronder in drie delen beschreven, waarbij het eerst deel ingaat op het definiëren van de basis en criteria vanuit bewaar- en representatie-eisen. Daarna worden alternatieve bewaarstrategieën opgesteld, in experimenten getest en vervolgens geëvalueerd. In het laatste deel van de methode worden de resultaten vergelijkbaar gemaakt en geïnterpreteerd.

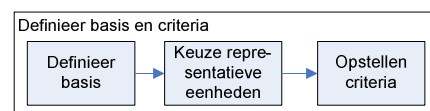


Figuur 22 - Bewaarstrategie evaluatie methode [Stro07-2]

5.1. Definiëren van de basis en de criteria

Het eerste deel van de bewaarstrategie-evaluatiemethode gaat in op het definiëren van de basis, het verkrijgen van representatieve eenheden (enkele te archiveren stukken in digitale vorm) en de criteria waaraan bewaring en representatie moeten voldoen. Onderstaande stappen gaan hier dieper op in.

Basis definiëren



In de eerste stap van de bewaarstrategie-evaluatiemethode wordt de basis vastgelegd waarin de overwegingen rondom duurzame bewaring worden genomen. Dit omvat de reikwijdte, het doel en andere metadata. Zo wordt er informatie op een semi-gestructureerde manier opgeslagen over informatietypen en bestandsformaten, hun aantallen en mogelijke problemen rondom intellectueel eigendom.

Keuze representatieve eenheden

Nadat de basis is gedefinieerd worden enkele representatieve eenheden van de te bewaren digitale informatie geselecteerd (steekproef). Deze worden gebruikt tijdens bewaarstrategieën experimenten. Er kunnen bijvoorbeeld enkele te bewaren MS Word 2003 bestanden (document informatietype) worden geselecteerd die representatief zijn voor de informatie die te vinden is in de andere archiefstukbestanden bij de ketenpartner.

Criteria opstellen en verkrijgen

De derde stap in de bewaarstrategie-evaluatiemethode is die van het definiëren van de criteria. Deze criteria worden opgesteld vanuit de bewaar- en representatie-eisen van de ketenpartner en criteria die JustID stelt aan het bewaaractieproces en de kosten. Deze zijn onder te verdelen naar de in hoofdstuk 4 gedefinieerde criteria categorieën en worden in de bewaarstrategie-evaluatiemethode weergegeven in een criteriaboom, zoals is weergegeven in bijlage 8.

Aan de hand van enkele case studies geven Strodl ea. aan dat een criteriaboom ergens tussen de 50 en enkele honderden criteria bevat, verdeeld over ongeveer 4 tot 6 niveaus. De criteria worden opgesteld via een workshop waarbij enkele experts aanwezig zijn van verschillende gebieden. Zo moet men in het kader van de bewaring in het CDD een workshop organiseren per zorgdragende ketenpartner en te bewaren dossiertype om de bewaareisen te inventariseren. Bij de raadplegende ketenpartners moet worden geïnventariseerd welke representatie-eisen zij stellen. Bij de workshops moeten domeinexperts aanwezig zijn op het gebied van wet- en regelgeving, archivistiek en het primaire proces.

Aan alle bladeren (criteria) van de criteriaboom worden meetbare eenheden gekoppeld. Waarmogelijk moeten deze criteria objectief meetbaar zijn, zoals euro per jaar, frames per seconde, seconde per conversie. Voor enkele criteria zal een subjectieve schaal noodzakelijk zijn, zoals voor de mate van openheid van de bestanden waar de informatie in gecodeerd is of de mate van ondersteuning van het bestandsformaat. Enkele voorbeeldcriteria met bijbehorende meeteenheden voor de beoordeling van bewaareisen en bewaaractiesproces eisen aan audiobestanden is weergegeven in tabel 9 [Rauc05].

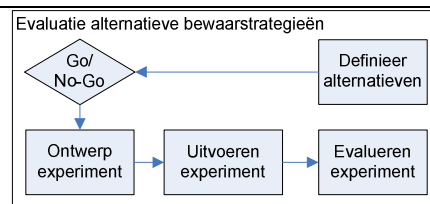
Criteria	Meeteenheid
Frame rate	Frames/seconde
Stereo	Ja/nee
Signaal representatie	Categorisering (bv. RGB,)
Personeel in dienst	€/jaar

Tabel 9 - Toekennen van meeteenheden per criteria [Rauc05]

De criteriaboom kan worden opgesteld door post-it te gebruiken, maar het opstellen zal aanzienlijk efficiënter gaan met behulp van mind-mapping software zoals deze geïntegreerd is in de PLANETS Preservation Planning Tool (PLATO) (zie bijlage 9).

5.2. Alternatieve bewaarstrategieën evalueren

Met de basis gedefinieerd, enkele representatieve informatie-eenheden geselecteerd en de criteria opgesteld, kunnen alternatieve bewaarstrategieën geëvalueerd worden. Daarvoor moeten er eerst alternatieve bewaarstrategieën worden gedefinieerd. Daarna moeten deze in een experiment getest worden om de meetwaarden van de eerder opgestelde criteria aan de representatie en het bewaaractieproces te verkrijgen. Onderstaande stappen gaan dieper in op dit proces.



Alternatieve bewaarstrategieën definiëren

Er worden in de literatuur rondom digitale duurzaamheid over enkele verschillende bewaarstrategieën gesproken die terug te voeren zijn op de twee overall bewaarstrategieën: migratie en emulatie (Hoofdstuk 3). Onder deze overkoepelende bewaarstrategieën zijn varianten te schalen die verschillend scoren op de eerder opgestelde criteria. Om alternatieve bewaarstrategieën te identificeren kan men zich richten op diensten van instanties die de technologisch trends in kaart brengen, zoals DigiCULT met haar Technology Watch Reports [13] en de Digital Preservation Coalition met een maandelijks blad [14].

Go / No-Go

Nadat er alternatieve bewaarstrategieën gedefinieerd zijn, moet er worden overwogen welke alternatieve daadwerkelijk getest moeten gaan worden. Dit zal per alternatief besloten moeten worden, rekening houdende met de beschikbare middelen en verwachtingen. Naast de 'Go' en 'No-Go' kan ook besloten worden om de overweging van het alternatief (voorlopig) uit te stellen. Bijvoorbeeld om voorliggende stappen opnieuw uit te voeren of om aanvullend extern onderzoek af te wachten.

Een voorbeeld van mogelijke resultaten uit de Go/No-Go stap zijn een tiental migratie-bewaarstrategieën voor audiobestanden, zoals de migratie naar Std. DV, Digi-Beta, PAL-VHS, SVHS, U-Matic, Beta Cam, MPEG, NTSC-VHS, DPS en Hi8 [Rauc05].

Ontwerp experiment

Om experimenten herhaaldelijk uit te kunnen voeren, moet er gebruik gemaakt worden van een gedocumenteerde omgeving. Dit omvat een beschrijving van de volgorde waarin het experiment wordt uitgevoerd, de gebruikte software/hardware en een beschrijving van tools die gebruikt worden voor het bepalen van de in- en uitgaande resultaten van tests. Alle onderdelen van de experimenten moeten worden ontwikkeld/aangeschaft, geïnstalleerd en getest.

Uitvoeren experiment

Tijdens het uitvoeren van een experiment worden een of meerdere aspecten van de toepassing van een bewaarstrategie op de steekproef getest. Daarnaast produceert het experiment resultaten in de vorm van gemigreerde bestanden, veranderde metadata of een geëmuleerde weergave van het origineel. Deze worden tijdens de volgende stap geëvalueerd.

Evalueren experiment

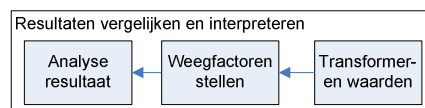
Tijdens het evalueren van het experiment worden de verkregen meetwaarden ingevuld in de criteriaboom en toegewezen aan de geteste bewaarstrategie. Enkele voorbeeldresultaten uit deze stap vanuit een experiment in het kader van de bewaring van audiobestanden [Rauc05] zijn weergegeven in tabel 10. Deze corresponderen met de toegekende meeteenheden per criteria in tabel 9.

Criteria \ Strategie	Std. DV	Digi-Beta	PAL-VHS	SVHS
Frame rate	30fps	30fps	30fps	30fps
Stereo	Ja	Ja	Ja	Ja
Signaal representatie	5	5	1	2
Personen in dienst	50	70	50	60

Tabel 10 – Voorbeeld van de meetwaarden van bewaarstrategieën op bewaarcriteria [Rauc05]

5.3. Resultaten vergelijken en interpreteren

Met de invulling van meetwaarden vanuit experimenten in de criteriaboorn zijn de resultaten nog niet algemeen te vergelijken ten opzichte van de meetwaarden op andere criteria. Onderstaande stappen gaan in op het vergelijkbaar krijgen van de meetwaarden, het stellen van weegfactoren aan de criteria en de uiteindelijke analyse en interpretatie van het resultaat. Dit laatste leidt tot de keuze voor de meest geschikte bewaarstrategie.



Transformeren van de meetwaarden

De meetwaarden in de criteriaboorn hebben allerlei verschillende waarden, bijvoorbeeld euro, seconde, aantal bits, mate van documentatie. Om de criteria onderling te kunnen vergelijken moeten de meetwaarden worden getransformeerd naar een gelijke schaal. Rauch ea. kiezen in de casestudy naar bewaarstrategieën voor audiobestanden [Rauc05] een vijf puntsschaal om de meetwaarden naar te transformeren. Hierbij kiezen zij een extra waarde voor niet acceptabele waarden. Tabel 11 ligt het gebruik van deze schaal toe enkele criteria. De schaal 0 of N.A. ("Niet Acceptabel") moet resulteren in dat de bewaarstrategie uiteindelijk afvalt of dat er expliciet kennis moet worden genomen van de afwijkende meetwaarde.

Criteria \ Schaal	5	4	3	2	1	0 / N.A.
Frame rate	≥ 30	≥ 20	≥ 17	≥ 15	≥ 10	< 10
Stereo	Ja	4	3	2	1	Nee
Signaal representatie	5	4	3	2	1	N.A.
Pers. in dienst	0	]0;20]	]20;40]	]40;80]	]80;150]	> 150

Tabel 11 – Voorbeeld van een transformatietabel voor meetwaarden per criteria [Rauc05]

Na de transformatie van de gemeten waarden wordt onderstaande tabel verkregen, waarin de getransformeerde waarden vermeld worden per geteste bewaarstrategie.

Criteria \ Strategie	Std. DV	Digi-Beta	PAL-VHS	SVHS
Frame rate	5	5	5	5
Stereo	5	5	4	4
Signaal representatie	5	5	1	2
Pers. in dienst	2	2	2	2

Tabel 12 – Voorbeeld van getransformeerde meetwaarden per criteria [Rauc05]

Weegfactoren toekennen

De criteria verschillen onderling in belangrijkheid en zijn afhankelijk van wat de ketenpartners aan bewaar- en representatie-eisen stellen. Om tot een goede verdeling van de belangrijkheid te komen worden er door ketenpartners weegfactoren toegekend aan de criteria. Onderstaande tabel geeft enkele voorbeeld niveaus en weegfactoren vanuit de casestudy naar bewaarstrategieën voor audiobestanden [Rauc05].

Niveau 1	%	Niveau 2	%	Niveau 3	%	Niveau 4	%	(Gewicht)
Bestands- eigenschappen	50	Uiterlijk	45	Functionaliteit	10	Stereo	20	0.0045
						Signaal repr.	35	0.007875
Kosten	25	Personeel	30	In dienst	50	0.0375		

Tabel 13 – Voorbeeld van het toekennen van weegfactoren (percentage en uiteindelijke fractie) [Rauc05]

De weegfactoren worden vervolgens vermenigvuldigd met de getransformeerde meetwaarden. Dit resulteert in een overzicht van getransformeerde gewogen waarden., zoals onderstaande weergave.

Criteria \ Strategie	Std. DV	Digi-Beta	PAL-VHS	SVHS
Frame rate	0.09	0.09	0.09	0.09
Stereo	0.0225	0.0225	0.018	0.018
Signaal representatie	0.039	0.039	0.007875	0.01575
Pers. in dienst	0.075	0.075	0.075	0.075

Tabel 14 – Voorbeeld van getransformeerde gewogen meetwaarden [Rauc05]

Analyseren van de resultaten

De getransformeerde gewogen meetwaarden kunnen nu per bewaarstrategie worden opgeteld en daarnaast kan ook het product worden berekend. Met de berekening van het product vallen bewaarstrategieën af; het product eindigt dan namelijk op 0 wanneer één van de criteria op 0 is gesteld, de waarde voor niet acceptabele meetwaarden op criteria.

Voordat vervolgens de bewaarstrategieën worden gerangschikt, kan een gevoeligheidsanalyse worden uitgevoerd om te kijken in welke mate kleine wijzigingen in de meetwaarden de uiteindelijke rangschikking beïnvloed. De uiteindelijke rangschikking geeft in volgorde aan wat de meest geschikte bewaarstrategie is voor de te bewaren digitale informatie. Onderstaande tabel geeft als voorbeeld de resultaten weer uit de casestudy naar bewaarstrategieën voor audiobestanden [Rauc05].

Rang	Strategie	Totaal score	
		Som	Product
1	Std. DV	3.38	3.16
2	Digi-Beta	3.37	3.15
3	SVHS	3.33	3.10
4	PAL-VHS	3.27	0

Tabel 15 – Voorbeeld van het rangschikken van bewaarstrategieën [Rauc05]

Strodl ea. [Stro07-2] stellen voor om naast het berekenen van de som en het product ook de Som van de Prioriteiten en Austin Slight te berekenen per bewaarstrategie.

5.4. Conclusie

In dit hoofdstuk is de bewaarstrategie-evaluatiemethode beschreven die gebruikt kan worden om een basis en criteria op te stellen, alternatieve bewaarstrategieën te definiëren, deze te testen binnen experimenten en uiteindelijk een uitspraak te doen over wat de meest geschikte bewaarstrategie is. Daarmee geeft dit hoofdstuk een antwoord op de derde onderzoeksvraag.

JustID moet de bewaarstrategie-evaluatiemethode gaan uitvoeren om samen met de ketenpartner(s) bewaar- en representatie-eisen op te stellen. Daarnaast moet zij zelf experimenten uitvoeren om te kunnen beoordelen welke bewaarstrategie de meest geschikte is. De volgende twee hoofdstukken gaan in op deze en vanuit voorgaande hoofdstukken geïdentificeerde processen die door de organisatie uitgevoerd moeten gaan worden en welke benodigde architectuur daarvoor nodig is.

Hoofdstuk 6 Organisatie-inrichting

In voorgaande hoofdstukken is beschreven welke bewaarstrategieën er zijn, hoe deze in de praktijk worden gebracht en hoe via de bewaarstrategie-evaluatiemethode de meest geschikte bewaarstrategie wordt bepaald. Geconcludeerd is dat JustID met haar organisatie en informatiesystemen zowel de migratie- als emulatiebewaarstrategie moet kunnen ondersteunen én dat zij de bewaarstrategie-evaluatiemethode moet uitvoeren in samenwerking met zowel zorgdragende- als raadplegende ketenpartners. Het realiseren van dit alles is de strategie voor het omgaan met digitale duurzaamheid. Mintzberg stelt dat strategie een patroon in een stroom van acties is, een consistente manier van handelen gedurende een bepaalde tijd [Mint78]. Datgene dat men daadwerkelijk doet is de gerealiseerde strategie en is onderdeel van de geplande/voorgenomen strategie en ontstane strategie. Dit laatste onderdeel van strategie is dus niet vooringenomen, maar wordt al doende en al lerende ingeslagen en omgezet naar consistent handelen. De in hoofdstuk 5 beschreven bewaarstrategie-evaluatiemethode voorziet in het constant herzien en aanpassen van de strategie om digitale informatie duurzaam toegankelijk te houden.

Om bovenstaande te realiseren moet een organisatiedeel worden ontworpen rondom het CDD die zich bezighoudt met de processen rondom digitale duurzaamheid. In dit hoofdstuk worden daarom deze processen uitgewerkt, waarbinnen taken uitgevoerd worden door functionarissen van verschillende afdelingen. De gehele processen rondom digitale duurzaamheid staan uitgewerkt in bijlage 10.

6.1. Services

Wanneer de services worden samengenomen die vanuit de praktijk zijn geïdentificeerd in het kader van het uitvoeren van de migratie- en emulatiebewaarstrategieën (hoofdstuk 3) en voor het komen tot een selectie daarvan (hoofdstuk 5), wordt onderstaande lijst van services verkregen.

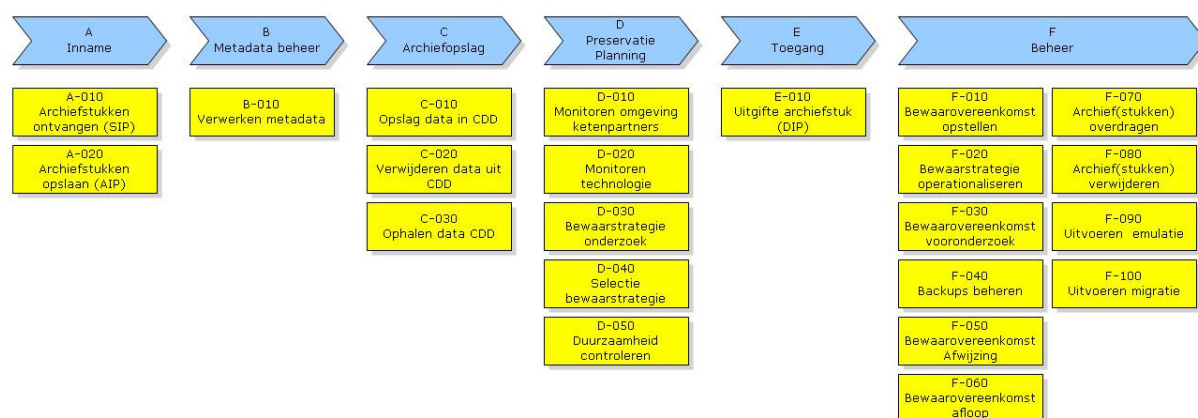
- Bestandsformaat herkenningsservice. Deze service moet de digitale bestanden herkennen (informatietype, software versie), maar ook de inhoud van de digitale bestanden bepalen en valideren;
- Aanbevelingsservice. Een service die tot een onderbouwde aanbeveling komt voor de te kiezen migratietool en het doel bestandsformaat of de te kiezen emulator en emulator module invulling (Welke applicatie, besturingssysteem, hardware?);
- Migratie-uitvoerservice. De service die de originele digitale bestanden converteert met een migratietool naar de doel bestandsformaten;
- Emulatie-uitvoerservice. De service die daadwerkelijk de emulatie uitvoert op het systeem van de klant of op afstand over een interactieve verbinding;
- Evaluatieservice. Een service die het informatieverlies bepaalt dat gemoeid is met de migratie of emulatie;
- Informatieservice. Een service die informatie geeft over welke bestandsformaten en software het risico lopen te verouderen en welke bestandsformaten en software aan te raden zijn;
- Vooronderzoekservice. Een services die bewaar- en representatie-eisen van ketenpartners opgesteld;
- Experimentsservice. Een service die experimenten met bewaarstrategieën inricht, uitvoert en op de

gestelde criteria meet.

Bovenstaande services zijn merendeel gericht aan klanten. Er is geen onderscheid gemaakt naar services aan de interne en externe organisatie. JustID zal echter het merendeel van bovenstaande services aan haar interne organisatie moeten leveren of gebruik maken van de beschreven services vanuit een extern systeem. Zodra dit is gerealiseerd kan JustID overwegen enkele interne services ook te gaan leveren aan haar ketenpartners.

6.2. Processen

Lavoie ea. [Lavo06] stellen processen binnen een digitaal archief op, verdeeld naar de onderdelen van het OAIS (paragraaf 2.3.3). Deze onderdelen stellen de procesgroepen voor, waaronder processen hangen. Ditzelfde principe, maar dan gericht op digitale duurzaamheid, levert onderstaande overzicht (figuur 16). Hierin komen de processen terug die beschreven staan in hoofdstuk 3 en 5, maar ook enkele algemene processen. De diepgang van de algemene processen is beperkt.



Figuur 16 - Processen in het kader van digitale duurzaamheid

De volgende procesgroepen en processen worden in het kader van digitale duurzaamheid onderscheiden:

A: Inname

A-010: Archiefstukken ontvangen (SIP) Archiefstukken worden door ketenpartners overgedragen naar het CDD.

A-020: Archiefstukken opslaan (AIP) Archiefstukken of delen hiervan worden verwerkt naar AIP's.

Tabel 17 - Procesgroep A: Inname

B: Metadata beheer

B-010: Verwerken metadata Metadata worden verwerkt in de database. Het kan hier gaan om het toevoegen, verwijderen of updaten van metadata.

Tabel 18 - Procesgroep B: Metadata beheer

C: Archiefopslag

C-010: Opslag data in CDD	Het daadwerkelijk opslaan van digitale bestanden.
C-020: Verwijderen data uit CDD	Het verwijderen van digitale bestanden.
C-030: Ophalen data CDD	Het ophalen van digitale bestanden.

Tabel 19 - Procesgroep C: Archiefopslag

D: Preservatie planning

D-010: Monitoren omgeving ketenpartners	De omgeving van de ketenpartner zal veranderen door de tijd heen. Het is dan ook van belang tijdig op handen zijnde veranderingen te constateren en hierop de anticiperen. Veranderingen in het domein van de ketenpartner zijn te vinden op het gebied van de techniek die gebruikt wordt (hard- en software) en de wet- en regelgeving die van kracht is.
D-020: Monitoren technologie	Ketenpartners zullen niet voorop lopen met de techniek. Veelal zijn soft- en hardware ver ontwikkeld voordat het wordt toegestaan binnen overheden. Om bekend te zijn met de veranderende techniek en de mogelijk daarmee gemoeide representatie problemen, moet de technologie goed in de gaten worden gehouden.
D-030: Bewaarstrategie onderzoek	Er zijn zeer veel verschillende bewaarstrategieën. Om te weten te komen welke bewaarstrategie de beste is, moet een veelvoud van bewaarstrategieën worden getest. De meetwaarden moeten bewaard worden om later te gebruiken bij het komen tot een onderbouwde keuze voor een bewaarstrategie gezien een context en informatietype.
D-040: Selectie bewaarstrategie	Er zijn bewaarexperimenten uitgevoerd, meetwaarden opgeslagen, getransformeerd naar te vergelijken waarden en weegfactoren zijn door ketenpartners toegekend aan criteria. Er kan nu berekend worden wat de meest geschikte bewaarstrategie is gezien een bepaalde context (die van de ketenpartner) en informatietype (bijvoorbeeld document).
D-050: Duurzaamheid controleren	De duurzaamheid moet na iedere monitorcycli worden gecontroleerd. Herberekening van bewaarstrategieën aan de hand van veranderingen van in de context een techniek.

Tabel 20 - Procesgroep D: Preservatie planning

E: Toegang

E-010: Uitgifte archiefstuk (DIP)

Archiefstukken moeten gerepresenteerd kunnen worden aan gerechtigden. Dit proces zorgt voor het opstellen van een DIP; de manifestatie van het archiefstuk die de aanvragen wil inzien.

Tabel 21 - Procesgroep E: Toegang

F: Beheer

F-010: Bewaarovereenkomst opstellen

Er moet niet zomaar digitale informatie in het CDD wordt gezet ter bewaring. Misschien kan het CDD helemaal niet de gegarandeerde bewaring waarmaken voor de verwachte bewaartermijn! Daarom moeten er bewaarovereenkomsten worden gesloten met ketenpartners. Deze overeenkomsten leggen onder andere vast voor welke termijn JustID zal voorzien in de bewaring van een bepaald type archiefstuk en welke digitale informatie deze archiefstukken mogen bevatten.

F-020: Bewaarstrategie operationaliseren

Om geteste en geselecteerde bewaarstrategieën te kunnen gebruiken moeten deze in het CDD operationeel gemaakt worden.

F-030: Bewaarovereenkomst vooronderzoek

Om te komen tot een bewaarovereenkomst moet het een en ander worden geïnventariseerd en bepaald worden. Dit kan worden uitgevoerd in een kort vooronderzoek bij de ketenpartner.

F-040: Backups beheren

Met regelmaat moeten er (externe) backups worden gemaakt van de informatie in het CDD. Daarnaast moet er hardware of complete systemen klaar staan om te voorzien in uptime en bitstream bewaring.

F-050: Bewaarovereenkomst Afwijzing

Het kan dat er (tijdelijk) niet tot een bewaarovereenkomst gekomen kan worden. Dit zou kunnen optreden wanneer er nog geen geschikte bewaarstrategie beschikbaar is voor de digitale informatie die een ketenpartner wil bewaren. Deze situatie ontstaat wanneer een ketenpartner zeer hoge eisen stelt aan de bewaring, bijvoorbeeld de eis dat de gehele "look-and-feel" duurzaam bewaard moet worden voor zeer lange tijd.

F-060: Bewaarovereenkomst afloop

Een bewaarovereenkomst is eindig. Ooit zal de digitale informatie naar het Nationaal Archief moeten worden overgedragen of zal de bewaarovereenkomst moeten worden herzien/verlengd.

F-070: Archief(stukken) overdragen	Archieven en archiefstukken moeten kunnen worden overgedragen naar alternatieve archiefsystemen, zoals die van het Nationaal Archief of een toekomstige opvolger van het CDD.
F-080: Archief(stukken) verwijderen	Na overdracht zouden archieven verwijderd mogen worden.
F-090: Uitvoeren emulatie	Emulatie moet daadwerkelijk worden uitgevoerd in de geselecteerde emulatie omgeving.
F-100: Uitvoeren migratie	Migraties moeten worden uitgevoerd met de daadvoor geselecteerde migratietool.

Tabel 22 - Procesgroep F: Beheer

De bovenstaande processen richten zich op digitale duurzaamheid en maken het mogelijk de eerder beschreven services te leveren. De processen die nodig zijn bij het kunnen leveren van de services staan in tabel 23 per services toegewezen. De processen onder “D: Preservation planning” komen overeen met de al in het functiemodel CDD+ [Delo06] opgestelde functies. Deze zijn in het functiemodel CDD+ aangemerkt als functies die uitgevoerd moeten worden door de organisatie.

Proces \ Service	Bestandsformat herkenningsservice	Aanbevelingsservice	Migratie- uitvoerservice	Emulatie- uitvoerservice	Evaluatieservice	Informatieservice	Vooronderzoekservice	Experimentsservice
A-010: Archiefstukken ontvangen (SIP)	X							
A-020: Archiefstukken opslaan (AIP)		X						
B-010: Verwerken metadata	X	X	X		X		X	
C-010: Opslag data in CDD			X					
C-020: Verwijderen data uit CDD								
C-030: Ophalen data CDD			X	X				
D-010: Monitoren omgeving ketenpartners		X				X	X	X
D-020: Monitoren technologie	X	X				X		X
D-030: Bewaarstrategie onderzoek					X			X
D-040: Selectie bewaarstrategie		X				X		
D-050: Duurzaamheid controleren		X				X		
E-010: Uitgifte archiefstuk (DIP)			X	X				
F-010: Bewaarovereenkomst opstellen							X	
F-020: Bewaarstrategie operationaliseren			X	X				
F-030: Bewaarovereenkomst vooronderzoek							X	
F-040: Backups beheren						X		
F-050: Bewaarovereenkomst Afwijzing							X	
F-060: Bewaarovereenkomst afloop						X	X	
F-070: Archief(stukken) overdragen								
F-080: Archief(stukken) verwijderen								
F-090: Uitvoeren emulatie				X				X
F-100: Uitvoeren migratie			X					X

Tabel 23 - Processen binnen services

6.3. Taken en afdelingen

De beschreven processen zijn uit te werken naar taken. Deze taken worden op hun beurt uitgevoerd door medewerkers of door het CDD systeem. Deze medewerkers vervullen een bepaalde functie en maken deel uit van een afdeling. In tabel 24 worden de taken per proces toegewezen aan afdelingen binnen JustID. Hiervoor is de beschrijving van de afdelingen gebruikt vanuit de recent binnen JustID uitgewerkte Administratieve Organisatie, weergegeven in paragraaf 1.2.2.

In onderstaande tabel wordt niet uitgediept naar specifieke functies binnen de afdelingen. Het ligt een stap dieper om binnen de afdelingen de taken toe te wijzen aan de functies binnen de afdelingen; deze scriptie heet niet tot doel een complete Administratieve Organisatie te beschrijven (paragraaf 2.5). Bij het invulling geven van het werk van individuele functies kan gebruik gemaakt worden van de

sociotechnical systems aanpak [Daft01]. Deze aanpak zorgt voor het rekening houden met de interactie tussen mens en techniek, waarbij de wensen van medewerkers afgestemd worden met de eisen die de organisatie stelt aan technische efficiëntie.

Afdeling	Groep	Proces	Taak
Systeem CDD	Archiefopslag	Opslag data in CDD	Opslaan bestanden juiste plek
		Verwijderen data uit CDD	Verwijderen bestanden
		Ophalen data CDD	Ophalen bestanden
	Beheer	Backups beheren	Kwaliteitscontrole hardware en beheer uitwijk hardware/systemen
		Uitvoeren emulatie	Samenstellen emulatoromgeving
			Emulator uitvoeren
	Inname	Archiefstukken ontvangen (SIP)	Emulator beschikbaar stellen aan uitgifte proces
			Bestandsformaat en inhoudbepaling/validatie en opslag
			Opslaan in quarantaine
	Metadata beheer	Archiefstukken opslaan (AIP)	Beslissen: Virus vrij?
			Beslissen: Binnen overeenkomst?
			Beslissen: Langdurige bewaring noodzakelijk?
Preservatie Planning	Verwerken metadata	Verwerken metadata	
		Duurzaamheid controleren	Duurzaamheid archief controleren
		Beslissen: Bewaaracties noodzakelijk?	
Toegang	Selectie bewaarstrategie	Koppel meest geschikte bewaarstrategie aan manifestatie	
		Geef overzicht van bewaarstrategieën en scores	
		Beslissen: Selectie in het kader van vooronderzoek?	
Toegang	Uitgifte archiefstuk (DIP)	Manifestatie aanbieden ter download	
		Interactieve emulatie afleveren aan verzoeker	
		Emulator bij download voegen	
		Beslissen: Download?	

			Beslissen: Manifestatie ter bewaring? Beslissen: Migratie uitvoeren? Beslissen: Migratiebewaarstrategie gekoppeld?
Informatiebeheer	Beheer	Archief(stukken) overdragen	Archief (AIPs) overdragen
		Archief(stukken) verwijderen	Beslissen: Correct?
	Inname	Archiefstukken ontvangen (SIP)	Meedelen aan ketenpartner Beslissen: Behoud archiefstuk?
	Metadata beheer	Verwerken metadata	Toevoegen metadata
	Preservatie Planning	Monitoren omgeving ketenpartners	Verzoek verificatie technische omgeving en wet-/regelgeving Beslissen: Verandering?
		Duurzaamheid controleren	Beslissen: Conform overeenkomst?
ICT en Innovatie	Beheer	Backups beheren	Externe backups maken
Informatieverwerking en Digitaal dossier	Beheer	Bewaarovereenkomst vooronderzoek	Beslissen: Geschikte bewaarstrategie?
		Uitvoeren migratie	Migreren bestanden met gekoppelde migratietool Controleren mate van behoud van informatie Afwijzen migratie & melden Voldoende informatiebehoud
	Preservatie Planning	Monitoren omgeving ketenpartners	Beslissen: Betere bewaarstrategie?
		Monitoren technologie	Technology assessments Updaten gegevens Beslissen: Verandering?
		Bewaarstrategie onderzoek	Definiëren bewaarstrategie alternatieven Ontwerpen bewaarexperiment Uitvoeren experiment

			Evalueren No-Go & actie plannen Evalueren experiment Transformeren meetwaarden Toevoegen in bewaarstrategie resultaten register Go/No-Go
		Selectie bewaarstrategie	Berekenen meest geschikte bewaarstrategie Beslissen: Bewaarstrategie geschikt?
Management	Beheer	Bewaarovereenkomst opstellen	Vastleggen overeenkomst Ketenpartner akkoord
		Bewaarovereenkomst Afwijzing	Beoordelen redelijkheid afwijzing en mogelijkheden tot aanpassing Afwijzing bewaarovereenkomst Positieve reactie ketenpartner? Concept overeenkomst aanpassen? Beslissen: Bestaande overeenkomst?
Projecten en Ontwikkeling	Beheer	Bewaarovereenkomst opstellen	Overeenkomst opstellen Beslissen: Overeenkomst correct?
		Bewaarstrategie operationaliseren	Gekozen migratietool installeren/testen Gekozen emulatie omgeving voorbereiden/testen Beslissen: Migratie uitvoeren? Beslissen: Migratiebewaarstrategie?
		Bewaarovereenkomst vooronderzoek	Context inventariseren Opstellen/koppelen criteria en weegfactoren Opstellen/koppelen criteria en weegfactoren
Relatiemanagement	Beheer	Bewaarovereenkomst opstellen	Beslissen: Is klant?
		Bewaarovereenkomst afloop	Informeren afloop bewaartermijn en vervolg acties Informeren afloop bewaartermijn en vervolg acties Aanbieden verlenging bewaarovereenkomst

	Beslissen: Positieve reactie ketenpartner?
	Beslissen: Bestaande overeenkomst herzien?

Tabel 24 - Taken per proces toegewezen aan afdelingen

De volgende afdelingen krijgen het meest nadrukkelijk te maken met processen rondom digitale duurzaamheid:

Projecten en Ontwikkeling

Processen die de basis vormen om überhaupt bewaarstrategieën te kunnen overwegen zijn ondergebracht bij de afdeling Projecten en Ontwikkeling. Zij zullen de bewaar- en representatie-eisen bij nieuwe ketenpartners moeten inventariseren, en, bij veranderingen door de tijd heen, deze routinematig herzien. Hiervoor moeten zij expertise hebben van de bedrijfsprocessen, ICT en archivistiek. Deze kennis hoeft niet noodzakelijk binnen de afdeling Projecten en Ontwikkeling te liggen, maar men moet deze kennis wel beschikbaar hebben tijdens het uitvoeren van vooronderzoek en het bijstellen van bewaarovereenkomsten. Deze afdeling moet zich dan ook richten op een gestroomlijnde communicatie om met regelmaat afstemming tussen experts te bereiken.

Informatieverwerking en Digitaal dossier

De processen die zich bezig houden met de onderzoekskant van digitale duurzaamheid zijn ondergebracht bij de afdeling Informatieverwerking en Digitaal dossier. Zij zullen experimenten gaan opzetten en onderzoek doen naar bewaarstrategieën. Dit onderzoek moet goed beschreven worden, herhaaldelijk worden uitgevoerd en gebruik maken van de eerder opgestelde criteria. Het werk is daarmee sterk analytisch. Daarnaast bevat het een hoge variëteit van taken en handelingen. Deze worden echter beschreven vanuit procedures en technieken, gelegen in de bewaarstrategie-evaluatiemethode. Daarmee zijn de onderzoeksgelateerde taken bij de afdeling Informatieverwerking en Digitaal dossier te typeren als “engineering” taken [Daft01]. Om de taken te kunnen uitvoeren is daarom veel kennis van zaken nodig en handelingen komen met enige regelmaat terug.

6.4. Gegevensdragers

Per taak worden verschillende gegevensdragers gecreëerd, geraadpleegd, geüpdate of verwijderd. Deze gegevensdragers worden per taak benoemd in bijlage 10. Onderstaande gegevensdragers zijn geïdentificeerd en weergegeven in twee kolommen (tabel 25).

Actieplanning	Concept bewaarovereenkomst
Afwijbsbericht	Digitale bestanden
AIP	Criteria
Archiefstuk	Emulator frame
Backups	Emulator modulen
Basisgegevens collectie	Evaluatie
Beschrijving huidige technische omgeving	Hardware onderdelen
Beschrijving opgeslagen technische omgeving	Huidige hardware systemen
Beschikbare geschikte bewaarstrategieën	Klantenbestand
Beschrijving huidige wet-/ regelgeving	Metadata
Beschrijving opgeslagen wet- en regelgeving	Migratietool
Bewaarovereenkomst	Steekproef collectie
Bewaarrapport	Technology reports
Bewaarstrategie alternatieven	Weegfactoren per criteria
Bewaarstrategie resultaten	

Tabel 25 - Gegevensdragers

6.5. Conclusie

In dit hoofdstuk is de organisatie inrichting beschreven die de organisatie toegespitst op de processen en taken die zij moet uitvoeren om services in het kader van digitale duurzaamheid te kunnen leveren aan de interne organisatie, dan wel aan ketenpartners. Afdelingen die met name taken zullen gaan uitvoeren gericht op het waarborgen van digitale duurzaamheid zijn de afdeling Projecten en Ontwikkeling en de afdeling Informatieverwerking en Digitaal dossier.

Het denken in services maakt het mogelijk dat ook vanuit de externe omgeving van JustID services kunnen worden aangesloten. Zo kan besloten worden om de onderzoeksgelateerde services, in het ontwerp ondergebracht bij de afdeling Informatieverwerking en Digitaal dossier, uit te besteden of samen te ondernemen met externe partijen.

In dit hoofdstuk zijn taken toegewezen aan de al aanwezige afdelingen en gegevensdragers aan taken. Het ligt een stap dieper om binnen de afdelingen de taken toe te wijzen aan de functies binnen de afdelingen. Dit dient te worden uitgevoerd binnen een vervolgonderzoek. De basis die binnen dit hoofdstuk is geschetst biedt inzicht in welke taken ondersteund moet worden met informatiesystemen. De architectuur die hierin voorziet wordt in het volgende hoofdstuk uitgewerkt.

Hoofdstuk 7 Architectuur

De in het voorgaande hoofdstuk beschreven organisatie voert taken uit in het kader van het operationaliseren van zowel de migratie- als emulatiebewaarstrategie. Daarnaast wordt beschreven hoe te bepalen wat de te volgen bewaarstrategie wordt op basis van bewaar- en representatie-eisen. Bij het uitvoeren van deze taken werken JustID medewerkers met het CDD systeem en met gegevensdragers. Het uitvoeren van deze taken en het werken met gegevensdragers dient zo veel mogelijk ondersteund te worden met informatiesystemen. Deze informatiesystemen worden in dit hoofdstuk behandeld en in een architectuur geplaatst, de informatiesysteem architectuur. Hierbij wordt informatie gebruikt van architectuuronderdelen die organisaties in de praktijk gebruiken (hoofdstuk 3).

7.1. Architectuuronderdelen

De aan de taken gekoppelde architectuuronderdelen worden in onderstaande paragrafen beschreven. In bijlage 10 staan deze architectuuronderdelen per taak benoemd.

7.1.1. Backup faciliteiten

Om te voorzien in de duurzame opslag van de bitstreams van de digitale informatie moeten er backup faciliteiten zijn. Te denken is aan een externe locatie waar de digitale archiefstukken ook met regelmaat opgeslagen worden. Onder backup-faciliteiten wordt ook het tijdig vernieuwen van hardware bedoeld.

7.1.2. Bewaaracties evaluatie tool

De resultaten die zijn ontstaan na het uitvoeren van bewaaracties moeten geëvalueerd worden. Dit wordt gedaan door te definiëren wat de digitaal gecodeerde informatie tot een unieke entiteit maakt en aan welke bewaareisen er voldaan moet worden. De resulterende representatie moet hieraan voldoen. Voorbeelden zijn het aantal karakters waar uit een document bestaat, de volgorde van deze karakters, het aantal pagina's en de grafische weergave en plaatsing hiervan. Bestandsformaat- en inhoudbepaling tools (paragraaf 7.1.6) worden gebruikt om de uniekheden te bepalen voor en na bewaaracties. De bewaaractie evaluatie tool moet zo veel mogelijk geautomatiseerd van de bestandsformaat- en inhoudbepaling tools gebruik maken en hier consequenties aan verbinden.

7.1.3. Bewaarstrategie adviseur

De bewaaradviseur is verantwoordelijk voor het bepalen van de beste bewaarstrategie. Het gebruikt hiervoor de bewaar- en representatie-eisen van ketenpartners, eisen aan het bewaaractieproces, en de resultaten van uitgevoerde bewaaracties (paragraaf 7.1.4). Al deze gegevens worden gebruikt om via een aangepaste utiliteitsanalyse te komen tot een score per bewaarstrategie (paragraaf 5.3).

7.1.4. Bewaarstrategie repository

In deze repository worden alleen bewaarstrategie meetresultaten opgeslagen. Deze worden door de bewaarstrategie adviseur gebruikt. Meetresultaten zijn afkomstig van bewaarstrategie experimenten en kunnen gedeeld worden met andere instellingen die zich bezighouden met digitale duurzaamheid.

7.1.5. Database

Hierin worden alle metadata opgeslagen in het metadatamodel (paragraaf 2.3.1).

7.1.6. Determinatie tools

Dit gedeelte van de architectuur moet voorzien in de herkenning van bestandsformaten en het bepalen en valideren van interne onderdelen van de digitale bestanden. Deze worden gebruikt tijdens de inname van bestanden en bij het valideren van uitgevoerde bewaaracties, bijvoorbeeld tijdens experimenten. Dit valideren geeft aan in welke mate de nieuwe representatievorm van de informatie gelijk is aan de originele informatie. Er zijn enkele applicaties geëvalueerd die in enige mate voorzien in bestandsformaat herkenning en inhoudbepaling (zie bijlage 6). De volgende worden aangeraden om in te zetten op dit onderdeel van de architectuur:

- DROID [15]. DROID staat voor Digital Record Object Identification en is een initiatief van The National Archives (UK). DROID werkt samen met bestandsformaat register PRONOM (zie paragraaf 7.1.12) om bestanden te kunnen herkennen aan de hand van enkele verschillende technieken.
- JHOVE [Jhov06][16]. JHOVE staat voor “JSTOR/Harvard Object Validation Environment” en is een initiatief van JSTORE en Harvard University. Deze tool richt zich op het bepalen van de interne onderdelen waaruit bestanden opgebouwd zijn en kan worden gebruikt om tot een uiteenraffing te komen van informatietypen en kan gebruikt worden bij het invullen van criteria. JHOVE2 beoogd gebruikt te gaan maken van DROID voor de bestandsformaat herkenning.

Er is helaas nog geen software om in het kader van een emulatiebewaarstrategie het geëmuleerde resultaat geautomatiseerd te valideren. Het geëmuleerde is namelijk een (interactieve) applicatie en er is nog geen software om de daarin gecodeerde informatie geautomatiseerd te bepalen, noch te vergelijken met de in de originele bestandsformaten gecodeerde informatie. Dit is een punt voor vervolgonderzoek.

7.1.7. Emulatie omgeving

Dit deel van de architectuur is verantwoordelijk voor het uitvoeren van de emulatie. De emulator modulen of emulatoren kunnen door derden worden geleverd of intern worden ontwikkeld. Daarnaast kan het zijn dat de emulatoren worden uitgevoerd op de PC van de ketenpartner of vanaf een centrale server. In het kader van de privacy gevoeligheid van de informatie in het CDD, heeft het de voorkeur om de emulaties intern uit te voeren.

Een punt waar vervolgonderzoek naar moet worden verricht is hoe een geëmuleerde representatie interactief bij de ketenpartner te krijgen, zonder dat de gehele emulator en originele bestanden moeten worden verzonden.

7.1.8. Migratie omgeving

Dit deel van de architectuur is verantwoordelijk voor het bieden van een omgeving om migraties in uit te voeren. Dit kan extern (service door derden) of intern worden uitgevoerd. In het kader van de privacy gevoeligheid van de informatie in het CDD, heeft het de voorkeur om de bestandsmigraties intern uit te voeren.

7.1.9. Mindmapping tool

Voor het opstellen van de criteriaboom (paragraaf 5.1) kan gebruik worden gemaakt van een mindmapping tool. Deze tool voorziet in het opstellen van de takken van de criteriaboom, waarbij meeteenheden zijn te definiëren en in te vullen. Daarnaast kunnen aan de takken van de criteriaboom weegfactoren worden gekoppeld. De software zou gebruik moeten kunnen maken van sjablonen per informatietype en/of context.

Een punt waar vervolgonderzoek naar moet worden verricht is wat de meest efficiënte software is om de criteriaboom op te stellen.

7.1.10. Planningtool

Een planningtool moet voorzien in het tijdig herinneren van te ondernemen acties. Een bewaar-overeenkomst kan bijvoorbeeld aflopen of er kunnen digitale bestanden in het CDD staan waaraan nog geen bewaarstrategie gekoppeld is. Daarnaast moet er met regelmaat gecontroleerd worden of er nog nieuw technologische ontwikkelingen zijn of dat ketenpartners veranderingen hebben doorgevoerd die van invloed zijn op de duurzame bewaring van informatie in het CDD.

7.1.11. Quarantaine omgeving

Binnengekomen digitale bestanden dienen gedurende een korte tijd in een quarantaine omgeving worden opgenomen. Dit voorziet in een periode waarin nieuwe virussen geïndexeerd worden door virusscanner leveranciers, zodat digitale bestanden ook op deze nieuwe virussen gescand worden voordat deze definitief in het archief worden opgenomen.

7.1.12. Representatie Informatie register

Het onderdeel is verantwoordelijk voor het beschikbaar hebben van feiten over bestandsformaten, softwareleveranciers, migratietools, emulator modulen en andere benodigde informatie noodzakelijk voor de representatie van digitale gecodeerde informatie. De volgende registers worden voorgesteld:

- GDFR registry [Abra05][20]. Een initiatief van Harvard University Library dat als doel heeft het ontdekken en leveren van representatie informatie over digitale bestandsformaten. Eind 2008 moet dit register operationeel zijn, waarbij PRONOM mogelijk een onderdeel van dit register gaat vormen. PRONOM [Darl03][Brow06][18] is een initiatief van The National Archives (UK) dat zich richt op het vergaren van informatie over alle bestandsformaten, software leveranciers, besturingssystemen, hardware componenten en andere technische onderdelen die spelen rond digitale duurzaamheid. Momenteel bevat PRONOM informatie rondom ongeveer 500 bestandsformaten en werkt samen met DROID (paragraaf 7.1.6);
- Representation Information Registry/Repository [Rusb05][Giar05][21]. Een initiatief van het Digital Curation Centre (UK) om te voorzien in representatie informatie op de manier van het OAIS. Deze register heeft de insteek om alle representatie informatie die noodzakelijke is ter representatie van de digitale informatie in registers op te nemen. Dit zal ook de informatie omvatten die noodzakelijk is voor de emulatiebewaarstrategie.

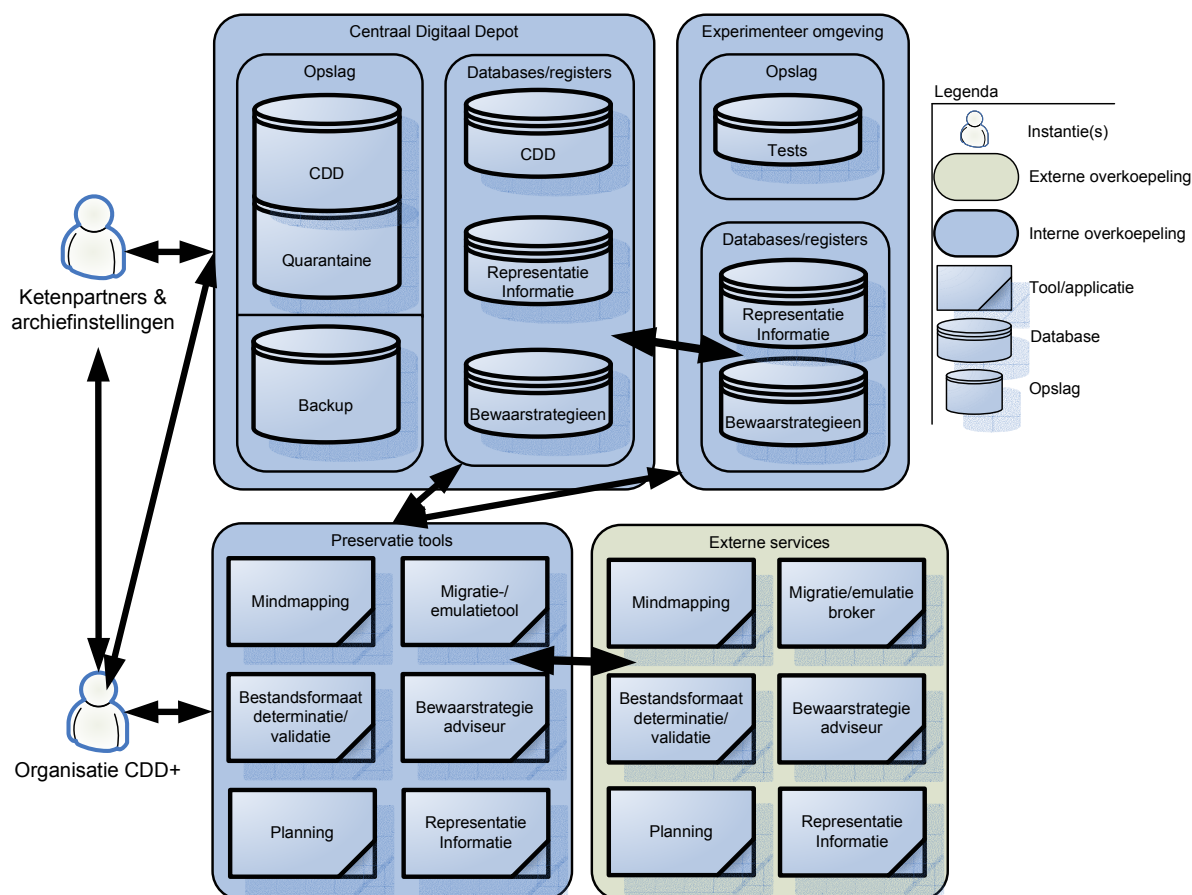
Het kan dat het GDFR niet geheel gaat voorzien in het kunnen koppelen van alle gewenste metadata aan de gedocumenteerde representatie informatie. In dit geval zal er intern een eigen database met aanvullende representatie informatie worden bijgehouden.

7.1.13. Storage omgeving

Dit is de omgeving waar de digitale bestanden worden opgeslagen en die voorziet in de bewaring van de bitstream. Daarvoor moet het media waarop deze bitstreams zich bevinden regelmatig worden verversd door bijvoorbeeld hardware te vervangen of backups te maken.

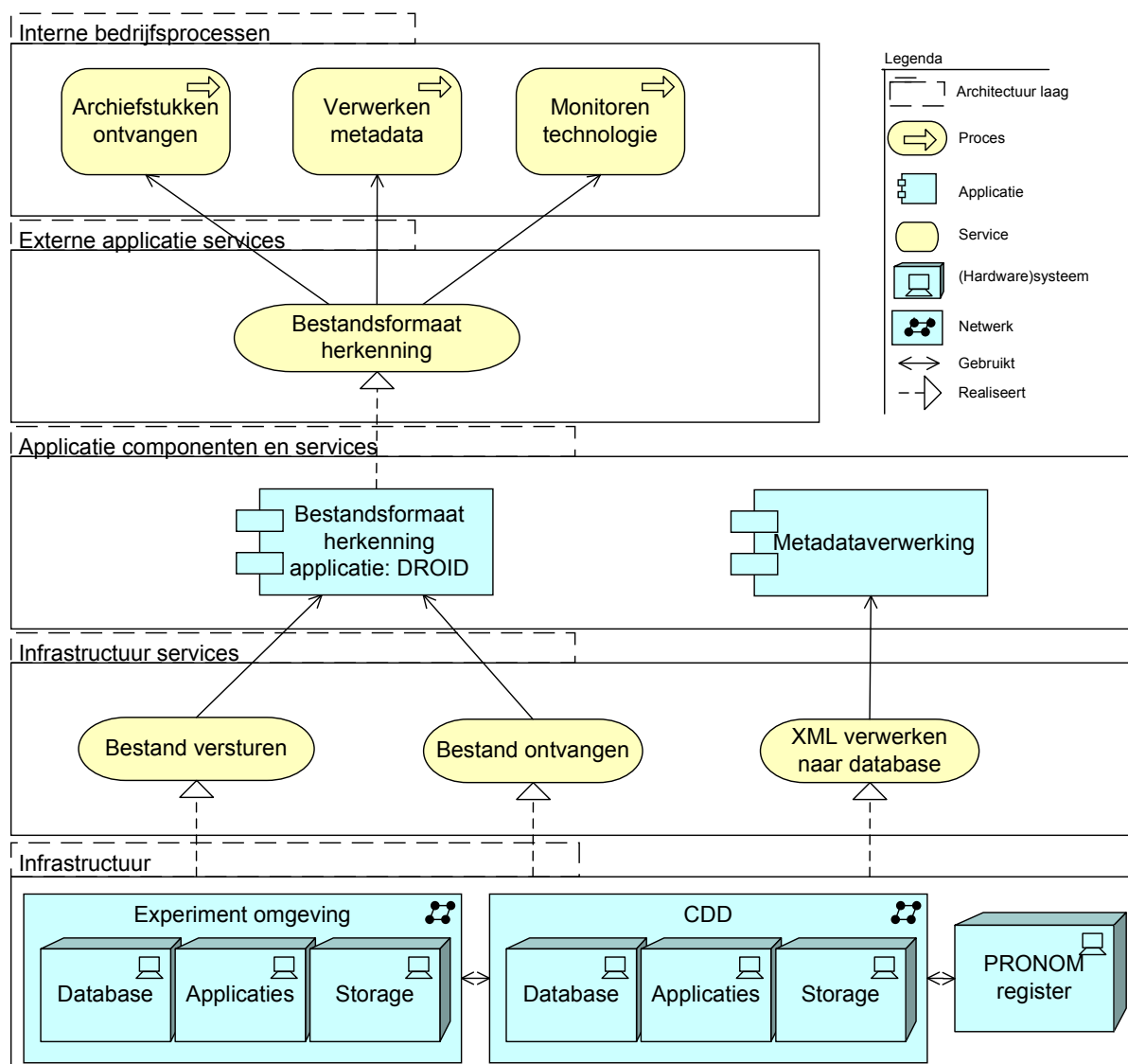
7.2. Architectuur

De besproken architectuuronderdelen zijn in onderstaande figuur (figuur 23) met elkaar in verband gebracht en vormen een globaal overzicht van de informatiesysteem architectuur die de organisatie zal ondersteunen in haar taken rondom digitale duurzaamheid. Ook eventuele externe services zijn in dit overzicht weergegeven. Deze spelen een zeer belangrijke rol bij het omlaag brengen van kosten die gemoeid gaan bij het ontwikkelen van bijvoorbeeld migratietools of emulator modules. Daarnaast wordt informatie bij elkaar gebracht in registers, zodat de zoektocht naar specificaties van bestanden en software verkort wordt.



Figuur 23 – Globale IS Architectuur

Bovenstaande globale IS architectuur is als voorbeeld in onderstaande figuur (figuur 24) dieper uitgewerkt vanuit de bestandsformaat herkenningsservice. Van onder naar boven levert fysieke hardware in een netwerk infrastructurele services aan die gebruikt worden door applicaties. Deze applicaties leveren de services die in de interne bedrijfsprocessen worden gebruikt



Figuur 24 - Architectuur lagen bestandsformaat herkenningsservice

7.3. Conclusie

In dit hoofdstuk is een overzicht geboden van de architectuuronderdelen die, verbonden via een architectuur, de organisatie zullen ondersteunen bij het verrichten van haar taken gericht op digitale duurzaamheid. Enkele delen van de architectuur kunnen ingevuld worden door services van derden. Door deze externe services te gebruiken kan geld bespaard worden omdat deze delen dan niet intern ontwikkeld dienen te worden. JustID zou echter ook kunnen bijdragen door zelf services te leveren aan externe partijen. De keuze tussen intern ontwikkelen, of gebruik maken van externe services, moet verder onderzocht worden.

Hoofdstuk 8 Conclusies en vervolgonderzoek

De centrale onderzoeksvraag binnen deze scriptie is de volgende:

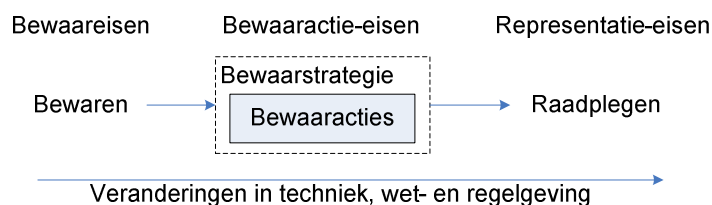
Hoe moet het Expertise Centrum CDD+ van JustID, dat zich binnen de Strafrechtsketen richt op digitale duurzaamheid van archieven in het CDD, organisatorisch en met ondersteunende informatiesystemen worden ingericht?

Deze wordt middels voorgaande hoofdstukken stap voor stap beantwoordt. Dit hoofdstuk geeft de overkoepelende conclusie die de centrale onderzoeksvraag beantwoordt. Daarnaast worden mogelijk onderwerpen voor vervolgonderzoek aangedragen.

8.1. Conclusies en aanbevelingen

De bewaarstrategieën om informatie digitaal duurzaam te bewaren worden geschaald onder twee overkoepelende bewaarstrategieën: de migratie- en emulatiebewaarstrategie. De keuze voor de meest geschikte bewaarstrategie is afhankelijk van een set van eisen. Deze eisen verschillen per creatiecontext (archiefstuk opstellende ketenpartner) en gebruikerscontext (JustID en de archiefstuk raadplegende ketenpartner). De volgende sets van eisen zijn onderscheiden (figuur 25):

- Bewaareisen, afkomstig van de ketenpartner die archiefstukken digitaal in bewaring wil stellen. Zij wil bepaalde visuele en contextuele ervaringen duurzaam digitaal bewaren en toegankelijk houden voor een bepaalde termijn;
- Representatie-eisen, afkomstig van de ketenpartner die archiefstukken digitaal willen raadplegen. Zij moet bij het raadplegen de door de bewaarder gestelde visuele en contextuele ervaringen, ervaren via de dan in gebruik zijnde technologie;
- Bewaaractieproces-eisen, afkomstig van raadplegende ketenpartner en JustID's interne organisatie. De ketenpartner wil dat het resultaat van het bewaaractieproces tijdig en authentiek aan hen gepresenteerd wordt. JustID's interne organisatie voert de bewaaractieprocessen uit, of laat deze extern uitvoeren. Zij stelt aanvullende eisen die ingaan op de kosten, transparantie en toepasbaarheid van het bewaaractieproces.



Figuur 25 - Eisen aan de bewaring van digitale informatie

De bewaarstrategie-evaluatiemethode wordt gebruikt om tot een keuze te komen voor de meeste geschikte bewaarstrategie. Een van de eerste stappen binnen deze methode is het in kaart brengen van de eisen en het hieruit opstellen van criteria. Dit wordt gedaan binnen een vooronderzoek of routinematige (her)inventarisatie van de bij de ketenpartners in gebruik zijnde technologie. Mindmapping tools helpen hierbij, maar het is aan te bevelen verder onderzoek te doen naar welke methode van inventariseren van criteria het beste werkt en wat daarvoor nodig is.

Vervolgens worden deze criteria gebruikt om bewaarstrategieën binnen experimenten te testen en te evalueren. Daartoe moet JustID een omgeving inrichten waarop migratie- en emulatieacties

uitgevoerd, getest en gemeten kunnen worden. Met behulp van de meetresultaten op de criteria wordt berekend wat de meest geschikte bewaarstrategie is.

Met de meest geschikte bewaarstrategie geselecteerd, dient deze te worden geoperationaliseerd. De experimenten, waarbinnen de bewaarstrategieën uitgevoerd zijn, dienen gedocumenteerd te worden. Deze documentatie maakt het mogelijk om de experimenten herhaaldelijk uit te kunnen voeren, en dient als basis om binnen de productieomgeving het bewaaractieproces te implementeren.

Bewaaractieprocessen kunnen meerdere bestanden als resultaat hebben, die samen één manifestatie vormen, zoals een website of een in XML opgeslagen document. Het datamodel van het CDD voorziet nog niet in het opslaan van deze manifestaties, anders dan deze op te slaan in een inpak bestandsformaat. Deze neemt de bestanden die opgeslagen moeten worden samen, wat maakt dat de bestanden niet individueel te benaderen zijn met relaties vanuit het metadatamodel. Het wordt aanbevolen het datamodel aan te passen.

Er zijn processen opgesteld en uitgewerkt in taken die gericht zijn op het garanderen van de duurzame toegankelijkheid van de in het CDD opgeslagen digitale informatie. Aan deze taken zijn benodigde gegevensdragers en ondersteunende architectuuronderdelen gekoppeld. De taken zijn onderverdeeld naar de afdelingen binnen JustID, dat de nodige veranderingen met zich meebrengt. Afdelingen die zich voornamelijk bezig zullen houden met het waarborgen van digitale duurzaamheid zijn de afdeling Projecten en Ontwikkeling en de afdeling Informatieverwerking en Digitaal dossier. Een volgende stap is om binnen de afdelingen de taken toe te wijzen aan de functies binnen de afdelingen en eventuele nieuwe functies te ontwerpen. Het is aan te bevelen dit nader uit te werken.

De mate waarin de ontworpen processen van invloed zijn op JustID is afhankelijk van de mate waarin JustID aansluiting vindt op services van derden. Zo geeft het PLANETS consortium aan verschillende architectuuronderdelen te gaan integreren en services te gaan bieden die aansluiten bij wat er in het kader van digitale duurzaamheid uitgevoerd moet worden door het CDD+. Zo kan besloten worden onderzoeksgerelateerde activiteiten, in het organisatie ontwerp ondergebracht bij de afdeling Informatieverwerking en Digitaal dossier, uit te besteden of samen te doen met externe partijen. Ook de migratie en emulatie bewaarprocesacties zouden extern kunnen worden uitgevoerd. Echter zal dit laatste niet de voorkeur genieten; JustID werkt met zeer privacy gevoelige. Een service die migratietools en emulatoren ter download aanbied, zodat deze lokaal uitgevoerd kunnen worden, zal echter wel uitkomst bieden. Dit zelfde geldt voor de inhoudelijke bepaling van digitale bestanden. Dit inhoudelijk bepalen is als essentieel aan te merken voor het CDD; men moet tenminste weten in welke coderingen informatie digitaal is opgeslagen in het CDD. In deze scriptie is aan bod gekomen welke services er te onderscheiden zijn en hoe deze te plaatsen zijn in de organisatie. Het is aan te bevelen verder te onderzoeken welke services extern ingevuld kunnen worden. In de discussie (paragraaf 8.2) wordt besproken wat de invloed kan zijn door samenwerking met andere archivistische instellingen op de focus van JustID.

Met het realiseren van de voorgestelde organisatie en ondersteunende architectuur voorziet JustID in het garanderen van digitale duurzaamheid van archieven in het CDD.

8.2. Discussie

Hoe relevant is het actief investeren in onderzoek naar bewaarstrategieën tegenwoordig eigenlijk en wat kan dit betekenen voor JustID? In welke mate verandert de rol van de archivist? Beide

onderwerpen worden hieronder kort bediscussieerd.

Relevantie investeren digitale duurzaamheid

Is het verstandig om actief te investeren in het uitvoeren van bewaarexperimenten en de ontwikkeling van representatiesoftware (migratietools/emulatoren) of kan de focus ook elders liggen?

Kijkende naar het verleden kan worden opgemerkt dat bijvoorbeeld 20 jaar geleden er veel meer verschillende besturingssystemen en applicaties waren dan nu. Zij hadden veelal eigen gesloten bestandsformaten en waren niet of nauwelijks onderling uit te wisselen. Met onwelwillendheid of het verdwijnen van een leverancier was het maar de vraag of er tijdig een migratie van de informatie uitgevoerd kon worden en in welke mate het resultaat dan overeen kwam met wat er ooit opgeslagen was.

Tegenwoordig is het aantal uiteenlopende besturingssystemen overzichtelijker, werken veel applicaties op meerdere besturingssystemen en kunnen zij overweg met bestandsformaten die steeds verder gestandaardiseerd zijn en waarvan de specificaties publiekelijk beschikbaar zijn. Dit resulteert in een afname van de variëteit van bestandsformaten per informatietype. Het standaard en open karakter van recente bestandsformaten zorgt er dan ook voor dat de daarin gecodeerde informatie langer toegankelijk blijft; de levensduur van informatie in één enkele manifestatie neemt toe. Iedere software ontwikkelaar kan namelijk de specificaties van het bestandsformaat gebruiken om informatie op te slaan in het gestandaardiseerde bestandsformaat. Eventuele bewaaracties zijn dan pas later nodig, zodra er noodzaak is om naar een andere standaard over te stappen.

Tegen de tijd dat een nieuwe standaard tot standaard verworden is en actief gebruikt gaat worden binnen de overheid, is het te veronderstellen dat er publiekelijk migratietools of emulatoren beschikbaar zijn die gebruikt kunnen worden om informatie te migreren. De activiteiten die ondernomen worden door (inter)nationale archieven en bibliotheken ondersteunen deze veronderstelling. Zij stellen zichzelf tot doel informatie in digitale vorm eeuwig te bewaren en toegankelijk te houden voor een veelvoud van mogelijke raadplegers. Dit resulteert in zeer strenge bewaar- en representatie-eisen, strengere dan ketenpartners, en daarmee het CDD+, zal stellen. Ketenpartners hebben namelijk niet de taak digitale informatie eeuwig te bewaren. De archiefstukken worden na een bewaartermijn overgedragen aan het Nationaal Archief, die een bepaald deel van het archief beoogt eeuwig te bewaren in het kader van het culturele erfgoed.

Dit verschil in bewaar- en representatie-eisen tussen eindige en eeuwige bewaring en het verschil in technologie tussen vroeger en nu heeft invloed op de focus van JustID. JustID zou zich in eerste instantie actief moeten richten op het adviseren van ketenpartners om open standaard bestandsformaten te gebruiken. Zij verminderd daarmee drastisch de eventueel in de toekomst uit te voeren bewaaracties, die op hun beurt al vergeleken kunnen zijn door archivistische instellingen. Om als archiverende instantie met kennis van zaken over te komen op ketenpartners zou JustID met andere archivistische instellingen samenwerken, om kennis richting ketenpartners actief beschikbaar te hebben.

Rol van de archivist

Is de rol van de archivistisch medewerker nog wel dezelfde als vroeger en in welke mate is zij voorbereid op de toekomst?

Met het zo goed mogelijk digitaal duurzaam houden van informatie verplaatst de rol van de archivist

meer en meer naar de digitale wereld. Het blijft echter nog altijd gaan over de creatiecontext, gebruikerscontext en de eigenschappen van de codering en opslag. Het onderzoeken van de gebruikte inkt en papiersoort worden ingeruild voor het onderzoeken van de verschillende eigenschappen van digitale informatie coderingen: het blijft gaan over eigenschappen van informatie en de manier van het coderen.

Men kan zich echter afvragen in welke mate archivistische medewerkers hierop voorbereid zijn in de context van digitalisering. Wanneer namelijk gekeken wordt naar de soort kennis die archivistische medewerkers óók nodig hebben, dan is dat bijvoorbeeld specialistische kennis van digitale coderingen en kennis over hoe hard- en software met de digitale coderingen omgaan. Wanneer dus archivistische medewerkers in de digitale wereld duurzaam archieven willen bijhouden, dienen zij gedegen kennis te hebben van alle tussenliggende (technische) lagen en de codering in en opslag van informatie in bits.

8.3. Vervolgonderzoek

Vanuit het uitgevoerde onderzoek zijn enkele onderwerpen geïdentificeerd waarin wetenschappelijk vervolgonderzoek wenselijk is. Onderstaande punten gaan hierop in.

Services

- Emulatiresultaat inhoudbepaling service. Emulaties die getoond worden op een systeem zijn niet noodzakelijk opgeslagen in digitale bestanden. De applicaties, besturingssystemen en hardwareonderdelen worden nagebootst. Daarbinnen wordt het originele digitale bestand geïnterpreteerd. Er is vervolgonderzoek noodzakelijk naar hoe de representatie-criteria (op een geautomatiseerde manier) te inventariseren zijn. Bij migratietools is het resultaat goed te vergelijken met het origineel; origineel en resultaat bestaan beiden uit digitale bestanden en de bitstreams daarin zijn met specialistische software te bepalen en valideren. Bij emulaties bevindt de informatie zich in het originele bestandsformaat en wordt het getoond via virtuele hardware en software, dat zich in het geheugen van het systeem bevindt. Een oplossingsrichting kan liggen in het bepalen van bitcode die zich bevindt in het geheugen van een systeem. Mogelijke onderzoeksvraag: “Hoe moet geëmuleerde digitale informatie gedetermineerd en gevalideerd worden, zodat deze gescoord kan worden op representatie-criteria en vergeleken kan worden met migratiresultaten?”;

Organisatie

- Emulatie aflevering. Bij het emuleren van digitale informatie moet een interactieve weergave worden afgeleverd bij de raadpleger. Het kan ingericht worden dat de emulator wordt uitgevoerd op het systeem van de raadpleger, maar dat zou inhouden dat grote hoeveelheden privacy gevoelige gegevens over het netwerk moeten worden verstuurd. De organisatie van een centraal platform, dat een soort van remote desktop mogelijkheid biedt voor het afleveren van interactieve ervaringen, kan hiervoor een oplossing bieden. Er is echter vervolgonderzoek nodig naar welke verschillende software interactieve ervaringen kan leveren aan raadplegers en op welke criteria deze verschillen. Daarnaast moet het centrale platform zelf duurzaam zijn. Zo zou de interactie afleverende software moeten kunnen samenwerken met de Universele Virtuele Computer (paragraaf 3.2.1), modulaire emulator (zie bijlage 7) of ander “duurzaam” systeem. Mogelijke onderzoeksvraag: “Hoe moet geëmuleerde digitale informatie zo efficiënt mogelijk interactief getoond worden aan raadplegers?”.

- Software Oriented Architecture (SOA) en emuleerbaarheid. Het gebruik van SOA wordt bedrijven aangeraden die te maken hebben met sterk veranderlijke bedrijfsprocessen en realiseert een bepaalde flexibiliteit door verschillende webservices te integreren. Welke gevolgen heeft dit echter voor de dan over de verschillende webservices versnipperde digitale informatie? Is de digitale informatie dan nog wel te emuleren?. Mogelijke onderzoeksvraag: “Welke invloeden heeft het gebruik van SOA op de emuleerbaarheid van digitale informatie?”.

Literatuur

Boeken/artikelen

- [Abra05] Abrams, S.L., *Establising a Global Digital Format Registry*, Library Trends, jrg. 54, nr. 1, pp. 125-143, 2005.
- [Barv07] Barve, S., *File Formats in Digital Preservation*, ARD Prasad & Devika P. Madalli (Eds.): ICSD-2007, pp. 239-248, 2007.
- [Beck07] Becker, C., Master thesis: *A Functional Model of a Digital Preservation Testbed*, Universiteit van Wenen, april 2007.
- [Boud02] Boudrez, F., *Standaarden voor digitale archiefdocumenten*, versie 4.1, Expertise centrum DAVID, Antwerpen-Leuven, 2002.
- [Boud04] Boudrez, F., Dekeyser, H., *Digitaal archiefbeheer in de praktijk: Handboek*, Stadsarchief Antwerpen, Antwerpen, 2004.
- [Boud05] Boudrez, F., *Digitale containers voor het digitaal archiefdepot*, Expertisecentrum DAVID, Antwerpen, 2005.
- [Brow06] Brown, A., Digital Preservation Technical Paper 1: Automatic Format Identification Using PRONOM and DROID, The National Archives, Australië, nr. 2, 2006.
- [Bush87] Bushoff, R., Oosterhaven, J.A., *Information Strategy Planning*, Informatie, jrg. 29, nr. 3, pp. 228-238, 1987.
- [Cons02] Consultative Committee for Space Data Systems, *Reference model for an Open archival Information System (OAIS)*, Blue Book, nr. 1, Washington, DC: CCSDS, 2002.
- [Daft01] Daft, R.L., *Organization Theory and Design*, 7de editie, Vanderbilt University, South-Western College Publishing, 2001.
- [Darl03] Darlington, J., *PRONOM – A Pratical Online Compendium of File Formats*, RLG DigiNews, jrg. 7, nr. 5, 2003.
- [Day03] Day, M., Preservation metadata initiatives: practicality, sustainability, and interoperability, UKOLN, In Proceedings Veröffentlichungen der Archivschule Marburg, 2003.
- [Delo06] Deloitte Consulting, Eindrapport Vooronderzoek CDD Strafrechtsketen voor de Justitiele Informatiedienst, Fase 1, 2006.
- [Dies02] Diessen, van, R.J., Steenbergen, J.F., *The Long-Term Preservation Study of the DNEP project - an overview of the results*, IBM / Koninklijke Bibliotheek, Amsterdam, 2002.
- [Eynd01] Eynde, van den, S., Wat archiveren en hoe? Op zoek naar de rol van PKI voor digitale archieven, Expertise centrum DAVID, Leuven, 2001.
- [Ferr05] Ferreira, M., Automatic Evaluation of Migration Quality in Distributed Networks of Converters, ECDL Doctoral Consortium Poster, 2005.
- [Ferr06] Ferreira, M., Baptista, A.A., Ramalho, J.C., *A Foundation for Automatic Digital Preservation*, ARIADNE, nr. 48, 2006.
- [Giar05] Giarretta, D., Patel, M., Rusbridge, A., Rankin, S., McIlwrath, B., *Supporting e-Research Using Representation Information*, In Proceedings All Hands Meeting, Nottingham, 2005.

- [Grij06] Grijpink, J.H.A.M., *Keteninformatisering in kort bestek*, 1^{de} druk, Lemma, 2006.
- [Hend93] Henderson, J.C., Venkatraman, N., *Strategic alignment: Leveraging information technology for transforming organizations*, IBM Systems Journal, jrg. 32, nr. 1, pp. 472-484, 1993.
- [Hoev05] Hoeven, J.R. van der, Wijngaarden, H.N. van, *Modular emulation as a long-term preservation strategy for digital objects*, International Web Archiving Workshop 2005 (IWA'05), Wenen, Oostenrijk, 2005.
- [Hoev06] Hoeven, J.R. van der, *Test Results Document: Experiment results and findings*, Koninklijke Bibliotheek, Nationaal Archief, Den Haag, 2006.
- [Hofm00] Hofman, H., *Een uitdijend heelal? Context van archiefbescheiden*, In: Context. Interpretatiekaders in de archivistiek, Stichting Archiefpublicaties, jaarboek 2000, Den Haag, pp. 45-61, 2000.
- [Hors04] Horsman, P.J., Jonker, A.E.M., *Archiveren. Een inleiding*, Archiefschrift 1, Stichting Archiefpublicaties, 's-Gravenhage, 2004/2006.
- [Hors05] Horsman, P.J., Waalwijk, H., van Bussel, G.-J., *Metadata voor Archiveringssytemen. Referentie model*, Stichting Archiefschool, Amsterdam, 2005.
- [Jonk05] Jonker, A.E.M., *Records Continuüm: een theoretische notitie*, Stichting Archiefschool, Amsterdam, 2005.
- [Knig07] Knight, G., SHERPA DP - Recommendations to ensure the long-term preservation of digital objects stored by institutional repositories, AHDS, 2007.
- [Lavo06] Lavoie, B., Henry, G., Dempsey, L., *A service Framework for Libraries*, D-Lib Magazine, jrg. 12, nr. 7/8, 2006.
- [Lawr00] Lawrence, G.W., Kehoe, W.R., Rieger, O.Y., Walters, W.H., Kenney, A.R., *Risk Management of Digital Information: A file format investigation*, Council on Library and Information Resources, DC: Washington, 2000.
- [Lori01] Lorie, R.A., *Long term preservation of digital information*, In Proceeding of the 1st ACM/IEEE-CS joint conference on Digital Libraries, VI: Roanoke, pp. 346-352, 2001.
- [Lori02] Lorie, R.A., *The UVC: a method for preserving digital documents – proof of concept*. LTP reports series, nr. 4, 2002.
- [Mani05] Maniatis, P., Roussopoulos, M., Giuli, T.J., Rosenthal, D.S.H., Baker, M., *LOCKSS: A Peer-to-Peer Digital Preservation system*, ACM Transactions on Computer Systems, jrg. 23, nr. 1, pp. 2-50, 2005.
- [Mint78] Mintzberg, H., *Patterns in Strategy Formation*, Management Science, jrg. 24, nr. 9, pp. 934-948, 1978.
- [Oltm05] Oltmans, E., Nanda, K., *A Comparison Between Migration and Emulation in Terms of Costs*, In RLG DigiNews, Jrg. 9, nr. 2, 2005.
- [Ploe02] Ploeg, van der, F., *Regeling geordende en toegankelijke staat archiefbescheiden*, OCW, Staatscourant 1 maart 2002, nr. 43, pp. 8, 2002.
- [Rauc04] Rauch, C., Rauber, A., *Preserving digital media: Towards a preservation solution evaluation metric*. In Proceedings of the 7th International Conference on Asian Digital Libraries, ICADL 2004, Springer, pp. 203-212, 2005.

- [Rauc05] Rauch, C., Strodl, S., Rauber, A., Preservation: A framework for Documenting the Behaviour and Functionality of digital Objects and Preservation Strategies, Project DELOS, University of Vienna, 2005.
- [Rooz92] Roozenburg, N.F.M., Eekels, J., Ontwerpmethodologie, TU Delft, Nederland, 1992.
- [Rose05] Rosenthal, D.S.H., T.S., Lipkis, Robertson, T., Morabito, S., *Transparent Format Migration of Preserved Web Content*, D-Lib Magazine, jrg. 11, nr. 1, 2005.
- [Roth99] Rothenberg, J., Bikson, T., Carrying Authentic Understandable and Usable Records Through Time, Den Haag, 1999.
- [Rusb05] Rusbridge, C., Burnhill, P., Ross, S., Buneman, P., Giarretta, D., Lyon, L., Atkinson, M., The Digital Curation Centre: A Vision for Digital Curation, In Proceedings From Local to Global: Data Interoperability - Challenges and Technologies, pp. 1-11, 2005.
- [Stam96] Stamper, R.K., *Organisational semiotics*, In Information Systems: An Emerging Discipline?, Stowell, F. en Mingers, J. (eds), McGraw-Hill, London/New York, pp. 267-283, 1996.
- [Stan04] Stanescu, A., Assessing the Durability of Formats in a Digital Preservation Environment, D-Lib Magazine, jrg. 10, nr 11, 2004.
- [Star97] Starreveld, R.W., Mare, H.B. de, Joëls, E.J., *Bestuurlijke informatieverzorging Deel 1: Algemene grondslagen*, 4^e druk, Samsom BedrijfsInformatie, Alphen aan den Rijn/Zaventem, 1994.
- [Stro06] Strodl, S., Rauch, C., Rauber, A., Hofman, H., Debole, F., Amato, G., *The DELOS Testbed for Choosing a Digital Preservation Strategy*, Proceedings of the 9th International Conference on Asian Digital Libraries, Springer, pp. 323-332, 2006.
- [Stro07-1] Strodl, S., Becker, C., Neumayer, R., Rauber, A., Bettelli, E.N., Kaiser, M., Hofman, H., Neuroth, H., Strathmann, S., Debole, F., Amato, G., *Evaluating Preservation Strategies for electronic Theses and Dissertations, Proceedings of the DELOS Conference on digital Libraries, Tirrenia, Italy*, bebruary 13-14, 2007.
- [Stro07-2] Strodl, S., Becker, C., Neumayer, R., Rauber, A., How to Choose a Digitale Preservation Strategy: Evaluating a Preservation Planning Procedure tijdens JCDL'07 june 18-23, pp. 29-38, 2007.
- [Task96] Task Force on Archiving of Digital Information, *Preserving digital information: report of the Task Force on Archiving of Digital Information*, Washington, DC: Commission on Preservation and Access, Nr.: 59, 1996.
- [Test03] Testbed Digitale Bewaring, Van digitale vluchtigheid naar digitale houvast: bewaren van tekstdocumenten, versie 1.1, Den Haag, 2003.
- [Test03-2] Testbed Digitale Bewaring, White Paper Emulatie: context en huidige stand van zaken, Den Haag, 2003.
- [Test05] Testbed Digitale Bewaring, *Functionele eisen bewaarsysteem*, versie 1.1, Den Haag, 2005.
- [Thib02] Thibodeau, K., Overview of Technological Approaches to Digital Preservation and challenges in coming years. In: The State of Digital Preservation, CLIR, 2002.
- [Upwa94] Upward, F., *Search of the Continuum in The Records Continuum*, National Archives of Australia, online herdruk, 1994.

[Wijn04] Wijngaarden, H.N., Oltmans, E., *Digital Preservation and Permanent Access: the UVC for Images*, Koninklijke Bibliotheek, Den Haag, 2004.

Websites

- [1] http://www.theregister.com/2006/07/06/microsoft_odf_interoperability, 06-07-2006, bezocht op 26-06-2007
- [2] http://www.regdeveloper.co.uk/2006/08/13/odf_debate, 13-08-2006, bezocht op 26-06-2007
- [3] http://www.regdeveloper.co.uk/2007/05/12/openxml_windup, 12-05-2007, bezocht op 26-06-2007
- [4] http://www.theregister.com/2007/05/16/norway_endorses_odf, 16-05-2007, bezocht op 26-06-2007
- [5] <http://www.opensourcnieuws.nl/index.php/content/view/3952/57>, 13-03-2007, bezocht op 26-06-2007
- [6] <http://blogs.microsoft.nl/raulpesch/archive/2006/06/19/1807.aspx>, 19-07-2006, bezocht op 26-06-2007
- [7] http://en.wikipedia.org/wiki/List_of_file_formats, 29-06-2007, bezocht op 04-07-2007
- [8] http://www.kb.nl/hrd/dd/dd_onderzoek/vp.gif, bezocht op 04-07-2007
- [9] <http://www.iana.org/assignments/media-types>, bezocht op 26-07-2007
- [10] http://www.justid.nl/producten_en_diensten, bezocht op 31-07-2007
- [11] <http://www.dpconline.org/graphics/events/0706jpeg2000wkshop.html>, bezocht op 07-08-2007
- [12] http://www.planets-project.eu/docs/comms/Planets_Project_Brochure.pdf, bezocht op 07-08-2007
- [13] <http://www.digicult.info/pages/techwatch.php>, bezocht op 09-08-2007
- [14] <http://www.dpconline.org/graphics/reports>, bezocht op 09-08-2007
- [15] <http://droid.sourceforge.net/wiki/index.php/Introduction>, bezocht op 23-04-2007
- [16] <http://hul.harvard.edu/jhove>, bezocht op 27-04-2007
- [17] <http://meta-extractor.sourceforge.net>, bezocht op 02-05-2007
- [18] <http://www.nationalarchives.gov.uk/pronom>, bezocht op 23-04-2007
- [19] <http://www.digitalpreservation.gov/formats>, bezocht op 09-08-2007
- [20] <http://hul.harvard.edu/gdfr>, bezocht op 01-05-2007
- [21] <http://dev.dcc.ac.uk/twiki/bin/view/Main/DCCRegRepV04>, bezocht op 16-04-2007
- [22] <http://www.videolan.org/vlc/features.html>, bezocht op 13-08-2007
- [23] <http://dioscuri.sourceforge.net>, bezocht op 13-08-2007
- [24] <http://xena.sourceforge.net>, bezocht op 02-05-2007

Bijlage 1 Verklarende woordenlijst

Begrip	Omschrijving
AIP	Archival Information Package. De wijze waarop de informatie is georganiseerd gedurende de opslag in een digitaal depot.
Archiefsysteem	Het geheel van gearchiveerde informatie. Substelsysteem van archiveringsstelsysteem. ISO-NEN 15489:2001 heeft dit begrip vertaald vanuit Record system.
Archiefstukken, -bescheiden	Informatie opgemaakt, ontvangen en onderhouden als bewijs en informatie door een organisatie of persoon bij het vervullen van wettelijke verplichtingen of bij zakelijke transacties.
Architectuur (Informatiesysteem-)	Een consistent geheel van principes, methoden en modellen dat gebruikt wordt om het landschap van en onderlinge samenhang tussen informatiesystemen aan te geven.
Archiveringsstelsysteem	Het geheel van procedures, methoden, kennis, mensen, middelen en documenten, waarmee een persoon of samenwerkingsverband van personen (organisatie) de archiveringsfunctie vorm geeft. [Hors04]. ISO-NEN 15489-1:2001 is de algemene norm voor een dergelijk stelsysteem, waarin het Recordkeeping system wordt genoemd.
Authenticiteit	De mate van gelijkheid aan het origineel, de 'look and feel'.
Bestandsformaat	Een opeenvolging van in bits (enen en nullen) gecodeerde informatie (bitstream) in een digitaal bestand.
Bewaarexperiment	Het testen van bewaarstrategieën door deze uit te voeren in een beheersbare en gedocumenteerde omgeving waarbinnen criteria gemeten kan worden.
Bewaarstrategie	Manieren om (digitale) informatie goed te bewaren en te beheren. Zie ook emulatie en migratie.
CDD+	Centraal Digitaal Depot, waarbij de plus staat voor de organisatie rondom het CDD.
Codering/Bitstream	Specifieke volgorde van bits (enen en nullen) die samen één gecodeerd informatie element beschrijven.
Conversie	Vorm van migratie, waarbij de oorspronkelijke bestandsformaten worden omgezet naar alternatieve bestandsformaten.
Creatiecontext	De omgeving (context) waarin informatie wordt opgesteld.
Determineren	Bepalen van eigenschappen.
DIP	Dissemination Information Package. De wijze waarop de informatie is georganiseerd tijdens verspreiding naar raadplegers.
DRM	Digital Rights Management. Een techniek om het beheer van digitale rechten van makers of uitgevers van werken (b.v. muziek, afbeeldingen, teksten) te vereenvoudigen.

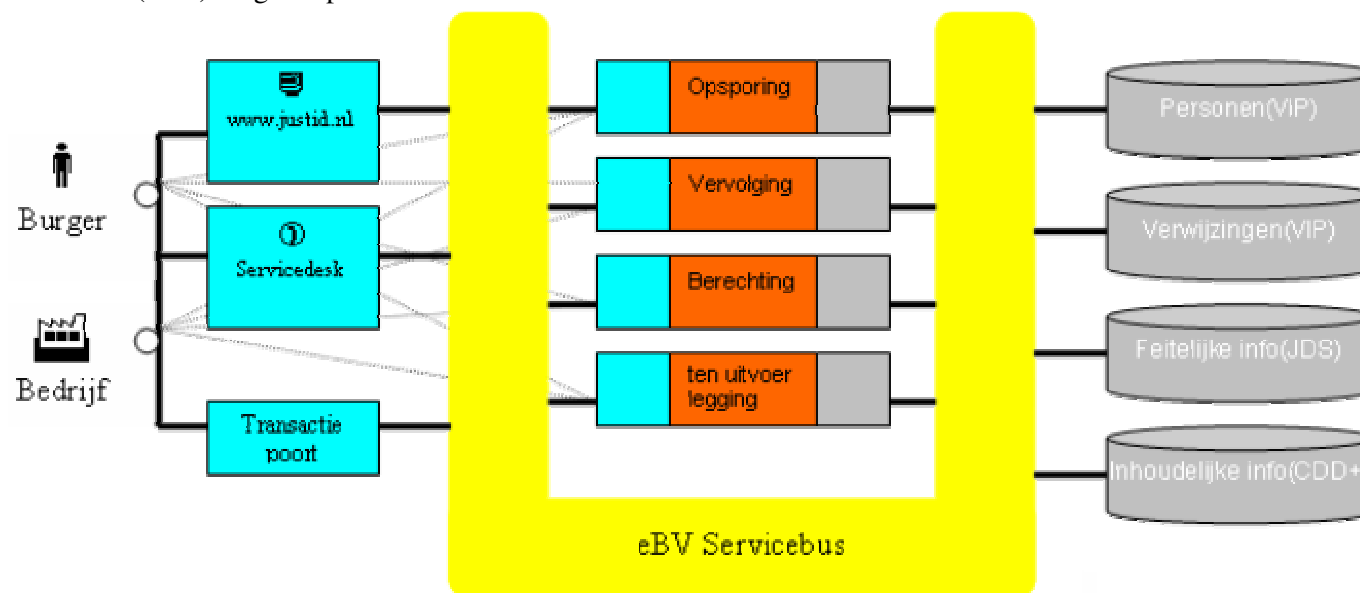
EBV	Elektronisch Berichten Verkeer. Berichten uitwisselingsprotocol binnen de Strafrechtsketen.
Emulatie	Aanpassing van de huidige omgeving (hard-/software) naar de oorspronkelijke omgeving waarin de digitale informatie was opgesteld, zodat de oorspronkelijke informatie weer kan worden gegeven.
Emulatiepaden	De opeenvolging van mogelijke emulaties tot het komen bij de gewenste emulatie.
Gebruikerscontext	De omgeving (context) waarin informatie wordt gebruikt/geraadpleegd.
Informatie- /gegevensdrager	Middel tot het overdragen van informatie.
Informatieobject	Informatie samen met alle metadata.
Informatiepakketten	Conceptuele containers met informatie over de inhoud van een archiefstuk en de informatie over hoe deze inhoud te representeren.
Informatiesysteem	Systeem voor informatieverzameling, -verwerking en -overdracht .
Informatietype	Classificatie van soorten informatie.
Informatievoorziening	Het geheel van mensen, middelen en maatregelen, gericht op de informatiebehoefte van de organisatie.
Integer persoons- /zaakbeeld	Persoons/zaakbeeld waarbij er zekerheid bestaat dat de persoon of zaak waar het over gaat de juiste is.
Integraal persoons- /zaakbeeld	Persoons/zaakbeeld waarbij er zo veel mogelijk relevante informatie over de persoon of zaak beschikbaar wordt gesteld.
JDS	Justieel Documentatie Systeem. Informatiesysteem van JustID informatie feitelijkheden over natuurlijke- en rechtspersonen bevat die met Justitie in aanraking zijn gekomen.
Justitiabele	Natuurlijke- of rechtspersoon (of samenhang van personen) onderhevig aan rechtsspraak.
Ketendossier	Overzichtelijke weergave van de voor de ketenpartner relevante archiefstukken, die al dan niet afkomstig zijn van andere ketenpartners.
Ketenpartners	Instanties die opereren binnen de Strafrechtsketen.
Ketens	Een netwerk van grote aantallen onafhankelijke organisaties en professionals die (in constant wisselende combinaties) samenwerken om een (maatschappelijk) product voort te brengen.
Koppelvlak	(Ook wel interface) Het raakvlak tussen twee organisaties of componenten aan.
LDB	Logische Data Beschrijving. Een taal die de inhoud van informatie beschrijft aan de hand van labels, zoals in XML.
Manifestatie	Een (alternatieve) codering voor dezelfde informatie.

Metadata	Gegevens over gegevens. Zoals gegevens over de te archiveren informatie.
Metadatamodel	Soort datamodel dat de entiteiten beschrijft waarmee metadata gearchiveerd wordt.
Migratie	Aanpassing van de manier waarop informatie digitaal gecodeerd wordt aan nieuwe hard- en/of software of aan een nieuwe generatie computer technologie.
Migratiepaden	De opeenvolging van migraties tot het komen tot de gewenste migratie.
Mindmap	Grafische weergave (of informatieboom) vanuit een centraal onderwerp dat daaraan bijzaken en verwante concepten linkt.
Normalisatie	De migratie van bestandsformaten naar open standaarden.
OAIS	Het Open Archive Information System referentie model [Cons02].
PDF/A	Bestandsformaat gericht op het duurzaam bewaren van digitale documenten. De variant 1b legt alleen de verschijningsvorm vast (het uiterlijk).
Register	Centrale plaats voor de opslag van metagegevens.
Repository	Centrale plaats voor de opslag van gegevens.
Selectielijst	Een selectielijst regelt welke categorieën archiefstukken voor altijd bewaard blijven, en welke categorieën op termijn vernietigd moeten worden. Aan de te vernietigen stukken kent de lijst bovendien een vernietigingstermijn toe.
SIP	Submission Information Package. De wijze waarop de informatie is georganiseerd tijdens de inname bij het archief.
Structuur	Het logische verband tussen de elementen van een document of van een archief (Regeling geordende en toegankelijke staat archiefbescheiden, art 1, lid m).
Substitutie / vervanging	De bevoegdheid archiefbescheiden te vervangen door reproducties, waarna de vervangen en worden vernietigd.
Utiliteitsanalyse	(Utility analysis) Kwantitatieve methode om alternatieven te vergelijken op te vergelijken criteria.
UVC	Universal Virtual Computer. Een computer die uit virtuele hardware bestaat.
Viewpath	Set van op elkaar voortbouwende lagen die nodig zijn om digitale informatie op te stellen, dan wel te representeren.
VIP	VerwijsIndex Personen. Het VIP is een verwijsindex voor personen. In VIP kan worden gezien welke ketenpartners binnen Justitie met een justitiabele bezig zijn. Daarnaast bevat het verwijzingen naar andere systemen en locaties.
Zorgdrager	Degene die bij of krachtens de wet belast is met de zorg voor de archiefbescheiden (Archiefwet 1995, art.1d).
XML	Extensible Markup Language. Een opmaaktaal op basis van tekst (Unicode ISO/IEC 10646) voor het beschrijven van de structuur en de betekenis van gegevens.

Bijlage 2 Basisvoorzieningen Strafrechtsketen

JustID biedt de Strafrechtsketen een vijftal basisvoorzieningen aan of is bezig deze te realiseren. Het hierbij om de volgende systemen (zie ook figuur 26):

- Verwijs Index Personen (VIP) (Personen): registratie van natuurlijke- en rechtspersonen die op enigerlei wijze met Justitie in aanraking zijn gekomen (justitiabelen);
- Verwijs Index Personen (VIP) (Verwijzingen): registratie van welke partijen in de Strafrechtsketen op welk tijdstip bezig zijn met welke justitiabelen;
- Justitieel Documentatie Systeem (JDS): registratie van strafrechtelijke gegevens over justitiabelen die op enigerlei wijze met Justitie in aanraking zijn gekomen;
- Centraal Digitaal Depot (CDD): registratie van digitale archieven rondom justitiabelen die op enigerlei wijze met Justitie in aanraking zijn gekomen;
- Elektronisch Berichten Verkeer (eBV): Digitaal protocol om berichten in de Strafrechtsketen uit te kunnen wisselen.



Figuur 26 – Stelsel van vijf basisvoorzieningen van JustID. Bron: JustID

Bijlage 3 Informatietypen

Digitale informatie is op verschillende manieren te coderen. Het is belangrijk te weten welke manieren dit zijn, wat de eigenschappen daarvan zijn en wat er hiervan bewaard moet worden. Op de eigenschappen van de verschillende informatietypen gaan onderstaande paragrafen in, vanuit een onderverdeling naar samengestelde formaten (complexe digitale objecten) en eenvoudige formaten (simpele digitale objecten) [7][9][Boud02][Barv07][Knig07]; welke hiervan bewaard moeten worden is enerzijds onderdeel van een proces dat samen moet worden uitgevoerd met de ketenpartners en anderzijds een rol voor wet- en regelgeving. Deze laatste stelt enkele bestandsformaten voor vanuit de Archiefwet 1995. Deze bestandsformaten worden waar mogelijk ook in onderstaande paragrafen besproken.

Enkelvoudige informatietypen

Enkelvoudige informatietypen, ook wel simpele objecten, bestaan veelal uit slechts één object dat gebruikt wordt binnen andere objecten, de zogenaamde samengestelde informatietypen, ook wel complexe digitale objecten. Zo komen tekstuele en statische grafische weergaven veelvuldig voor in documenten en worden audio en bewegende beelden veelal samen gepresenteerd. Onderstaande enkelvoudige informatietypen worden geïdentificeerd.

Tekstueel

Een digitaal bestand dat alleen tekst bevat, zonder al te veel opmaak, wordt gezien als tekstueel. Het informatietype tekstueel is een basis voor de beschrijving van vele andere formaten. De wet- en regelgeving rondom archiveren schrijft voor tekensets het gebruik van ASCII (ISO/IEC 8859-1) of Unicode (ISO/IEC 10646-1) voor.

Audio

Bestanden van het informatietype audio kunnen worden gebruikt voor het opslaan van audio, zoals de opname van het verhoor van een verdachte of opgenomen geluid van een beveiligingscamera. Technisch gezien kunnen bestanden van dit informatietype worden onderverdeeld naar bestanden die na omzetting inhoud verloren hebben, of juist geen opname verlies hebben (lossy, dan wel lossless). Verder zijn er bestanden die compressie toestaan of de mogelijkheid bieden encryptie te gebruiken, zoals dit mogelijk is bij Digital Right Management (DRM).

Bestanden van het informatietype audio zijn an sich niet als complex aan te merken. Ze worden wel complexer door het gebruik van compressie en encryptie.

Visueel

Bestanden van het informatietype visueel kunnen stilstaande beelden bevatten of bewegende beelden, zoals een eenvoudige afbeelding of de beelden van beveiligingscamera of mobile telefoon. Op twee manieren kunnen visuele representaties op worden geslagen. Dit kan gedaan worden in vectoren of in een raster van pixels. De meeste gebruikte formaten voor statisch visuele opslag zijn JPEG, GIF, TIFF en PNG. Deze en anderen bestandsformaten van het informatietype visueel verschillen in kwaliteit en enkele andere eigenschappen, zoals het kunnen representeren van animaties (GIF), weergave van doorschijnende vlakken of gehanteerde compressie.

De meeste gebruikte formaten van het informatietype visueel die bewegende beelden kunnen opslaan zijn MPEG, AVI, Flash, MOV. Ook deze formaten verschillen in kwaliteit en enkele andere

eigenschappen, zoals de mogelijkheid tot interactie, aantal frames per seconde, aantal key-frames en gehanteerde compressie of gebruik van encryptie.

De wet- en regelgeving rondom het archiveren schrijft de opslag in PDF voor wanneer het gaat om afbeeldingen, en indien hierbij gebruik wordt gemaakt van compressie schrijft het ITU T4 of ITU T6 voor. Er wordt niet ingegaan op de opslag van bewegende beelden in digitale vorm.

Bestanden van het informatietype visueel zijn over het algemeen niet complex, maar kunnen wel complexer worden naar mate er gedrag/interactie wordt opgeslagen, zoals mogelijk is in een Flash object. Ook maakt het gebruik van bijvoorbeeld compressie en encryptie de formaten veel complexer.

Samengestelde informatietypen

De zogenaamde samengestelde informatietypen, ook wel complexe digitale objecten, maken gebruik van de enkelvoudige informatietypen. Zo kan het voor komen dat de informatietypen waaruit digitale informatie is samengesteld een verschillende ‘levensverwachting’ hebben, en dat er daarmee verschillende bewaarstrategieën zouden moeten worden toegepast op het samengestelde informatietype. Onderstaande samengestelde informatietypen kunnen worden geïdentificeerd.

Inpak informatietype

Bestandsformaten van het inpak informatietype worden gebruikt voor de overdracht van digitale bestanden of het aan elkaar binden hiervan. Zij kunnen ook gebruikt worden voor het beveiligen van de ingepakte digitale bestanden. Het meest gebruikte inpak bestandsformaat is ZIP, maar ook email kan worden gezien als gebonden/ingepakte informatie in een MSG formaat. Samenhangende informatie kan gebonden worden via een inpakformaat.

Een inpakapplicatie kan één of meerdere digitale bestanden inpakken tot één of enkele opeenvolgende bestanden. Er kan compressie worden toegepast waardoor de som van de bits waarin de digitale informatie is gecodeerd kleiner is dan dat van het origineel. Van compressie wordt verwacht dat deze lossless is, wat inhoudt dat de originele bestanden weer kunnen worden verkregen zonder verlies van informatie door het inpak informatietype uit te pakken met een daarvoor ontwikkelde uitpakapplicatie.

Naast de mogelijkheid tot compressie, kan er encryptie op het ingepakte worden toegepast. Dit maakt het voor onbevoegden lastig tot onmogelijk de digitale informatie te kunnen raadplegen.

Het inpak informatietype kan worden gezien als een complex informatieobject vanwege de alternatieve manieren van opslag van bestanden: bestanden worden uit de originele context gehaald en geplaatst in het inpak informatietype. Compressie is een complex proces en encryptie bemoeilijkt de interpretatie. Wanneer er echter geen compressie of beveiliging wordt toegepast en het inpak informatietype goed gedocumenteerd is, kan deze worden gezien als niet complex. Dan dient het inpak informatietype bijvoorbeeld alleen een overdrachts- of bindingsnut.

Database informatietype

Databases worden gebruikt om grote hoeveelheden gestructureerde informatie te bewerken en beheren. Daarvoor bestaat een database uit de database zelf (inhoud), het databasemanagement systeem (DBMS) en de databaseapplicaties en toepassingen. Deze laatste twee stellen eindgebruikers in staat om mutaties in de database handmatig of geautomatiseerd door te voeren en gegevens te bekijken. Het DBMS op zijn beurt is de laag tussen gebruiker en fysieke database. Het handelt de verzoeken af en voert ze daadwerkelijk door in de databasebestanden.

Er zijn verschillende gegevensmodellen waarop de DBMS'en zijn gebaseerd. Het relationele

gegevensmodel domineert momenteel de markt [Test03]. Deze bestaat uit tabellen, die bestaan uit kolomnamen en veldinhoud. Kolomnamen kunnen bijvoorbeeld een sleutel zijn en zodanig refereren naar andere sleutelkolomnamen in andere tabellen. Op deze manier kunnen relaties worden gelegd tussen gegevens. Een speciale taal, Standard Query Language (SQL), kan worden gebruikt om de database te bevragen en om gebruikelijke bevestigingen te programmeren.

Bij het digitaal duurzaam houden van databases kan er niet generiek gesteld worden wat er relevant is om op te slaan. Zo kan het relevant zijn om het complete databasesysteem op te slaan, alleen de database of een database-dump (inhoud), een enkele rij uit een tabel of tabellen of de gegevens zoals deze op het scherm van de gebruiker worden getoond door een (web)applicatie.

Naast de vraag wat als relevant wordt gezien om op te slaan verschillen de databasesystemen in complexiteit. Zo worden Oracle-producten, SQL Server en DB2 DBMS'en als zeer complex gezien en zijn MS Access en dBase IV minder complex [Test03].

De wet- en regelgeving rondom het archiveren van databases schrijft de opslag in de oorspronkelijke bestandsformaten voor, of in ASCII (tekstueel bestandsformaat), met veldscheidingstekens, vergezeld van documentatie over de structuur van de database (XML-DTD) en tenminste het logisch datamodel met beschrijving van de entiteiten; queries dienen opgeslagen te worden in de vraagtaal SQL (SQL2).

Document informatietype

Veruit de meeste digitale informatie bij de overheid wordt opgeslagen in een document informatietype [Test03]. De informatie ontstaat uit werkprocessen en representeert bijvoorbeeld een beleidsnota, voortgangsrapportages of een projectbeschrijving. De bestanden worden veelal opgesteld in grafische tekstverwerkers en maken gebruik van een WYSIWYG (What You See Is What You Get) editor. Deze zorgen er voor dat de gebruiker gemakkelijk kan opstellen wat zij wil en dat zij ook direct ziet hoe de informatie wordt weergegeven. Dit maakt het digitale object echter complex. Er kan van alles in het document informatietype bestand worden gezet: tabellen, formulieren, koppen en paragrafen, secties, opsommingstekens, voet- en eindnoten, tekens en symbolen, velden en kaders, automatische waarden, grafische objecten (afbeeldingen), commentaar en wijzigingen en zelfs gekoppelde of geïntegreerde objecten. Daarnaast kan er bij bepaalde document informatietypen ook een digitale handtekening worden toegevoegd of kan encryptie en wachtwoordbeveiliging worden toegepast.

De mate van opmaak en toevoeging van complexe elementen, maken het uiteindelijke digitale bestand complexer. Het document informatietype kan dus eenvoudig in elkaar zitten, maar kan ook erg complex zijn.

De wet- en regelgeving rondom het archiveren van documenten schrijft voor dat er wordt opgeslagen in PDF of SGML, dan wel in XML vergezeld van een stylesheet (XSL, CSS), dan wel TIFF of PDF met de metadata in een XML-wrapper.

Script, bron-, objectcode, executables & codebibliotheek informatietype

Applicaties bestaan uit een of meerdere bestanden die gebruik maken van object bibliotheken en andere referenties. Zij hebben meestal een informatie verwerkende of representerende functie. Vaak bestaan applicaties uit een gebruikersinterface met backend en zijn uit te voeren op een combinatie van computerarchitectuur en besturingssysteem. Het kan ook gaan om achtergrond processen die een informatieverwerkende functie hebben.

Applicaties kunnen daarnaast uit leesbare (tekstuele) broncode bestaan wanneer deze nog niet

gecompileerd is en nog geïnterpreteerd moet worden tijdens het uitvoeren. Dit maakt het mogelijk middels het tekstuele informatietype applicaties te beschrijven. Het merendeel van de applicatie informatietypen zal echter niet langer uit broncode bestaan (al gecompileerd), en is daarmee te classificeren als complex. De complexiteit is hiermee beschreven door de afhankelijkheid van het al dan niet gecompileerd zijn en het gebruik van referenties naar andere objecten.

Pagina beschrijvend informatietype

Dit informatietype beschrijft documenten en grafische weergaven op een statische manier waardoor de bestandsformaten onder verschillende besturingssystemen en printers hetzelfde worden weergegeven. De meest bekende zijn PDF, PostScript en CSS. Enkele zijn gesloten, en gebruiken een specifieke applicatie die de bestanden kan inlezen en kan weergeven op het computerscherm of kan printen via printerdrivers. Steeds vaker wordt er echter een beschrijvende taal gebruikt, zoals XML, en worden onderdelen binnen het pagina beschrijvend informatietype ingepakt tot één bestand of mappen structuur. Een voorbeeld hiervan is PDF/A. De statische opslag en het gebruik van meer en meer open bestandsformaten maakt het dat het pagina beschrijvend informatietype als niet complex kan worden beschouwd.

Presentatie informatietype

Voor het presenteren van informatie wordt steeds vaker de computer als hulpmiddel gebruikt. De meest bekende presentatieprogramma's zijn MS PowerPoint en Adobe PDF. Deze laatste is tevens een pagina beschrijvend informatietype en daarmee minder complex dan het MS PowerPoint bestandsformaat. MS PowerPoint kan complexer zijn dan bestanden van het document informatietype. Hierbij gaat het om de informatie eigenschap gedrag dat in hogere mate kan worden toegevoegd in bestanden van het presentatie informatietype.

Spreadsheet informatietype

Bestandsformaten van het spreadsheet informatietype worden gebruikt om makkelijk opstapelende berekeningen uit te kunnen voeren en de resultaten grafisch te presenteren. Spreadsheets worden veel gebruikt door financiële- en productiebedrijven. De applicaties zorgen er voor dat de resultaten van achterliggende berekeningen te zien zijn en dat ook de achterliggende berekening getoond kan worden. De spreadsheet applicatie zorgt dus voor een bepaald gedrag, iets dat niet op eenzelfde interactieve manier op papier is vast te leggen. Daarnaast worden de grafische presentaties van de gegevens opgesteld uit de gegevens, bijvoorbeeld uit rekenresultaten. Dit kan worden gezien als een afbeelding met logica er achter. Vanwege de logica en berekeningen in een spreadsheet, zijn bestanden van het spreadsheet informatietype complex. Het is daarom van belang te weten wat er bewaard moet worden. Zonder het gedrag is een spreadsheet niets anders dan een raster met afbeeldingen, tekst en cijfers: een statische visuele weergave.

Webpagina/opmaak informatietype

Websites maken gebruik van bestandsformaten die te schalen zijn onder het webpagina/opmaak informatietype. De meest gebruikte is HTML, een taal die te vergelijken is met de broncode van applicaties en net als het document informatietype complexer is naar mate er naar meerdere informatie objecten wordt gerefereerd. Er wordt bijvoorbeeld gerefereerd naar een bestand van het visuele informatietype, externe database of een flash of java element. Er kan daarnaast een complex gedrag opgeslagen zijn waarbij de inhoud afhankelijk is van zoektermen, gebruikers login of een IP-adres

waarvan verbonden uitgevoerd.

Naast dat websites gebruik maken van referenties naar elkaar, zijn er dus tal van complexe elementen die niet volledig kunnen worden vastgelegd zonder dat de originele broncode of applicatie beschikbaar is en kan worden gedraaid.

Lettertype informatietype

Verschillende digitale informatietypen gebruiken in hun informatie tekens, zoals letters om woorden te vormen. Deze worden vorm gegeven in een lettertype. Dit lettertype informatietype wordt op zijn beurt opgeslagen in het informatie object of er wordt naar een extern bestand van het lettertype informatietype gerefereerd. Daar waar het lettertype wordt gebruikt in een applicatie, kan gerefereerd zijn naar algemene lettertypen van het besturingssysteem of naar lettertypen die alleen binnen de applicatie beschikbaar zijn.

Expertise informatietype

Specialistische applicaties gebruiken veelal zelf ingerichte bestandsformaten voor bijvoorbeeld het opslaan van expertinformatie, zoals wetenschappelijke meetresultaten, signaal modellering, boekhouding of technische tekeningen. De bestandsformaten van het expertise informatietype zijn niet noodzakelijk complex, maar de interpretatie er van kan wel uiterst complex zijn. Meet- en onderzoeksresultaten kunnen bijvoorbeeld in een tekstbestand worden geplaatst, maar het kan zijn dat de interpretatie ervan gedaan moet worden via applicaties. Zo kan een applicatie voor het werken met statistische gegevens, zoals SPSS, de interpretatieslag maken van datasets naar te begrijpen gegevens. Technische tekeningen kunnen bestaan uit 2D weergaven die samen een interactief 3D model vormen.

De wet- en regelgeving rondom archiveren schrijft voor dat technische tekeningen moeten worden opgeslagen in PDF en STEP (Standard for the exchange of product data, ISO 10303). Voordat er kan worden besloten op welke manier er moet worden opgeslagen/bewaard, is het dus van belang te weten wat van de digitale informatie bewaard moet blijven.

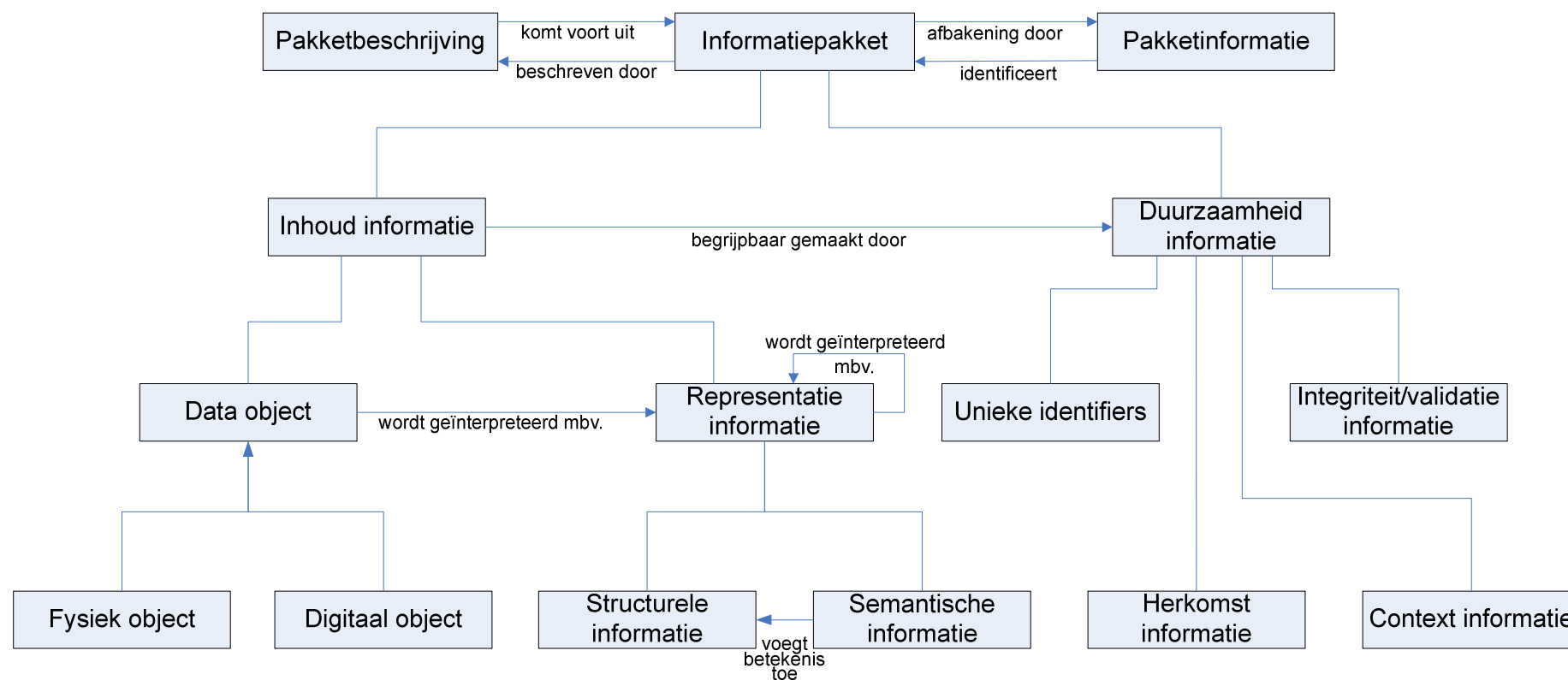
Bijlage 4 Entiteiten metadatamodel CDD+

Onderstaand wordt het metadatamodel van het CDD+ besproken op hoofdlijnen [Delo06]. Hierbij legt de laatste kolom de link met het eigenlijke datamodel.

Onderwerp	Toelichting	Entiteiten
Archief en archiefregime	In het CDD+-datamodel is een onderscheid gemaakt tussen een Archiefregime (het stelsel van regels rondom een bepaald soort archief) en een Archief. De verantwoordelijkheid van bijvoorbeeld een Rechtbankpresident (zorgdrager voor archief arrondissement X) wordt hiermee goed onderscheiden van de algemene richtlijnen voor de archieven van de rechterlijke organisatie. Deze laatste liggen voor alle rechtbanken vast.	Archief Archiefregime
Archiefrechterlijke structuur	De Handelingenstructuur implementeert de door het Nationaal Archief geïdentificeerde archiefstructuur van een bepaald werkterrein van de overheid. De beschrijving hiervan wordt vastgelegd in een Rapport Institutioneel Onderzoek. De gegevens uit een dergelijk onderzoek worden gestructureerd vastgelegd in de entiteiten Handeling, Mandaat, Actor soort, Bron en Procedure. Het vaststellen van de bewaartermijnen van producten uit een beleidsterrein gebeurt via basisselectiedocumenten (BSD's). Deze documenten worden in het datamodel per actor gerepresenteerd in de entiteiten Selectielijst en Selectieregel.	Handeling Mandaat Actor soort Bron Procedure Selectielijst Selectieregel
Overige structuren: classificaties en dossierstructuur	Classificaties zijn een tweede mechanisme om documenten te categoriseren. Dit mechanisme is in te richten per archief. Het kan gebruikt worden om een ordening aan te geven voor eigen gebruik. Door het onderkennen van een synoniemrelatie binnen rubrieken, wordt tevens implementatie van een thesaurus mogelijk gemaakt. Classificaties zijn geïmplementeerd met de entiteiten Classificatie en Rubriek. Dossierstructuur wordt opnieuw bij het archiefregime vastgelegd. In de rechterlijke organisatie betekent dit: landelijk beheerd. Als voorbeeld hiervan kan de landelijk uniforme dossierstructuur van GPS gelden. De dossierstructuur is geïmplementeerd in de entiteiten Bestanddeel en Bestanddeel soort. Binnen een Archief kunnen subdossiers gedefinieerd worden binnen bestaande dossiers.	Classificatie Rubriek Bestanddeel Bestanddeel soort

Betrokkenen	Betrokken personen en organisatie rondom het dossier vallen uiteen in twee groepen. Allereerst zijn er de personen en instantie binnen de ketenpartners die werken aan het archief. Deze organisaties zijn vormgegeven middels de entiteiten Actor, Functionaris, Instantie en Naamalternatieven. Daarnaast zijn er de externe organisatie (bijvoorbeeld procespartijen) die verbonden zijn aan een bepaald dossier. De externe betrokkenen liggen vast als Relatie, onderverdeeld in Natuurlijke persoon en Rechtspersoon. Van beide groepen kunnen adressen vastgelegd worden. Daarnaast kunnen beide groepen op verschillende manieren betrokken worden bij een document. Deze betrokkenheid ligt vast via de entiteiten Rol en Rol soort	Relatie Natuurlijk persoon Rechtspersoon Actor Functionaris Instantie Rol Rol soort Adres
Beheer	De mogelijke handelingen die binnen een archiefregime kunnen worden uitgevoerd, liggen vast in Beheershandeling soort. Het uitvoeren van de handelingen wordt ondersteund met Beheershandeling en Selectie. Dit laatste bundelt de betrokken documenten bij een bepaalde handeling. De entiteit Autorisatie legt vast wie welke handelingen mag uitvoeren binnen een bepaald archief. Daarbij kan nog nader onderscheid gemaakt worden met behulp van de drie bovengenoemde ordeningsstructuren.	Beheershandeling Beheershandeling soort Selectie Autorisatie
Bestandsformaten	Via bestandsformaten ligt vast wat de opmaak van een elektronisch document is. Middels Toegestane formaten zijn de mogelijke Bestandsformaten binnen een Archiefregime te beperken. Met Conversietools is aan te geven hoe een document van een niet geaccepteerd formaat geconverteerd wordt naar een wel geaccepteerd formaat.	Toegestaan formaat Bestandsformaat Conversietool
Fysieke verblijfplaatsen	Het beheer van fysieke delen van een archief is ondersteund met de entiteiten Verblijf en Verblijfplaats.	Fysieke verblijfplaats Verblijf

Bijlage 5 Informatie pakket / Archival information package (AIP)



Bijlage 6 Bestandsformaat herkenning en inhoudbepaling

Notitie Technische metadata CDD+

14-05-2007 - Rob Berentsen

Samenvatting en conclusie

Het proces van digitale bewaring start met de juiste bepaling van het bestandsformaat. Op het punt van de verkregen technische metadata kan dit proces binnen het CDD nog verbeterd worden. Deze verbetering is van direct belang voor toekomstige vervolgprocessen als migratie/conversie en emulatie.

Er spelen een tweetal problemen, voortvloeiend uit het feit dat het CDD+ zelf geen bestandsformaat kan bepalen:

- Het CDD+ kan zelf geen onderscheid maken tussen een normale PDF en de archiveringsvariant PDF/A. Bestanden worden daarom door een extern conversie systeem gestuurd (AdLib). Wanneer er echter een veelvoud van of zeer grote bestanden aan dit systeem worden gestuurd, leidt dit tot hoge systeembelasting;
- Bij 'digital born' documenten (originelen) die naast de PDF/A (representatieversie) dienen te worden opgeslagen, is het van belang technische metadata op te slaan. AdLib biedt als extern conversie systeem hiervoor niet de juist gegevens.

Deze notitie neemt software die geautomatiseerde bestandsformaat bepaling uitvoert onder de loep en raad een product aan om te gebruiken tijdens de invoer in het CDD: DROID v1.1.

DROID v1.1 bepaald als beste het bestandsformaat van de testbestanden en geeft deze tevens een unieke referentiecode naar een online bestandsformaat register (PRONOM), waarin meer relevante informatie te vinden is over het bestandsformaat in het kader van digitale duurzaamheid.

Gebruik van DROID v1.1 zal het externe conversie systeem (AdLib) ontlasten en de bestandsformaat bepaling van 'digital born' bestanden mogelijk maken.

DROID biedt de eerste stap richting classificatie van digitale bestanden. In een later stadium wordt het bepalen van de interne elementen van digitale bestanden belangrijk. Dit wordt gebruikt om te komen tot een geschikte bewaarstrategie per digitaal informatieobject.

Inleiding

Met het oog op digitale duurzaamheid van de archiefobjecten is het van belang dat er exact bekend is wat er in het archief staat. Alleen dan kunnen goede vervolgprocessen zoals migratie/conversie of emulatie worden uitgevoerd (bewaarstrategie). De technische metadata wordt gezien als een grondbeginsel voor toekomstige interpretatie van archiefobjecten. Momenteel wordt er in het CDD+ wel technische metadata toegevoegd, maar dient het bestand eerst te worden aangeboden aan een externe server waarop AdLib draait (conversieserver). Door AdLib worden de bestanden uit voorzorg geconverteerd om er zeker van te zijn dat het CDD+ een PDF/A krijgt aangeboden en niet een normale PDF. Het CDD+ zelf kan momenteel niet dit onderscheid maken, noch andere bestandsformaten voldoende bepalen. AdLib biedt alleen een deeloplossing door, onder het risico van hoge systeembelasting, te verzekeren dat het CDD+ een PDF/A krijgt. AdLib bepaald niet de bestandsformaten van 'digital born' originelen: het is conversiesoftware.

Er is specifieke software beschikbaar die ingezet kan worden om bestandsformaten te bepalen. Deze notitie neemt de beschikbare software, ontwikkeld door universiteiten en archiefinstanties, onder de loep. Enkele criteria waaraan alle producten dienen te voldoen zijn:

- ontwikkeld in JAVA;
- te benaderen via de commandline;
- kan overweg met XML voor zowel in- al output.

Overzicht tools

DROID - <http://droid.sourceforge.net/wiki/index.php/Introduction>

DROID staat voor Digital Record Object Identification en is een initiatief van The National Archives (UK). DROID werkt samen met FileFormats database PRONOM om bestanden te kunnen identificeren aan de hand van enkele verschillende technieken. De gegenereerde output is XML. Binnenkort komt een nieuwe versie uit (v2.0) die meer technieken zal bevatten om een betere/diepere herkenning te realiseren.

Versie: 1.1	Datum: 08-2006
-------------	----------------

JHOVE - <http://hul.harvard.edu/jhove>

JHOVE staat voor “JSTOR/Harvard Object Validation Environment” en is een initiatief van JSTORE en Harvard University. Deze tool richt zich niet zozeer op het bepalen van bestandsformaten, maar meer op het bepalen van de interne onderdelen(!). Als output kan net als bij DROID gekozen worden voor XML. JHOVE geeft in tegenstelling tot DROID een uiteenraffeling naar de interne onderdelen van het bestand (preservation metadata extraction), dus welke fonts, images, filters, metadata-info, etc worden gebruikt en dat ook per pagina of object (images, font). Deze uiteenraffeling kan worden gebruikt bij het scoren van de bestanden om te komen tot een geschikte bewaarstrategie. Dit proces heeft niet directe prioriteit en kan later ook worden gedaan. Belangrijk is wel dat het archief zowel de migratie als emulatie bewaarstrategie kan ondersteunen. Er bestaat niet een directe koppeling met een repository als PRONOM. Er wordt gewerkt aan versie 2.

Versie: 1.1f (pre-release. Stable: v1.0 - 26-05-2005)	Datum: 08-01-2007
---	-------------------

Metadata extraction tool v3 - <http://www.natlib.govt.nz/about-us/current-initiatives/metadata-extraction-tool/> / <http://meta-extractor.sourceforge.net>

Metadata extration tool is een initiatief van National Library of New Zealand en is vergelijkbaar aan JHOVE: het bepaald de interne onderdelen.

Versie: 3	Datum: ??-??-2007 (test versie, nog niet uit)
-----------	---

XENA - <http://xena.sourceforge.net>

XENA staat voor XML Electronic Normalising of Archives en kan bestanden normaliseren en vervolgens bijvoorbeeld omzetten naar OpenDocument/XML. Het is een initiatief van het nationale archief van Australië.

Versie: 3.0	Datum: 05-06-2006
-------------	-------------------

Resultaten	DROID v1.1				JHOVE v1.1f		
Testbestanden	PUID	mimetype	format	version	mimetype	format	version
scanner PDF (1.2!)	fmt/16	application/pdf	Portable Document Format	1.2	application/pdf	PDF	1.2
PDF 1.2	fmt/95	application/pdf	Portable Document Format - Archival	1	application/pdf	PDF	1.2
PDF 1.4	fmt/18	application/pdf	Portable Document Format	1.4	application/pdf	PDF	1.4
PDF/A	fmt/95	application/pdf	Portable Document Format - Archival	1	application/pdf	PDF	1.4
MS-ODI-TIFFprint	fmt/7	image/tiff	Tagged Image File Format	3	image/tiff	TIFF	6.0
	fmt/8	image/tiff	Tagged Image File Format	4			
	fmt/9	image/tiff	Tagged Image File Format	5			
	fmt/10	image/tiff	Tagged Image File Format	6			
MS PowerPoint	fmt/126	application/vnd.ms-excel	Microsoft Powerpoint Presentation	97-2002	application/octet-stream	bytestream	
MS Excel	fmt/61	application/vnd.ms-excel	Binary Interchange File Format (BIFF) Workbook	8	application/octet-stream	bytestream	
	fmt/62	application/vnd.ms-excel	Binary Interchange File Format (BIFF) Workbook	8X			
MS Word 2007	x-fmt/263	application/zip	ZIP Format				
MS Word 2002	fmt/40	application/msword	Microsoft Word for Windows Document	97-2003	application/octet-stream	bytestream	
MS Word 97-2002en6.0-95	fmt/52	text/rtf	Rich Text Format	1.7	neemt te veel tijd in beslag (time-out) 6mb>		
	fmt/53	text/rtf	Rich Text Format	1.8			

MS Word 6.0-95	fmt/39	application/msword	Microsoft Word for Windows 6.0/95 Document	application/octet-stream	bytestream
MS Word 5.0 (mac)	fmt/64	image/vnd.dxf	Microsoft Word for Macintosh 4.0 Document	application/octet-stream	bytestream
	fmt/65	image/vnd.dxf	Microsoft Word for Macintosh 5.0 Document		
Wordperfect 5.x (win32)	x-fmt/394		WordPerfect for MS- 5.1 DOS/Windows Document	application/octet-stream	bytestream
Wordperfect 5.1 (dos)	x-fmt/394		WordPerfect for MS- 5.1 DOS/Windows Document	application/octet-stream	bytestream
Wordperfect 5.0	x-fmt/393		WordPerfect for MS-DOS 5.0 Document	application/octet-stream	bytestream
Works 2000	fmt/111		OLE2 Compound Document Format	application/octet-stream	bytestream
Works 4.0	fmt/111		OLE2 Compound Document Format	application/octet-stream	bytestream
RTF	fmt/52	text/rtf	Rich Text Format 1.7	neemt te veel tijd in beslag (time-out) 6mb>	
	fmt/53	fmt/53	Rich Text Format 1.8		
InkWriter/Note Taker	x-fmt/95		Inkwriter/Notetaker Document	application/octet-stream	bytestream
Pocket Word Document	x-fmt/94		Pocket Word Document	application/octet-stream	bytestream

Testbestanden	Metadata extraction tool v3			XENA v3.0		
	mimetype	format	version	mimetype	format	version
scanner PDF (1.2!)	application/pdf	Adobe PDF	PDF-1.2		PDF	
PDF 1.2	application/pdf	Adobe PDF	PDF-1.2		PDF	
PDF 1.4	application/pdf	Adobe PDF	PDF-1.4		PDF	
PDF/A	application/pdf	Adobe PDF	PDF-1.4		PDF	
MS-ODI-TIFFprint	image/tiff	Tagged Information File Format	6.0		PDF	
MS PowerPoint	application/powerpoint	Microsoft Power Point	6.0+		Presentation	
MS Excel	application/excel	Microsoft Excel	4.1+		Spreadsheet	
MS Word 2007	application/open-office-1.x	Open Office,			Binary	
MS Word 2002	application/msword	Microsoft Word	Microsoft Word 10.0		Word Processor	
MS Word 97-2002en6.0-95	text/ms-word				Word Processor	
MS Word 6.0-95	application/msword				Word Processor	
MS Word 5.0 (mac)	file/unknown	Microsoft Word	Microsoft Word for Windows 95		Binary	
Wordperfect 5.x (win32)	application/wordperfect	Corel WordPerfect	5.1		Word Processor	
Wordperfect 5.1 (dos)	application/wordperfect	Corel WordPerfect	5.1		Word Processor	
Wordperfect 5.0	application/wordperfect	Corel WordPerfect	5.0		Word Processor	
Works 2000	application/msworks	Quill96StoryGroupClass			Word Processor	
Works 4.0	application/msworks		4.0		Word Processor	
RTF	application/rtf				Word Processor	
InkWriter/Note Taker	file/unknown				PlainText	
Pocket Word Document	file/unknown				PlainText	

De resultaten geven een behoorlijk uiteenlopend beeld. Dit wordt onder andere veroorzaakt door het doel van de tools. Eigenlijk is DROID namelijk de enige tool die als hoofddoel heeft bestandsformaten te bepalen. De overige tools hebben meer de **inhoudelijke** bitstream bepaling als doel, dus de uiteenrafeling van de bestanden ter bepaling van deze delen (door JHOVE/Metadata extraction tool) of zelfs de normalisatie naar XML (door XENA).

Puur op het gebied van de herkenning van het bestandsformaat en de versie doet, naast DROID, de Metadata extraction tool v3 het behoorlijk goed. Deze mist echter enkele bestandsformaten en heeft geen aansluiting met PRONOM.

Hier ligt ook juist het grote voordeel van DROID boven andere software: het is verbonden met een bestandsformaten register die continu geüpdate wordt. Het geeft daarnaast dus ook aanvullende informatie, zoals de software vendor en de levensverwachting van het bestand.

De enige “fouten” van DROID zijn te vinden bij de herkenning van “MS Word 5.0 (mac)”, waarbij het een verkeerd minetype lijkt terug te geven en de Wordperfect formaten, waarbij geen mimetype wordt gevonden. Daarnaast wordt MS Word 2007 bestanden herkend als ZIP bestanden. Dit zijn ook ZIP bestanden, maar wel met een .docx extensie. Gezien de context zou deze wellicht moeten worden herkend als MS Word 2007 bestand.

Advies: DROID

DROID herkent als beste de test bestandsformaten. Bij de nog missende informatie na herkenning, geeft DROID in ieder geval het bestandsformaat een eigen PIUD. Dit nummer correspondeert met PRONOM. Aanvullende informatie kan dan later nog worden opgehaald vanuit dit fileformaat register (PRONOM).

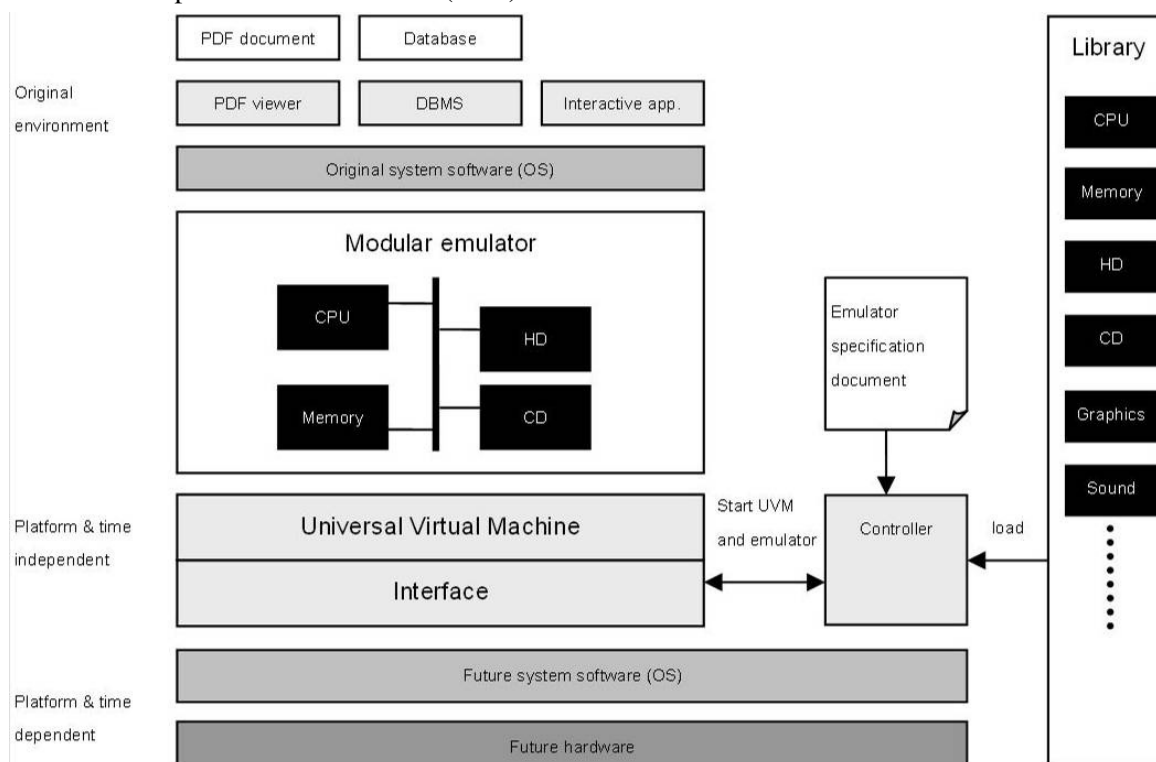
In een later stadia kan worden besloten bestanden uit een te rafelen en wellicht om te zetten naar XML (naast de PDF/A representatieversie en het eventuele digitale origineel). Dit proces heet normalisatie. Voordat dit een te overwegen stap wordt, moet de normalisatie software hiervoor veel verder worden ontwikkeld en ook om kunnen gaan met bijvoorbeeld gesloten bestandsformaten. Zelfs wanneer dit relevant wordt, zal DROID zo correct mogelijk inzicht blijven bieden in de bestandsformaten die in het archief aanwezig zijn.

Eind mei wordt versie 2.0 van DROID gelanceerd, samen met PRONOM 5. Deze nieuwe versie voorziet in meerdere fileformaat herkenningsmethoden en in een vollediger weergave van bijvoorbeeld het minetype.

Bijlage 7 Modulaire emulator

Van der Hoeven ea. hebben een conceptueel model uitgewerkt voor de modulaire emulator [Hoev05]. Dit model bestaat uit de volgende onderdelen (zie ook figuur 27):

- Modular emulator;
- Universal Virtual Machine (UVM);
- Controller;
- Module library;
- Emulator specification document (ESD).



Figuur 27 - Conceptueel model van de modulaire emulator

Modular emulator

De emulator vormt de kern van het schema. Met de emulator is het mogelijk het bovenliggende computerplatform te draaien op het huidige/toekomstige computerplatform. De benaming 'modular' komt voort uit de opbouw van deze emulator. Elk computercomponent (processor, geheugen, toetsenbord, etc.) vormt een aparte module. Deze module bootst de originele functionaliteit van het werkelijke component na. Door al deze modules aan elkaar te koppelen, ontstaat een emulator die een compleet computerplatform kan imiteren.

Universal Virtual Machine (UVM)

Door deze laag tussen het huidige platform en de emulator te plaatsen, wordt de emulator losgekoppeld van het onderliggende systeem. Met andere woorden, veranderingen aan de computer waarop de emulator draait hebben geen effect meer op de emulator zelf. Hierdoor wordt de emulator veel stabiel en behoeft veel minder onderhoud.

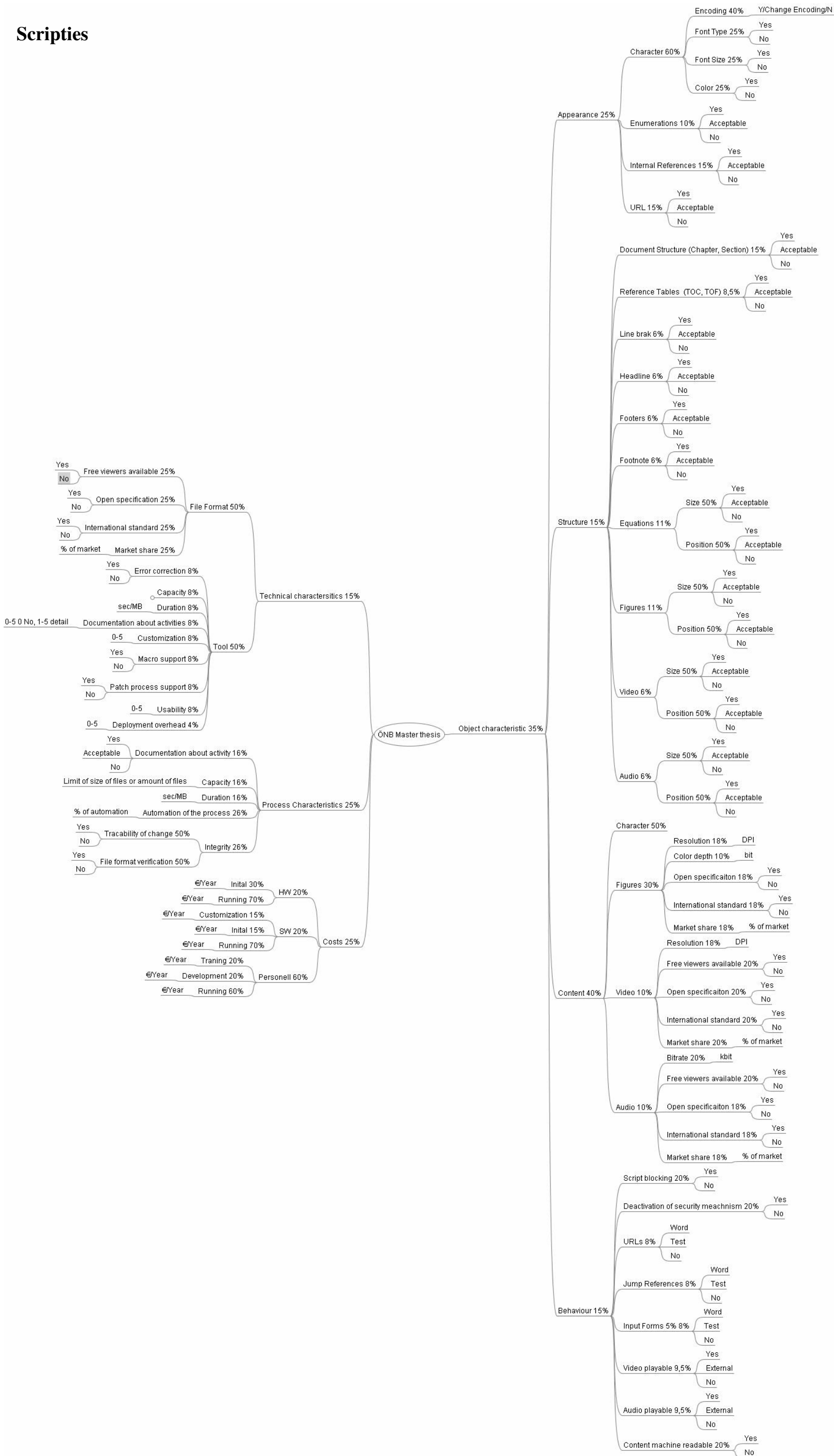
Controller, module library en ESD

De controller zorgt voor het configureren en uitvoeren van het emulatie proces. Het Emulator Specification Document (ESD) beschrijft welke modules er in de emulator nodig zijn om het gewenste platform te kunnen imiteren. Op basis van het ESD haalt de controller de benodigde modules uit de module library en configureert deze waar nodig.

Bijlage 8 Voorbeelden object bomen Websites



Scripties



Bijlage 9 PLATO

De PLANETS Preservation Planning Tool (PLATO) integreert mindmapping met de utiliteitsanalyse [11]. Daarnaast biedt de tool ondersteuning voor alle stappen van preservation planning die beschreven zijn door Strodl ea. [Stro07-2].

PLANETS Preservation Planning Tool (PLATO)

Home LoadProject New Project

Define Requirements
Define Basis
Define Sample Records
Identify Requirements

Evaluate Requirements
Define Alternatives
Go/No-Go
Develop Experiment
Run Experiment
Evaluate Experiment

Consider Results
Transform Measured Values
Set Importance Factors
Analyse Results
Sum
Multiplication
Sum of Priority
Austin Slight

Expand All | Collapse All
X Website

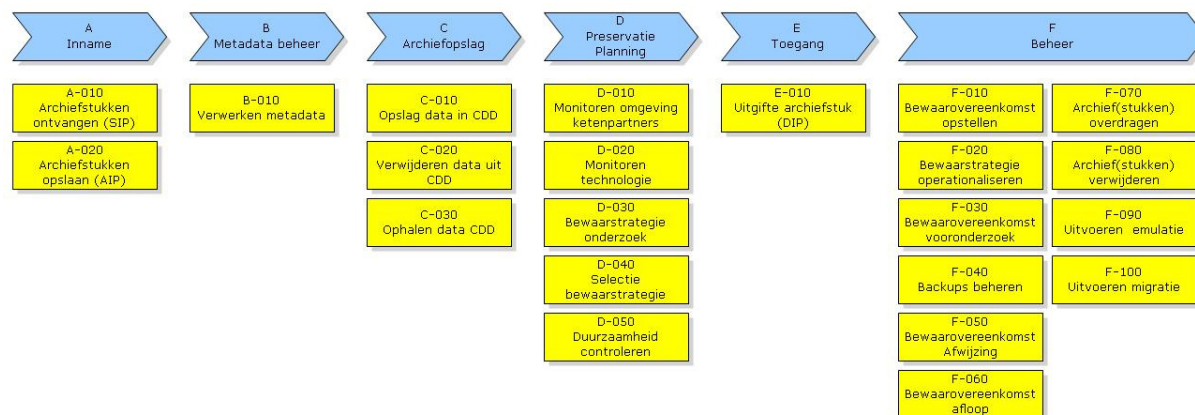
Focus	Node	Scale	Restriction	Unit
X	Website			
X	Record characteristics			
X	Appearance			
X	Content			
X	Structure			
X	Behaviour			
X	Context			
X	Technical characteristics			
X	Previous			
X	Ubiquity	Ordinal	Ubiquitous/Widespread/Specialised/Obsolete	number of tools
X	Support	Ordinal	0/1-5/6-10/10+	
X	Documentation			
X	Stability			
X	Ease of identification	Ordinal	Automatic/Manual/No	
X	Ease of validation	Ordinal	Automatic/Manual/No	
X	Lossiness	Ordinal	Lossy/Lossless	
X	IPR	Boolean		
X	Next 9 - 11 of 11			
X	Infrastructure characteristics			
X	Process characteristics			
X	Automation	Ordinal	Pre-configured/API/Scriptable/None	
X	Performance			
X	Streamable	Boolean		
X	Throughput	Float		files per hour
X	Scalability	Boolean		
X	Documentation			
X	Format	Ordinal	Compatible/Transformable/Incompatible	
X	Content	Ordinal	Comprehensive/Satisfactory/Minimal/None	
X	Usability	Ordinal	Good/Acceptable/Poor	
X	Validation			
X	Syntactic	Ordinal	Full/Partial/None	
X	Semantic	Ordinal	Full/Partial/None	

Quick Access: [Help]

© 2007 Institute of Software Technology and Interactive Systems: «office bears»

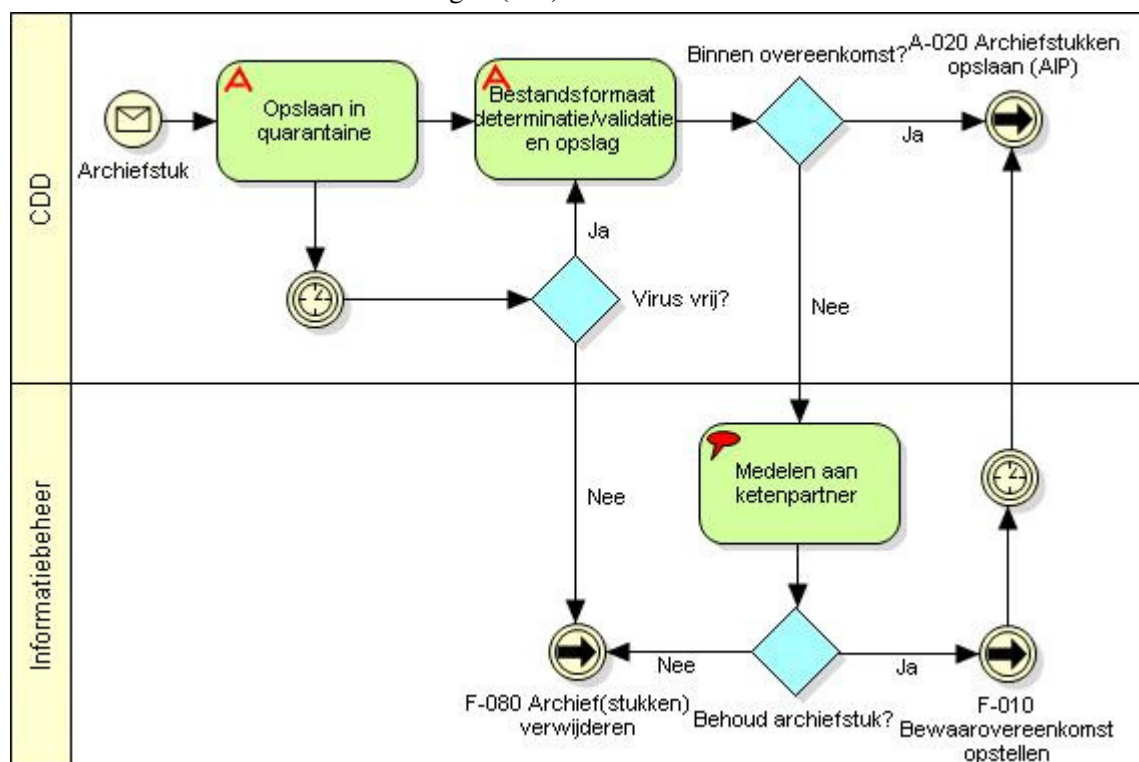
Bijlage 10 Inrichting digitale duurzaamheid CDD+

Onderstaande staat ook interactief op het intranet gepubliceerd.



1.1. A: Inname

1.1.1. A-010: Archiefstukken ontvangen (SIP)



Archiefstukken worden door ketenpartners overgedragen naar het CDD.

Taken:

A-010-010: Bestandsformaat bepaling/validatie en opslag

Architectuuronderdelen:	Determinatie tools Omschrijving: DROID, JHOVE
Gegevensdragers:	Archiefstuk (Read Update) Criteria (Read)

A-010-030: Meedelen aan ketenpartner

Gegevensdragers:	Bewaarovereenkomst (Read) Klantenbestand (Read)
------------------	--

A-010-070: Opslaan in quarantaine

Architectuuronderdelen:	Quarantaine omgeving
Gegevensdragers:	Archiefstuk (Create)

Beslismomenten:

Binnen overeenkomst?

Gegevensdragers:	Bewaarovereenkomst (Read) Klantenbestand (Read)
------------------	--

Behoud archiefstuk?

Virus vrij?

Architectuuronderdelen:	Quarantaine omgeving Omschrijving: Virusscanner
-------------------------	--

Gebeurtenissen:

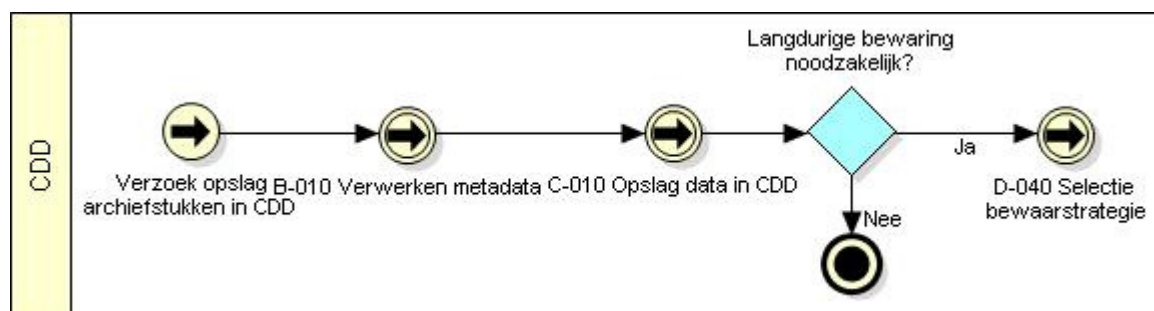
Archiefstuk

F-010: Bewaarovereenkomst opstellen

A-020: Archiefstukken opslaan (AIP)

F-080: Archief(stukken) verwijderen

1.1.2. A-020: Archiefstukken opslaan (AIP)



Archiefstukken of delen hiervan worden verwerkt naar AIP's.

Beslismomenten:

Langdurige bewaring noodzakelijk?

Architectuuronderdelen:	Database
Gegevensdragers:	Bewaarovereenkomst (Read) Metadata (Read)

Gebeurtenissen:

Verzoek opslag archiefstukken in CDD

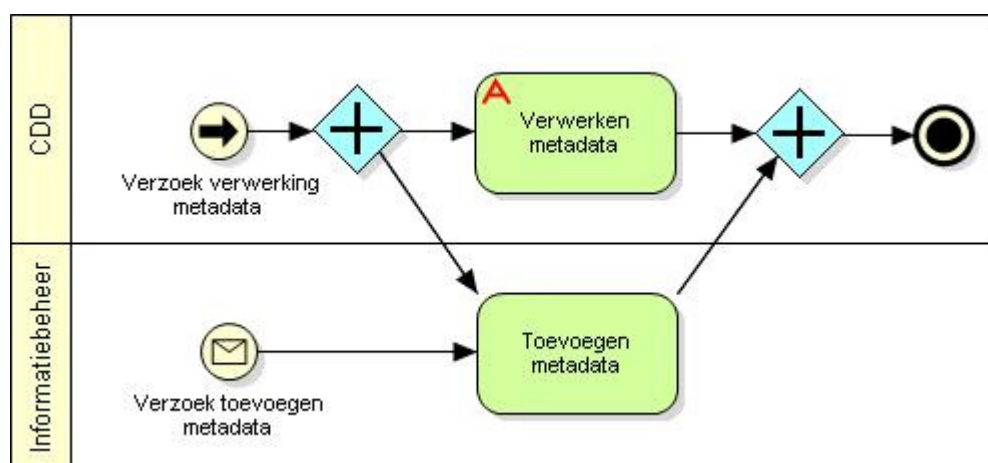
B-010: Verwerken metadata

D-040: Selectie bewaarstrategie

C-010: Opslag data in CDD

1.2. B: Metadata beheer

1.2.1. B-010: Verwerken metadata



Metadata wordt verwerkt in de database. Het kan hier gaan om het toevoegen, verwijderen of updaten van metadata.

Taken:

B-010-010: Verwerken metadata

Architectuuronderdelen:	Database Omschrijving: Oracle database voor de opslag van metadata Representatie Informatie register
Gegevensdragers:	Metadata (Create Update Delete)

B-010-020: Toevoegen metadata

Architectuuronderdelen:	Database Omschrijving: Oracle database voor de opslag van metadata Representatie Informatie register
Gegevensdragers:	Metadata (Create Update Delete)

Gebeurtenissen:

Verzoek verwerking metadata

Verzoek toevoegen metadata

1.3. C: Archiefopslag

1.3.1. C-010: Opslag data in CDD



Het daadwerkelijk opslaan van digitale bestanden.

Taken:

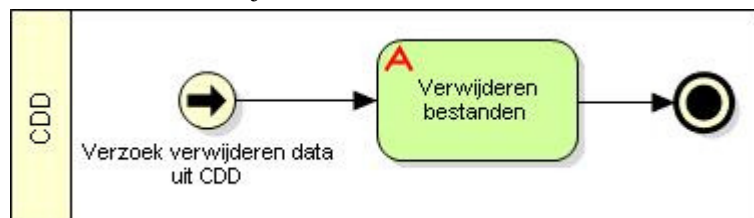
C-010-010: Opslaan bestanden juiste plek

Architectuuronderdelen:	Storage omgeving
Gegevensdragers:	Digitale bestanden (Create Read)

Gebeurtenissen:

Verzoek opslag data in CDD

1.3.2. C-020: Verwijderen data uit CDD



Het verwijderen van digitale bestanden

Taken:

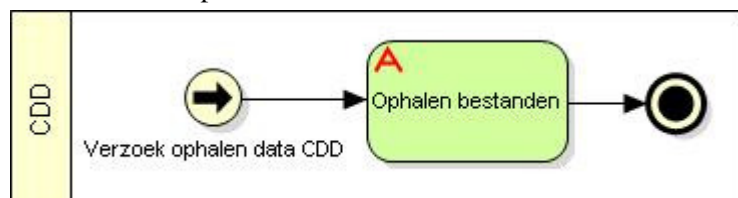
C-020-010: Verwijderen bestanden

Architectuuronderdelen:	Storage omgeving
Gegevensdragers:	Digitale bestanden (Delete)

Gebeurtenissen:

Verzoek verwijderen data uit CDD

1.3.3. C-030: Ophalen data CDD



Het ophalen van digitale bestanden.

Taken:

C-030-010: Ophalen bestanden

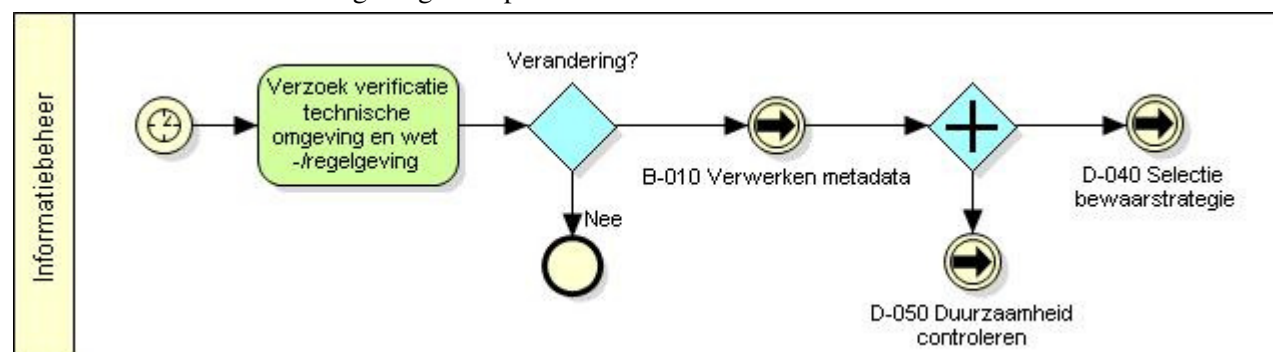
Architectuuronderdelen:	Storage omgeving
Gegevensdragers:	Digitale bestanden (Read)

Gebeurtenissen:

Verzoek ophalen data CDD

1.4. D: Preservatie Planning

1.4.1. D-010: Monitoren omgeving ketenpartners



De omgeving van de ketenpartner zal veranderen door de tijd heen. Het is dan ook van belang tijdig op handen zijnde veranderingen te constateren en hierop de anticiperen. Veranderingen in het domein van de ketenpartner zijn te vinden op het gebied van de techniek die gebruikt wordt (hard- en software) en de wet- en regelgeving die van kracht is.

Taken:

D-010-010: Verzoek verificatie technische omgeving en wet-/regelgeving

Architectuuronderdelen:	Database
Gegevensdragers:	Beschrijving huidige technische omgeving (Create) Beschrijving huidige wet-/ regelgeving (Create)

Beslismomenten:

Verandering?

Architectuuronderdelen:	Database
Gegevensdragers:	Beschrijving huidige technische omgeving (Read) Beschrijving huidige wet-/ regelgeving (Read) Beschrijving opgeslagen technische omgeving (Read) Beschrijving opgeslagen wet- en regelgeving (Read)

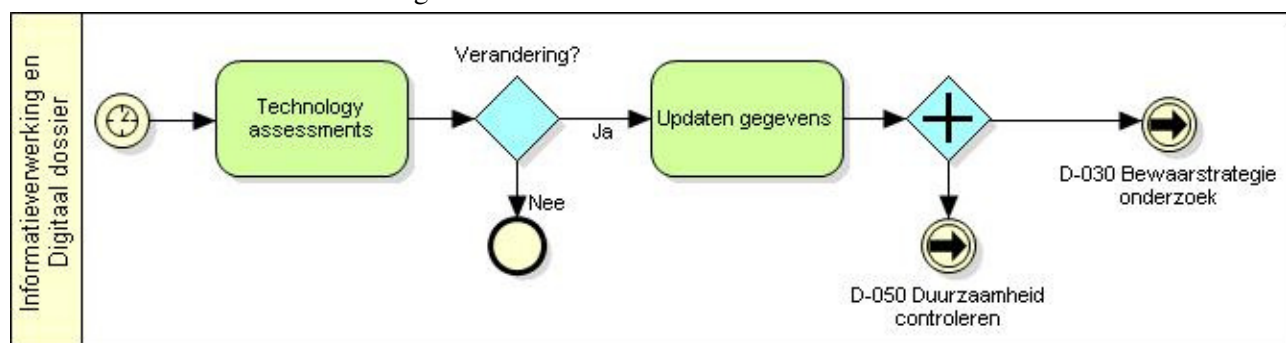
Gebeurtenissen:

D-040: Selectie bewaarstrategie

B-010: Verwerken metadata

D-050: Duurzaamheid controleren

1.4.2. D-020: Monitoren technologie



Ketenpartners zullen niet voorop lopen met de techniek. Veelal zijn soft- en hardware ver ontwikkeld voordat het wordt toegestaan binnen overheden. Om bekend te zijn met de veranderende techniek en de mogelijk daarmee gemoeide representatie problemen, moet de technologie goed in de gaten worden gehouden.

Taken:

D-020-010: Technology assessments

Architectuuronderdelen:	Database Determinatie tools
Gegevensdragers:	Technology reports (Create Read Update)

D-020-020: Updaten gegevens

Architectuuronderdelen:	Database Emulatie omgeving Migratie omgeving Representatie Informatie register
Gegevensdragers:	Emulator frame (Create Update) Emulator modulen (Create Update) Metadata (Create Update) Migratietool (Create Update) Technology reports (Read)

Beslismomenten:

Verandering?

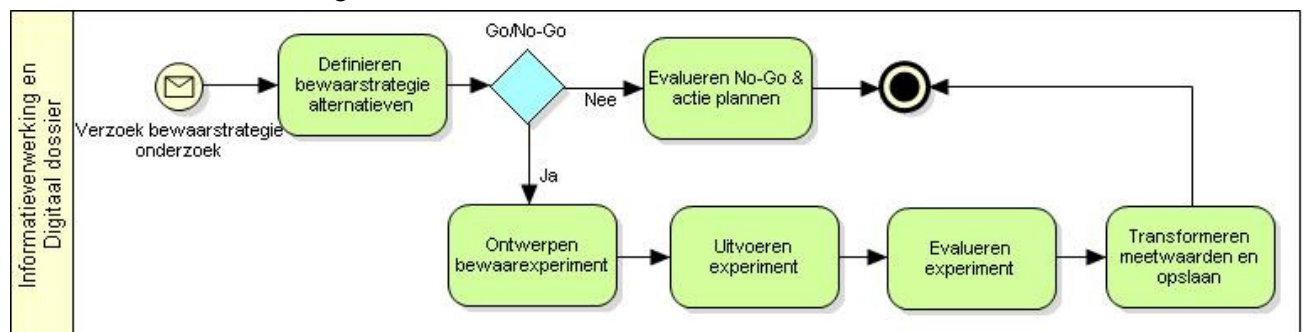
Architectuuronderdelen:	Database
Gegevensdragers:	Technology reports (Read)

Gebeurtenissen:

D-030: Bewaarstrategie onderzoek

D-050: Duurzaamheid controleren

1.4.3. D-030: Bewaarstrategie onderzoek



Er zijn zeer veel verschillende bewaarstrategieën. Om te weten te komen welke bewaarstrategie de beste is, moet een veelvoud van bewaarstrategieën worden getest. De meetwaarden moeten bewaard worden om later te gebruiken bij het komen van een onderbouwde keuze voor een bewaarstrategie gezien een context en informatietype.

Taken:

D-030-010: Definieren bewaarstrategie alternatieven

Architectuuronderdelen:	Bewaarstrategie repository Representatie Informatie register
Gegevensdragers:	Bewaarstrategie alternatieven (Create) Steekproef collectie (Read) Technology reports (Read)

D-030-020: Ontwerpen bewaarexperiment

Architectuuronderdelen:	Determinatie tools Emulatie omgeving Migratie omgeving Representatie Informatie register Storage omgeving
-------------------------	---

Gegevensdragers:	Bewaarstrategie alternatieven (Read) Criteria (Read) Emulator frame (Create) Emulator modulen (Create) Migratietool (Create) Steekproef collectie (Read) Technology reports (Read)
------------------	--

D-030-030: Uitvoeren experiment

Architectuuronderdelen:	Determinatie tools Emulatie omgeving Migratie omgeving
Gegevensdragers:	Bewaarstrategie alternatieven (Read) Bewaarstrategie resultaten (Create) Emulator frame (Read) Emulator modulen (Read) Migratietool (Read) Steekproef collectie (Read)

D-030-050: Evalueren No-Go & actie plannen

Architectuuronderdelen:	Planningtool
Gegevensdragers:	Actieplanning (Create) Bewaarstrategie alternatieven (Read) Evaluatie (Create)

D-030-060: Evalueren experiment

Architectuuronderdelen:	Determinatie tools Emulatie omgeving Migratie omgeving
-------------------------	--

Gegevensdragers:	Bewaarstrategie alternatieven (Read) Bewaarstrategie resultaten (Read) Criteria (Read) Evaluatie (Create) Steekproef collectie (Read)
------------------	---

D-030-070: Transformeren meetwaarden en opslaan

Architectuuronderdelen:	Bewaaracties evaluatie tool Bewaarstrategie repository Representatie Informatie register
Gegevensdragers:	Bewaarstrategie alternatieven (Read) Bewaarstrategie resultaten (Create Read) Criteria (Read) Evaluatie (Read) Weegfactoren per criteria (Read)

Beslismomenten:

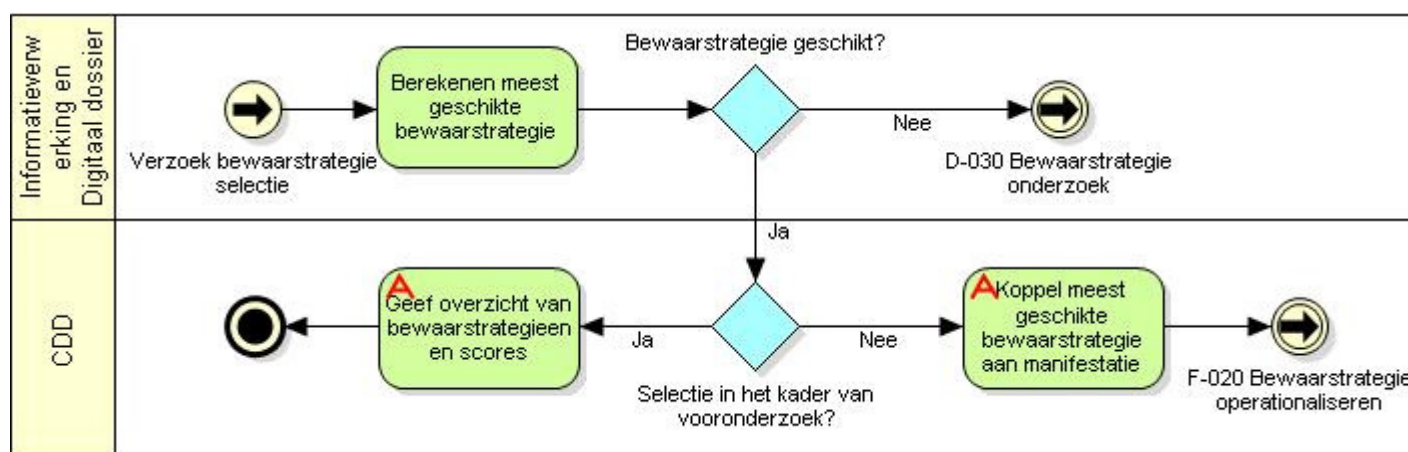
Go/No-Go

Architectuuronderdelen:	Bewaarstrategie repository
Gegevensdragers:	Bewaarstrategie alternatieven (Read) Technology reports (Read)

Gebeurtenissen:

Verzoek bewaarstrategie onderzoek

1.4.4. D-040: Selectie bewaarstrategie



Er zijn bewaarexperimenten uitgevoerd, meetwaarden opgeslagen, getransformeerd naar te vergelijken waarden en weegfactoren zijn door ketenpartners toegekend aan criteria. Er kan nu berekend worden wat de meest geschikte bewaarstrategie is gezien een bepaalde context (die van de ketenpartner) en informatietype (bijvoorbeeld document).

Taken:

D-040-010: Berekenen meest geschikte bewaarstrategie

Architectuuronderdelen:	Bewaaracties evaluatie tool Bewaarstrategie adviseur Bewaarstrategie repository Database
Gegevensdragers:	Basisgegevens collectie (Read) Beschikbare geschikte bewaarstrategieën (Create) Bewaarstrategie resultaten (Read) Digitale bestanden (Read) Criteria (Read) Metadata (Read) Steekproef collectie (Read) Weegfactoren per criteria (Read)

D-040-020: Koppel meest geschikte bewaarstrategie aan manifestatie

Architectuuronderdelen:	Bewaarstrategie adviseur Database Representatie Informatie register
-------------------------	---

Gegevensdragers:	Beschikbare geschikte bewaarstrategieën (Read) Metadata (Update)
------------------	---

D-040-030: Geef overzicht van bewaarstrategieën en scores

Architectuuronderdelen:	Bewaarstrategie adviseur
Gegevensdragers:	Beschikbare geschikte bewaarstrategieën (Read)

Beslismomenten:

Bewaarstrategie geschikt?

Architectuuronderdelen:	Bewaarstrategie adviseur
Gegevensdragers:	Beschikbare geschikte bewaarstrategieën (Read)

Selectie in het kader van vooronderzoek?

Architectuuronderdelen:	Database
-------------------------	----------

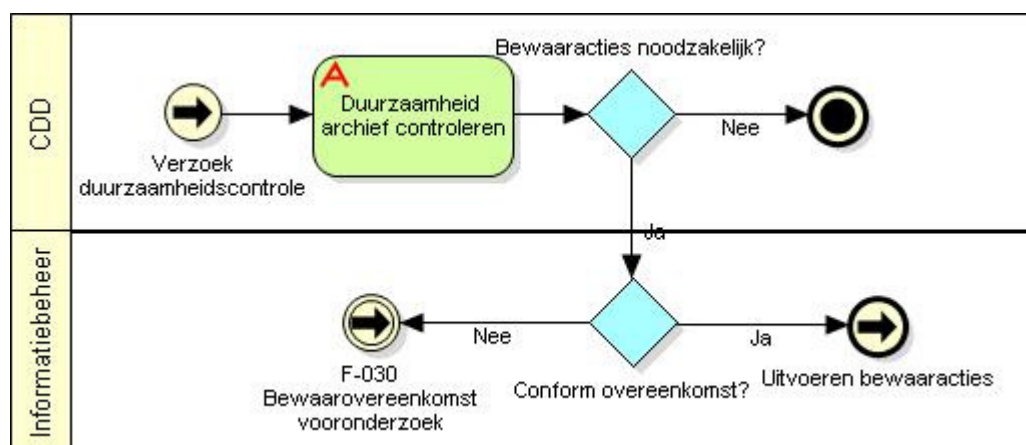
Gebeurtenissen:

Verzoek bewaarstrategie selectie

D-030: Bewaarstrategie onderzoek

F-020: Bewaarstrategie operationaliseren

1.4.5. D-050: Duurzaamheid controleren



De duurzaamheid moet naar iedere monitorcycli worden gecontroleerd. Herberekening van bewaarstrategieën aan de hand van veranderingen van in de context een techniek.

Taken:

D-050-010: Duurzaamheid archief controleren

Omschrijving:	Herberekenen of er betere bewaarstrategieën zijn toe te passen op de opgeslagen digitale bestanden.
Architectuuronderdelen:	Bewaarstrategie repository Database Storage omgeving
Gegevensdragers:	Beschikbare geschikte bewaarstrategieën (Read) Bewaarovereenkomst (Read) Digitale bestanden (Read)

Beslismomenten:

Bewaaracties noodzakelijk?

Architectuuronderdelen:	Bewaarstrategie adviseur
-------------------------	--------------------------

Conform overeenkomst?

Gebeurtenissen:

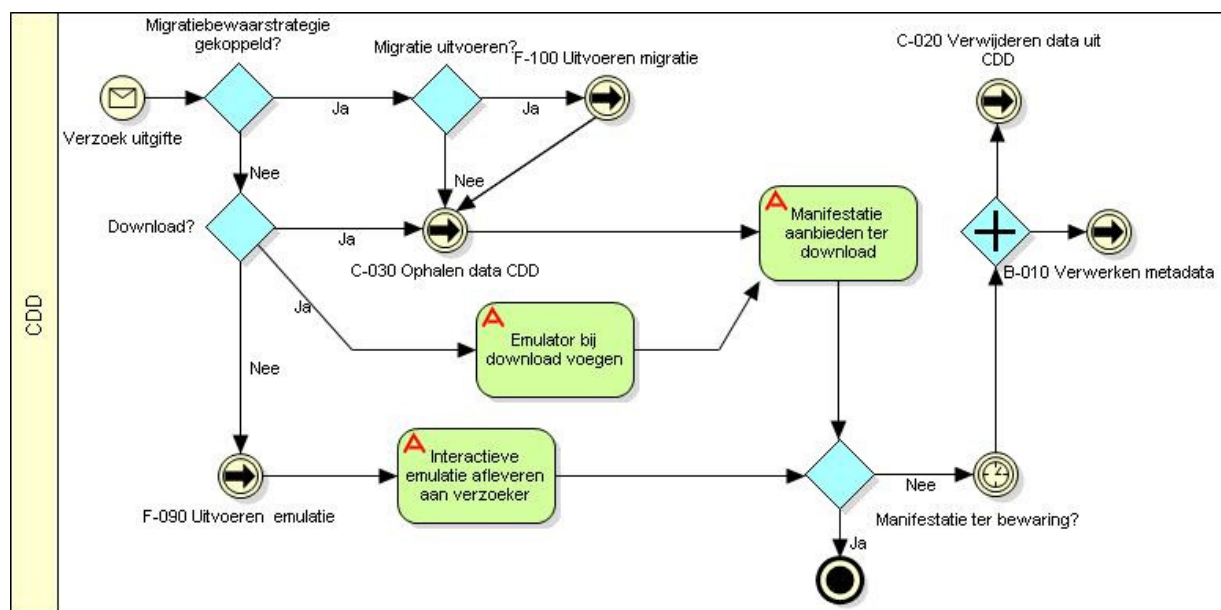
F-030: Bewaarovereenkomst vooronderzoek

Uitvoeren bewaaracties

Verzoek duurzaamheidscontrole

1.5. E: Toegang

1.5.1. E-010: Uitgifte archiefstuk (DIP)



Archiefstukken moeten gerepresenteerd kunnen worden aan gerechtigden. Dit proces zorgt voor het opstellen van een DIP; de manifestatie van het archiefstuk die de aanvragen wil inzien.

Taken:

E-010-060: Manifestatie aanbieden ter download

Architectuuronderdelen:	Storage omgeving
Gegevensdragers:	Digitale bestanden (Read)

E-010-070: Interactieve emulatie afleveren aan verzoeker

Architectuuronderdelen:	Emulatie omgeving
Gegevensdragers:	Emulator frame (Read) Emulator modules (Read)

E-010-080: Emulator bij download voegen

Architectuuronderdelen:	Representatie Informatie register Storage omgeving
Gegevensdragers:	Digitale bestanden (Create) Emulator frame (Read) Emulator modules (Read)

Beslismomenten:

Migratie uitvoeren?

Architectuuronderdelen:	Database
Gegevensdragers:	Steekproef collectie (Read)

Migratiebewaarstrategie gekoppeld?

Architectuuronderdelen:	Database
Gegevensdragers:	Metadata (Read)

Manifestatie ter bewaring?

Gegevensdragers:	Metadata (Read)
------------------	-------------------

Download?

Omschrijving:	Verzoek door raadpleger
---------------	-------------------------

Gebeurtenissen:

Verzoek uitgifte

F-100: Uitvoeren migratie

C-020: Verwijderen data uit CDD

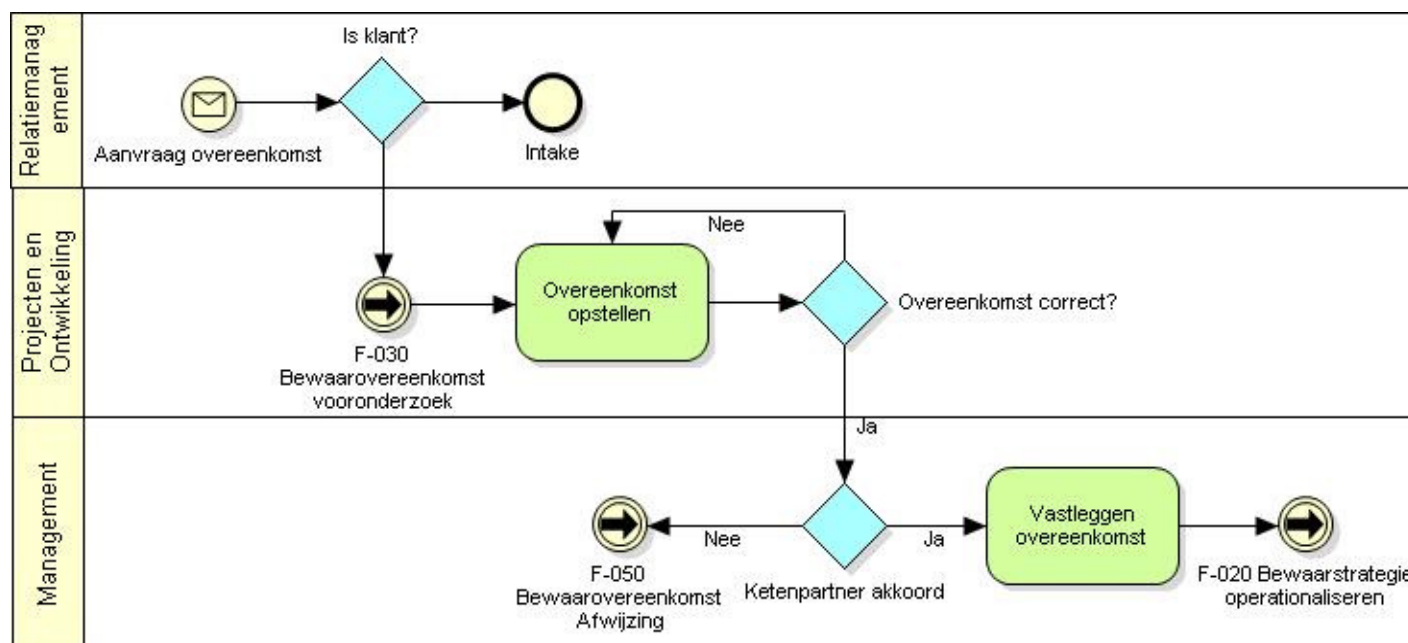
F-090: Uitvoeren emulatie

B-010: Verwerken metadata

C-030: Ophalen data CDD

1.6. F: Beheer

1.6.1. F-010: Bewaarovereenkomst opstellen



Er moet niet zomaar digitale informatie in het CDD wordt gezet ter bewaring. Misschien kan het CDD helemaal niet de gegarandeerde bewaring waarmaken voor de verwachte bewaartermijn! Daarom moeten er bewaarovereenkomsten worden gesloten met ketenpartners. Deze overeenkomsten leggen onder andere vast voor welke termijn JustID zal voorzien in de bewaring van een bepaald type archiefstuk en welke digitale informatie deze archiefstukken mogen bevatten.

Taken:

F-010-030: Overeenkomst opstellen

Architectuuronderdelen:	Database
Gegevensdragers:	Basisgegevens collectie (Read) Beschikbare geschikte bewaarstrategieën (Read) Beschrijving huidige technische omgeving (Read) Beschrijving huidige wet-/ regelgeving (Read) Bewaarovereenkomst (Read) Concept bewaarovereenkomst (Create Update) Klantenbestand (Create Read)

F-010-040: Vastleggen overeenkomst

Gegevensdragers:	Bewaarovereenkomst (Create) Concept bewaarovereenkomst (Read)
------------------	--

Beslismomenten:

Is klant?

Architectuuronderdelen:	Database
Gegevensdragers:	Klantenbestand (Read)

Overeenkomst correct?

Architectuuronderdelen:	Database
Gegevensdragers:	Basisgegevens collectie (Read) Beschrijving opgeslagen technische omgeving (Read) Beschrijving opgeslagen wet- en regelgeving (Read) Concept bewaarovereenkomst (Read) Klantenbestand (Read)

Ketenpartner akkoord

Gegevensdragers:	Bewaarovereenkomst (Read)
------------------	-----------------------------

Gebeurtenissen:

Aanvraag overeenkomst

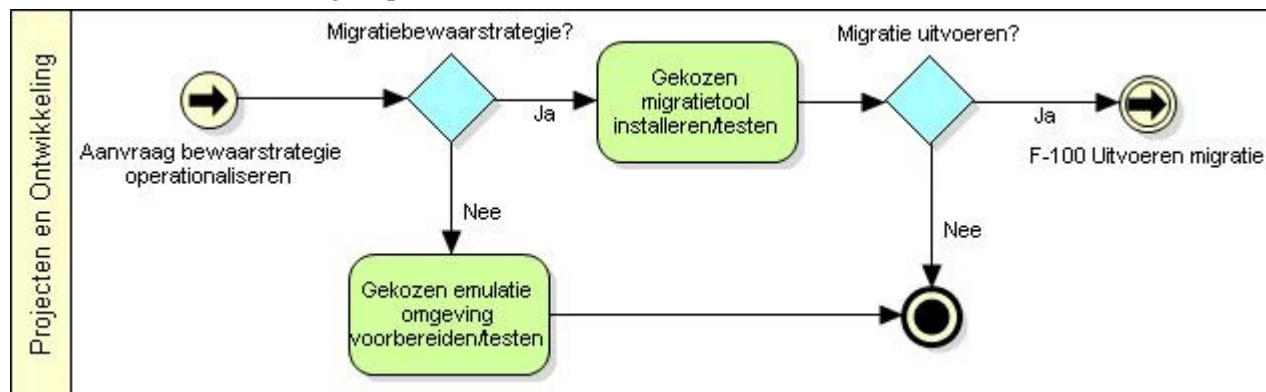
F-030: Bewaarovereenkomst vooronderzoek

F-050: Bewaarovereenkomst Afwijzing

F-020: Bewaarstrategie operationaliseren

Intake

1.6.2. F-020: Bewaarstrategie operationaliseren



Om geteste en geselecteerde bewaarstrategieën te kunnen gebruiken moeten deze in het CDD

operationeel gemaakt worden.

Taken:

F-020-010: Gekozen migratietool installeren/testen

Architectuuronderdelen:	Database Migratie omgeving Omschrijving: Biedt ruimte aan het uitvoeren van migratietools Representatie Informatie register Omschrijving: Bevat de benodigde tools en schema's die noodzakelijk zijn om digitale informatie te kunnen representeren.
Gegevensdragers:	Metadata (Read Update) Migratietool (Create Read) Steekproef collectie (Read)

F-020-020: Gekozen emulatie omgeving voorbereiden/testen

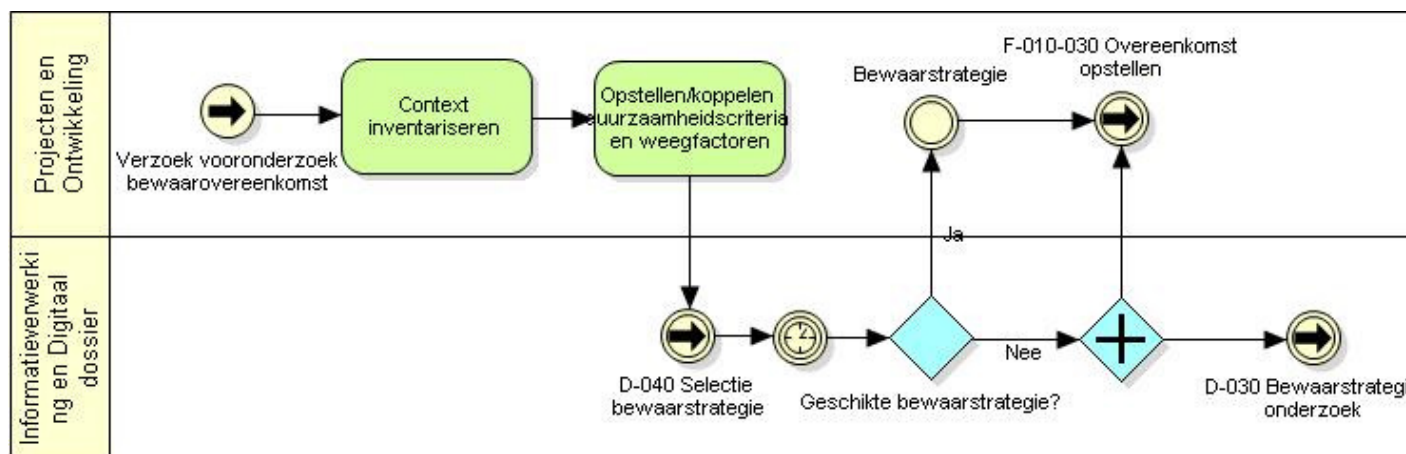
Architectuuronderdelen:	Database Emulatie omgeving Omschrijving: Omgeving die ruimte biedt aan de emulator Representatie Informatie register Omschrijving: Bevat de benodigde tools en schema's die noodzakelijk zijn om digitale informatie te kunnen representeren.
Gegevensdragers:	Emulator frame (Create Read) Emulator modulen (Create Read) Metadata (Read Update) Steekproef collectie (Read)

Beslismomenten:

Migratiebewaarstrategie?

Architectuuronderdelen:	Database
-------------------------	----------

Migratie uitvoeren?	
Architectuuronderdelen:	Database Omschrijving: Geeft aan of er initieel al een migratie moet worden uitgevoerd



Architectuuronderdelen:	Database Mindmapping tool
Gegevensdragers:	Basisgegevens collectie (Create) Beschrijving huidige technische omgeving (Create) Beschrijving huidige wet-/ regelgeving (Create) Metadata (Create) Steekproef collectie (Create)

F-030-020: Opstellen/koppelen criteria en weegfactoren

Architectuuronderdelen:	Database Mindmapping tool
Gegevensdragers:	Basisgegevens collectie (Read) Beschrijving opgeslagen technische omgeving (Read) Beschrijving opgeslagen wet- en regelgeving (Read) Criteria (Create) Steekproef collectie (Read) Weegfactoren per criteria (Create)

Beslismomenten:

Geschikte bewaarstrategie?

Architectuuronderdelen:	Bewaaracties evaluatie tool Database
Gegevensdragers:	Beschikbare geschikte bewaarstrategieën (Read) Metadata (Read)

Gebeurtenissen:

Verzoek vooronderzoek bewaarovereenkomst

D-030: Bewaarstrategie onderzoek

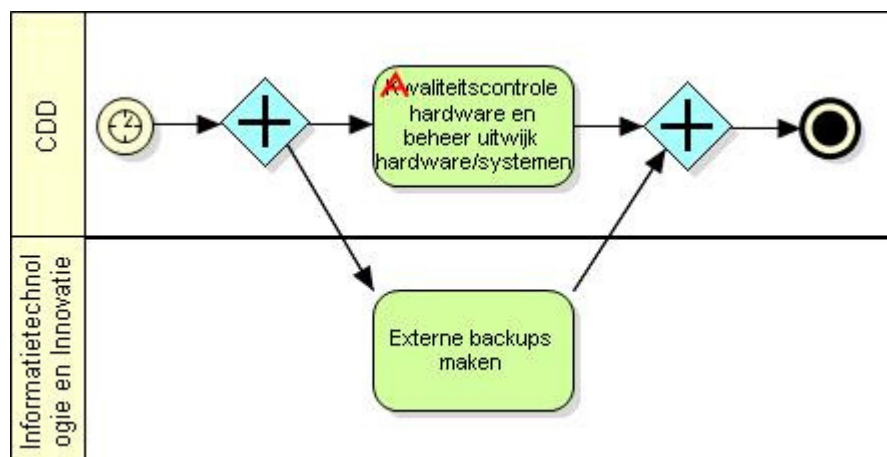
D-040: Selectie bewaarstrategie

Bewaarstrategie

Omschrijving:	De meest geschikte bewaarstrategie, gezien de context en digitale informatie
---------------	--

F-010-030: Overeenkomst opstellen

1.6.4. F-040: Backups beheren



Met regelmaat moeten er (externe) backups worden gemaakt van de informatie in het CDD. Daarnaast moet er hardware of complete systemen klaar staan om te voorzien in uptime en bitstream bewaring.

Taken:

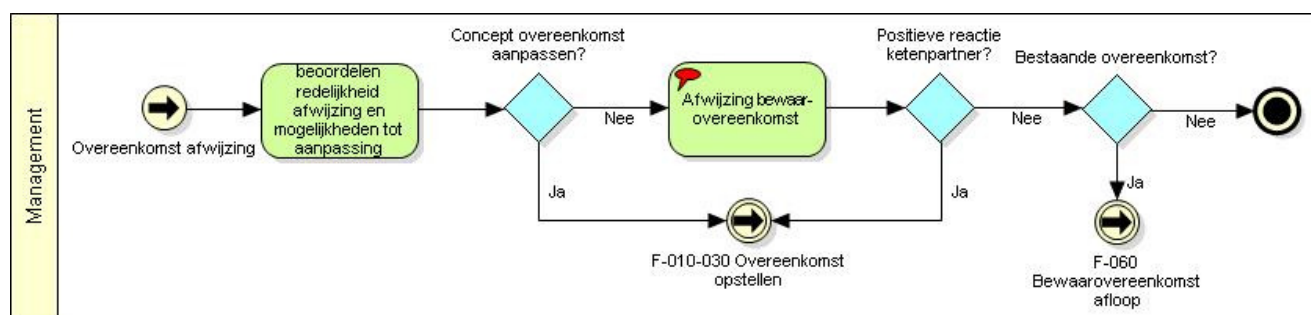
F-040-010: Externe backups maken

Architectuuronderdelen:	Backup faciliteiten Database Storage omgeving
Gegevensdragers:	Backups (Create Update) Digitale bestanden (Read) Metadata (Read)

F-040-020: Kwaliteitscontrole hardware en beheer uitwijk hardware/systemen

Architectuuronderdelen:	Backup faciliteiten
Gegevensdragers:	Hardware onderdelen (Create) Huidige hardware systemen (Read Update Delete)

1.6.5. F-050: Bewaarovereenkomst Afwijzing



Het kan dat er (tijdelijk) niet tot een bewaarovereenkomst gekomen kan worden. Dit zou kunnen optreden wanneer er nog geen geschikte bewaarstrategie beschikbaar is voor de digitale informatie die een ketenpartner wil bewaren. Deze situatie ontstaat wanneer een ketenpartner zeer hoge eisen stelt aan de bewaring, bijvoorbeeld de eis dat de gehele "look-and-feel" duurzaam bewaard moet worden voor zeer lange tijd.

Taken:

F-050-030: Beoordelen redelijkheid afwijzing en mogelijkheden tot aanpassing

Architectuuronderdelen:	Database
Gegevensdragers:	Afwijksbericht (Read) Bewaarovereenkomst (Read) Concept bewaarovereenkomst (Read) Klantenbestand (Read)

F-050-040: Afwijzing bewaar-overeenkomst

Gegevensdragers:	Afwijksbericht (Create)
------------------	---------------------------

Beslismomenten:

Bestaande overeenkomst?

Architectuuronderdelen:	Database
Gegevensdragers:	Klantenbestand (Create Read Update) Metadata (Read)

Concept overeenkomst aanpassen?

Positieve reactie ketenpartner?

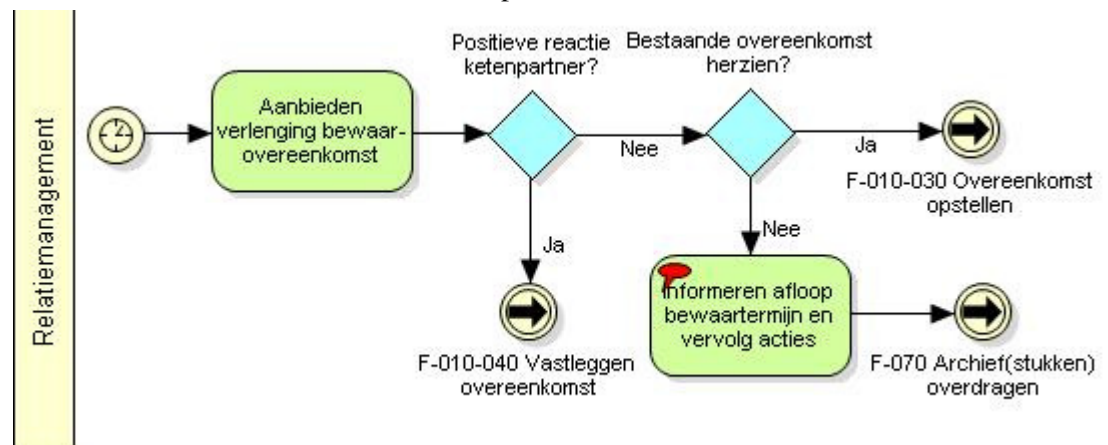
Gebeurtenissen:

Overeenkomst afwijzing

F-010-030: Overeenkomst opstellen

F-060: Bewaarovereenkomst afloop

1.6.6. F-060: Bewaarovereenkomst afloop



Een bewaarovereenkomst is eindig. Ooit zal de digitale informatie naar het Nationaal Archief moeten worden overgedragen of zal de bewaarovereenkomst moeten worden herzien/verlengt.

Taken:

F-060-040: Informeren afloop bewaartermijn en vervolg acties

Gegevensdragers:	Bewaarovereenkomst (Read) Klantenbestand (Read)
------------------	--

F-060-080: Aanbieden verlenging bewaar- overeenkomst

Architectuuronderdelen:	Database
Gegevensdragers:	Bewaarovereenkomst (Read) Klantenbestand (Read) Metadata (Read)

Beslismomenten:

Bestaande overeenkomst herzien?

Architectuuronderdelen:	Database
Gegevensdragers:	Klantenbestand (Create Read Update)

Positieve reactie ketenpartner?

Architectuuronderdelen:	Database
Gegevensdragers:	Klantenbestand (Update)

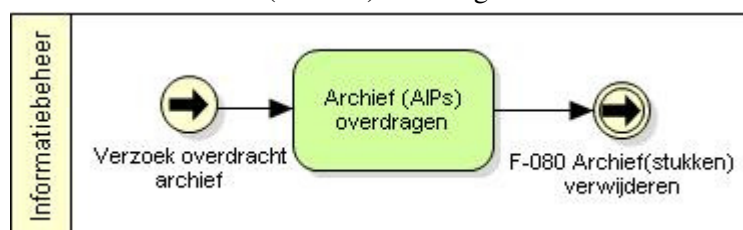
Gebeurtenissen:

F-010-030: Overeenkomst opstellen

F-010-040: Vastleggen overeenkomst

F-070: Archief(stukken) overdragen

1.6.7. F-070: Archief(stukken) overdragen



Archieven en archiefstukken moeten kunnen worden overgedragen naar alternatieve archiefsystemen, zoals die van het Nationaal Archief of een toekomstige opvolger van het CDD.

Taken:

F-070-010: Archief (AIPs) overdragen

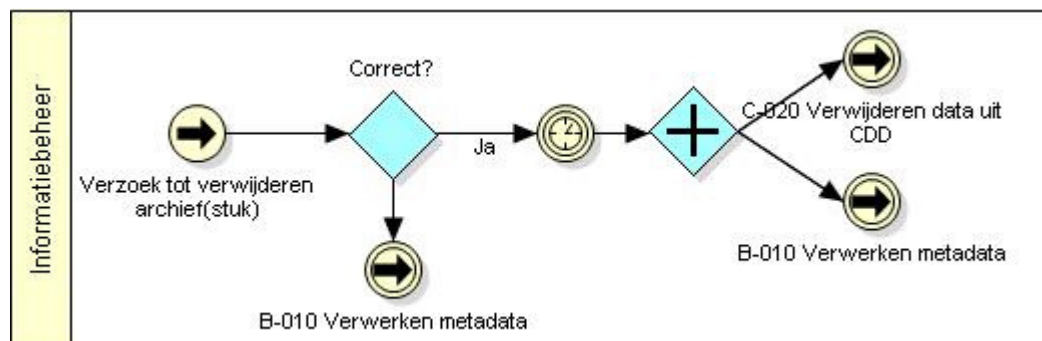
Architectuuronderdelen:	Database Storage omgeving
Gegevensdragers:	AIP (Create) Beschrijving opgeslagen technische omgeving (Read) Beschrijving opgeslagen wet- en regelgeving (Read) Bewaarovereenkomst (Read) Digitale bestanden (Read) Klantenbestand (Read) Metadata (Read)

Gebeurtenissen:

Verzoek overdracht archief

F-080: Archief(stukken) verwijderen

1.6.8. F-080: Archief(stukken) verwijderen



Na overdracht zouden archieven verwijderd mogen worden.

Beslismomenten:

Correct?

Architectuuronderdelen:	Database
Gegevensdragers:	Metadata (Read)

Gebeurtenissen:

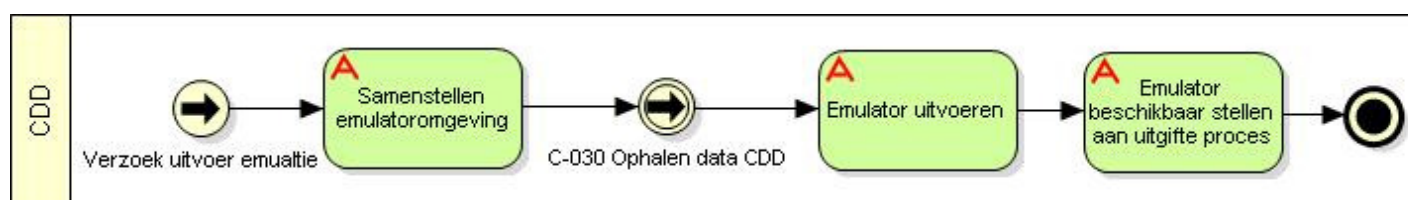
Verzoek tot verwijderen archief(stuk)

C-020: Verwijderen data uit CDD

B-010: Verwerken metadata

B-010: Verwerken metadata

1.6.9. F-090: Uitvoeren emulatie



Emulatie moet daadwerkelijk worden uitgevoerd in de geselecteerde emulatie omgeving.

Taken:

F-090-060: Samenstellen emulatoromgeving

Architectuuronderdelen:	Database Emulatie omgeving Representatie Informatie register
-------------------------	--

Gegevensdragers:	Digitale bestanden (Read) Emulator frame (Read) Emulator modulen (Read) Metadata (Read)
------------------	--

F-090-070: Emulator uitvoeren

Architectuuronderdelen:	Database Emulatie omgeving Storage omgeving
Gegevensdragers:	Digitale bestanden (Read) Metadata (Read)

F-090-080: Emulator beschikbaar stellen aan uitgifte proces

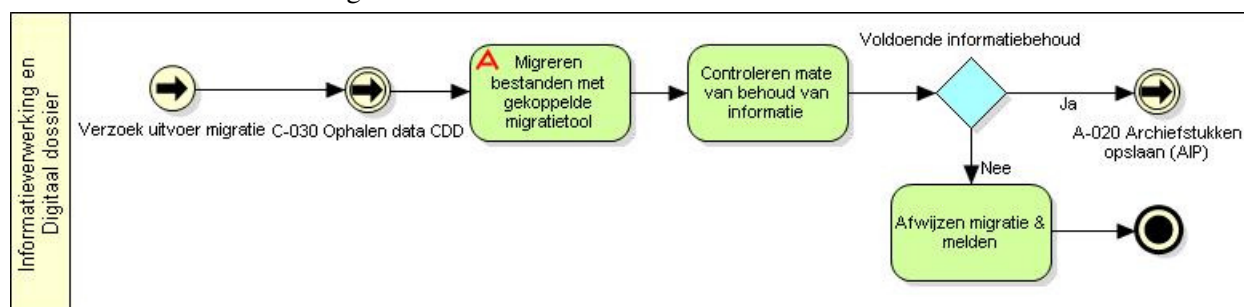
Architectuuronderdelen:	Emulatie omgeving
Gegevensdragers:	Emulator frame (Read) Emulator modulen (Read)

Gebeurtenissen:

Verzoek uitvoer emualtie

C-030: Ophalen data CDD

1.6.10. F-100: Uitvoeren migratie



Migraties moeten worden uitgevoerd met de daadvoor geselecteerde migratietool.

Taken:

F-100-030: Migreren bestanden met gekoppelde migratietool

Architectuuronderdelen:	Database Migratie omgeving
Gegevensdragers:	Digitale bestanden (Create Read) Metadata (Read) Migratietool (Read)

F-100-050: Controleren mate van behoud van informatie

Architectuuronderdelen:	Database Determinatie tools Migratie omgeving
Gegevensdragers:	Bewaarovereenkomst (Read) Bewaarrapport (Create) Digitale bestanden (Read) Metadata (Create Read)

F-100-060: Afwijzen migratie & melden

Architectuuronderdelen:	Bewaaracties evaluatie tool Database Migratie omgeving
Gegevensdragers:	Bewaarovereenkomst (Read) Bewaarrapport (Read) Digitale bestanden (Delete) Klantenbestand (Read Update) Metadata (Update)

Beslismomenten:

Voldoende informatiebehoud

Architectuuronderdelen:	Bewaaracties evaluatie tool Database
Gegevensdragers:	Bewaarrapport (Read) Metadata (Read)

Gebeurtenissen:

Verzoek uitvoer migratie

A-020: Archiefstukken opslaan (AIP)

C-030: Ophalen data CDD

Bijlage 11 Reflectie

Reflecterend op het onderwerp digitale duurzaamheid van archieven realiseer ik mij dat er een nieuwe taak ligt voor de archivaris in het digitale tijdperk. Er zal zich geen laag stof meer ophopen op de archieven, want die wordt door koeling en afzuiging de serverruimtes uitgewerkt. De archivaris wordt een expert in de verschillende manieren van coderen van informatie in digitale vorm.

Naast deze verschuiving staan de veranderingen in technologie de laatste 20 jaar niet stil, maar ik denk niet dat deze rechtevenredig kunnen worden doorgetrokken naar de toekomst. Meer en meer verschillende manieren van coderen en representatie methoden van informatie in digitale vorm worden transparanter, verschuiven naar open standaarden en maken de informatie daarmee duurzamer. Het relateert het onderwerp van digitale duurzaamheid, maar maakt het voor eeuwige bewaring niet minder relevant.

Reflecterend op het afstudeerproject ben ik van mening dat ik erg veel heb geleerd. Een paar onderdelen wil ik specifiek noemen: het managen van een eigen onderzoek, initiatief nemen, het constant speuren naar literatuur, deze eigen maken, onderzoek doen. Allemaal kenmerken waarmee ik gedurende mijn afstuderen mee ben geconfronteerd en die ik niet uit de weg ben gaan. Soms was het verleidelijk om even iets totaal anders te doen dan het focussen op mijn afstuderen, maar al met al is hetgeen dat voor u ligt gerealiseerd.

De werkomgeving die JustID en Deloitte mij boden is mij goed bevallen. JustID met haar ketenbrede visie en initiatieven verdient in toenemende mate de aandacht van ketenpartners binnen én buiten de Strafrechtsketen. Het was een zeer leerzame context om in te participeren, namelijk die van keteninformatisering en overheid, maar helaas tevens een context die niet of onvoldoende terug kwam in mijn studie. Vanuit mijn studie was het afstuderen via Deloitte een zeer goede omgeving om toch ook de bedrijfskundige kant van mijn studie sterker tot zijn recht te laten komen. De omgeving binnen Deloitte heb ik over het algemeen als zeer leerzaam en ondernemend ervaren. Er is veel kennis aanwezig in zowel de medewerkers als de verschillende kennissystemen en ondersteunende software.

Ik hoop met mijn voorstel voor de inrichting van de organisatie van het CDD+ en de noodzakelijke informatiesysteem architectuur op het gebied van digitale duurzaamheid, een goede bijdrage te hebben geleverd voor de omgang met een vaak niet onderkend probleem.