



***Van digitale
vluchtigheid naar
digitaal houvast***

Bewaren van e-mail

Testbed Digitale Bewaring is een initiatief van de Rijksarchiefdienst en het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Het is een onderzoeksprogramma waar de praktische toepasbaarheid getoetst wordt van verschillende manieren om (overheids-) informatie te bewaren en toegankelijk te houden voor de toekomst. Testbed Digitale Bewaring maakt deel uit van de stichting ICTU waar verschillende programma's zijn ondergebracht die alle ten doel hebben de digitale overheid op te bouwen.

ICTU
Nieuwe Duinweg 24-26
2587 AD Den Haag

Tel. 070 888 77 77
Fax: 070 888 78 88

E-mail testbed@ictu.nl
www.digitaleduurzaamheid.nl

Testbed Digitale Bewaring *Van digitale vluchtigheid naar digitaal houvast. Bewaren van e-mail.*

Den Haag, april 2003.

ISBN 90-807758-1-9

© Testbed Digitale Bewaring, Den Haag 2003

Alle rechten voorbehouden. Niets uit deze uitgave mag openbaar gemaakt of verveelvoudigd door middel van druk, fotokopie, microfilm of welke andere wijze dan ook zonder voorafgaande toestemming van het programmabureau. Het gebruik van (delen van) deze publicatie als toelichting of ondersteuning bij artikelen, boeken en scripties en dergelijke is toegestaan, mits de bron duidelijk wordt vermeld.

Inhoud

Voorwoord/5

Leeswijzer/7

1. De Nederlandse digitale overheid/9

- 1.1. Ontwikkelingen binnen de digitale overheid/9
- 1.2. Bewust werken: digitaal duurzaam beheren/10
- 1.3. Digitaal werken betekent ook digitaal bewaren/10
- 1.4. Digitaal bewaren en de wet/11
- 1.5. Een technische oplossing voorhanden?/12
- 1.6. De opdracht voor Testbed Digitale Bewaring/13

2. Digitale documenten en authenticiteit/14

- 2.1. Definitie van een digitaal document/14
- 2.2. Authenticiteit als kernbegrip/15
- 2.3. Digitale documenten, digitale kenmerken/16

3. E-mail bewaren in authentieke staat/19

- 3.1. De vlucht van e-mail: problemen bij het bewaren/19
- 3.2. De status van e-mail/20
- 3.3. Kenmerken van e-mail/22
- 3.4. E-mail en het transmissiebestand/22
- 3.5. Authenticiteitseisen voor e-mail/26
- 3.6. De digitale handtekening/29
- 3.7. Samenvatting/29

4. Drie bewaarstrategieën getoetst/31

- 4.1. Inleiding/31
- 4.2. Migratie als bewaarstrategie/31
 - 4.2.1. *Achterwaartse compatibiliteit*/32
 - 4.2.2. *Interoperabiliteit*/33
 - 4.2.3. *Conversie naar standaarden*/35
- 4.3. XML als bewaarstrategie/36
 - 4.3.1. *Wrapper en kapstok*/37
 - 4.3.2. *XML als bestandsformaat*/40
- 4.4. Emulatie als bewaarstrategie/42
 - 4.4.1. *Hardware-emulatie*/43
 - 4.4.2. *Software-emulatie*/43
 - 4.4.3. *De Universele Virtuele Computer-strategie (UVC)*/44
- 4.5. Conclusie/48

5. Aanbevolen e-mailbenadering/50

- 5.1. Inleiding/50
- 5.2. E-mail als transmissiebestand/50
- 5.3. Conversieprocedures/51
- 5.4. Duurzaam bewaarde e-mail/55

6. Concrete acties/61

- 6.1. Actieplan voor managers/63
- 6.2. Actieplan voor records managers/65
- 6.3. Actieplan voor ICT-ers/71
- 6.4. Actieplan voor eindgebruikers/77

Verklarende woordenlijst/85

Bibliografie/89

- Appendix A: Technische beschrijving van de e-mail/XML demo;
- Appendix B: XML schema, behorend bij de e-mail/XML demo;
- Appendix C: Bewaar Transactie Logbestand en
- Appendix D: Verschillende representaties van een e-mailbericht (HTML-representatie, transmissiebestand, platte tekst en XML-representatie).

Voorwoord

In de serie *Van digitale vluchtigheid naar digitaal houvast* die Testbed uitgeeft, was het deel *Bewaren van e-mail* oorspronkelijk niet als eerst gepland. De problemen en vragen rondom het bewaren van e-mail heeft ons aanleiding gegeven om dit deel als eerste uit te geven.

In de beginfase van het project heeft het Testbedteam tijd nodig gehad om elkaars verschillende disciplines te leren kennen. Dit was wel eens moeilijk, maar het leverde uiteindelijk wel de kwaliteit aan dit advies. De multi-disciplinaire aanpak is terug te zien in deze publicatie, immers ook binnen uw organisatie moeten verschillende medewerkers met uiteenlopende achtergrond met elkaar aan de slag.

Het werk van Testbed had niet kunnen worden gedaan zonder de actieve hulp en steun, niet alleen van de enthousiaste teamleden, maar ook van vele anderen in binnen- en buitenland. Ook de departementen Verkeer en Waterstaat, VROM, LNV en Binnenlandse Zaken en Koninkrijksrelaties hebben een steentje bijgedragen door materiaal af te staan waarmee wij konden experimenteren.

Er is heel wat te doen voor overheden die op een verantwoordelijke manier met hun digitale informatie om willen gaan. Testbed heeft gepoogd zo concreet mogelijk aan te geven welke (technische) oplossingen het meest voor de hand liggen en welke acties de verschillende partijen moeten nemen. Ik hoop dat deze publicatie hierbij de nodige houvast zal bieden.

Jacqueline Slats
Programma Manager
Testbed Digitale Bewaring

Leeswijzer

Deze publicatie van *Van digitale vluchtigheid naar digitaal houvast* bestaat uit vier afzonderlijk te lezen delen, waarvan u nu deel 4 *Bewaren van e-mail* in handen heeft. De delen 3 tot en met 1 verschijnen in aflopende volgorde in de loop van 2003. De titels van die delen zijn:

Deel 3: Bewaren van tekstdocumenten;

Deel 2: Bewaren van spreadsheets;

Deel 1: Kosten- en beslissingsmodellen/Functionele specificaties
Bewaren van databases.

Deze publicatie is geschreven voor iedereen die betrokken is bij het goed beheren en bewaren van digitale informatie bij de overheid. Testbed heeft geprobeerd het gebruik van jargon zoveel mogelijk te vermijden, of waar dat niet kon in ieder geval uit te leggen. De acties die de verschillende personen binnen een organisatie moeten ondernemen om digitale informatie nu en straks goed te bewaren zijn per doelgroep uitgesplitst en zijn gemakkelijk te vinden via de tabbladen.

Deel 1 van de reeks is het sluitstuk van het onderzoek dat door Testbed is verricht naar het bewaren van digitale informatie. Dit deel verschijnt als laatste in de loop van 2003 en completeert de serie, omdat het aanvullende informatie bevat over alle delen, zoals kosten- en beslissingsmodellen en functionele specificaties.

Dit deel 4, over het bewaren van e-mail, is als volgt opgebouwd. Hoofdstuk 1 is een inleidend hoofdstuk over de digitale overheid, een probleemschets van digitaal bewaren en de opdracht voor Testbed Digitale Bewaring om via praktische experimenten de juiste bewaarstrategie vast te stellen.

In hoofdstuk 2 kunt u lezen waarin digitale documenten zich onderscheiden van papieren documenten. Ingegaan wordt op de specifieke eigenschappen van digitale documenten, waarbij de vijf hoofdkenmerken van een digitaal document worden toegelicht, te weten inhoud, context, structuur, opmaak en gedrag.

Hoofdstuk 3 gaat in op het documenttype dat in deze uitgave centraal staat, te weten e-mail. Wat is e-mail precies en welke authenticiteitseisen zijn relevant? Oftewel waaraan moeten e-mailberichten voldoen om 'niet aan echtheid in te boeten', zodat het voor een ieder helder is dat het e-mailbericht is wat het beweert te zijn.

In hoofdstuk 4 wordt u meegenomen op een onderzoek langs de verschillende bewaarstrategieën die wereldwijd in de belangstelling staan. Testbed beoordeelt de strategieën in relatie tot e-mail.

Hoofdstuk 5 gaat vervolgens in op die bewaarstrategie die op basis van ons onderzoek als het meest veelbelovend uit de bus is gekomen voor e-mail, te weten XML (Extensible Markup Language). Tevens wordt een wijze van implementeren beschreven voor het duurzaam bewaren van e-mail in XML.

In hoofdstuk 6 vindt u een concreet actieplan voor de verschillende doelgroepen binnen een (overheids)organisatie, te weten managers, records managers, ICT-ers en eindgebruikers. Iedere doelgroep heeft zijn eigen verantwoordelijkheden in deze en krijgt op deze manier die informatie aangereikt waarmee hij of zij een steentje kan bijdragen aan het bouwen van een betrouwbare digitale overheid.

De publicatie wordt besloten met een verklarende woordenlijst, een bibliografie en vier appendices met de volgende onderdelen:

- Appendix A: Technische beschrijving van de e-mail/XML demo;
- Appendix B: XML schema, behorend bij de e-mail/XML demo;
- Appendix C: Bewaar Transactie Logbestand en
- Appendix D: Verschillende representaties van een e-mailbericht (HTML-representatie, transmissiebestand, platte tekst en XML-representatie).

1. De Nederlandse digitale overheid

Er zijn de laatste jaren grote ambities als het gaat om een beter presterende overheid. Op veel fronten is de digitale overheid in opbouw en zijn er uiteenlopende initiatieven op lokaal, regionaal en nationaal niveau. Digitaal bewaren krijgt echter niet altijd de aandacht die het verdient. Actie is nodig want: een digitale overheid *kán* niet zonder digitaal geheugen.

1.1 Ontwikkelingen binnen de digitale overheid

De Nederlandse overheid werkt steeds meer met digitale documenten en bestanden. Het kabinet Kok II formuleerde als doelstelling dat in 2002 25% van de transacties tussen overheid en burger langs digitale weg moet kunnen plaatsvinden, een doelstelling die destijds ruimschoots is gehaald. Inmiddels zijn er nieuwe streefcijfers door de overheid genoemd; eind 2006 moet 65% van alle transacties tussen overheid en burger op elektronische wijze kunnen worden afgehandeld. Het noemen van dit streefcijfer past in het beeld van een effectief opererende overheid, waarin regelgeving vereenvoudigd is en de bureaucratie tot een minimum beperkt. Dit beleid, door minister Remkes van Binnenlandse Zaken en Koninkrijksrelaties samengevat als Beter Bestuur voor Burger en Bedrijf staat of valt met de juiste inzet van ICT binnen de overheid.

De voordelen van het digitaal werken zijn, voor zover ze nog ter discussie staan, zeer groot. Digitale informatie is ten eerste beter *toegankelijk*, voor burgers, maar ook voor andere overheden. De betekenis van het world wide web, *www*, als informatiebron is hierbij ook van belang. Overheden zijn voorts, als zij hun digitale informatie goed beschikbaar maken, beter te *controleren*, bijvoorbeeld door de Algemene Rekenkamer, of Inspecties. Zij kunnen in principe *beter werk* afleveren, omdat informatie vollediger beschikbaar is en bijvoorbeeld vaker dan een keer gebruikt kan worden. De *dienstverlening* aan de burger kan sneller, en *beter*, denk bijvoorbeeld aan het aanvragen van officiële documenten, of het in kaart brengen van riskante bedrijven in een regio (zoals de provincie Friesland dat doet op hun website www.fryslan.nl) om burgers en bedrijven beter te informeren. Tot slot levert digitaal werken niet alleen organisatorische voordelen op, maar ook financiële. Miljoenen euro's kunnen zodoende bespaard worden¹.

Nu is iedereen zo langzamerhand wel overtuigd van de voordelen van een digitale overheid, maar de problemen ervan zijn soms moeilijk te onderkennen of aan te pakken. Meer digitale transacties tussen de burger en de overheid, betekent een omvangrijke omslag in de backoffices, oftewel de informatiehuishouding, van de overheidsorganisaties. Naast het goed functioneren van de backoffice is de transparantie in het werk en de blijvende toegankelijkheid van informatie een probleem dat dringend oplossing behoeft. Dit laatste punt, de blijvende toegankelijkheid van digitale informatie, wordt hieronder verder uitgelicht.

¹

Zie de publicatie *Winst met ICT in uitvoering*, A. Zuurmond, K. Mies; Zenc, Den Haag, juni 2002.

1.2 *Bewust werken: digitaal duurzaam beheren*

Het feit dat de overheid niet alleen op papier, maar ook digitaal moet gaan bewaren, dringt bij steeds meer organisaties door. Digitaal duurzaam werken is het credo. Dit staat voor het op zodanige wijze vastleggen, bewaren, beheren en beschikbaar stellen van digitale documenten en bestanden dat deze ook na verloop van tijd raadpleegbaar, toegankelijk en authentiek zijn.

Digitaal duurzaam beheren is geen kwestie van techniek alleen. Overheidsorganisaties moeten (voor zover zij dat nog niet doen) het probleem van digitale duurzaamheid onderkennen en bereid zijn hier iets aan te doen. Dat betekent budget en aandacht vrijmaken; beleid, regels en procedures formuleren en uitvoeren; (technische) hulpmiddelen aanschaffen en implementeren; medewerkers opleiden en instrueren. Ook de individuele medewerkers moeten de noodzaak van beleid, regels en procedures onderkennen en bereid zijn deze na te leven. Dat zal alleen het geval zijn, als deze hen niet of nauwelijks hinderen bij de uitvoering van hun normale werk en als de ondersteunende technische hulpmiddelen het hen gemakkelijk maken.

Voorts is het belangrijk dat overheidsorganisaties volop kunnen kiezen uit op de markt beschikbare softwaretoepassingen waarin de duurzame bewaring van de teksten, images, beelden, geluiden en combinaties hiervan, van meet af aan (dat wil zeggen al bij de creatie van die informatie) geïntegreerd zijn opgenomen.

1.3 *Digitaal werken betekent ook digitaal bewaren*

De overheid heeft met papieren documenten en registraties enkele eeuwen ervaring opgebouwd; digitale documenten kent zij echter pas sinds enkele decennia. Door de specifieke eigenschappen van digitale documenten kan de papieren oplossing niet worden gebruikt (in het volgende hoofdstuk wordt hierop verder ingegaan). Op bepaalde punten is digitale informatie wezenlijk anders dan papieren informatie. Digitale documenten hebben geen vaste vorm en zij worden vaak gemaakt door meerdere personen. In het verleden zorgden speciale archiefafdelingen ervoor dat documenten overeenkomstig de wet en de functieverantwoordelijkheden werden beheerd. Tegenwoordig beschikken overheidsmedewerkers dankzij ICT over allerlei nieuwe mogelijkheden om documenten te maken, variërend van tekstbestanden en e-mailberichten tot spreadsheets en databases. Daarbij is het beheer van die documenten steeds verder onttrokken aan het toezicht van de daarvoor verantwoordelijke afdeling. De bestaande procedures en regels voor papieren documenten worden niet toegepast op digitale documenten, die daardoor een riskant bestaan leiden.

Hoewel deze witte vlek in de procesgang te maken heeft met het leerproces bij de overgang van papieren naar digitale documenten, mag deze ontwikkeling zich niet voortzetten. Ook in het digitale tijdperk moeten documenten worden gemaakt die de tand des tijds kunnen doorstaan. Ook dienen zij goed beheerd te worden. Met de meeste documenten die tegenwoordig worden gemaakt, inclusief e-mailberichten, is dat niet het geval.

Enerzijds heeft het probleem dus te maken met de informatiehuishouding binnen organisaties. Anderzijds ligt het probleem van het behouden van digitale documenten in de snelle veroudering van hard- en software. Wordt hier niets aan gedaan, dan zal digitale informatie verloren gaan, omdat deze niet meer leesbaar en toegankelijk is. De termijn waarover we spreken is kort: al na twee jaar of meer kan informatie waardeloos zijn geworden.

De consequenties hiervan kunnen zijn dat belangrijke informatie verdwijnt en het bijvoorbeeld niet meer mogelijk is een besluitvormingsproces van de overheid te reconstrueren. Een recent voorbeeld hiervan is te vinden in het parlementaire enquête over Srebrenica door de commissie Bakker (januari 2003). Getuigenverklaringen worden soms per e-mail afgelegd, maar hoe moeten deze bewaard worden? Het is niet genoeg ze uit te printen. Een formeel digitaal document moet immers digitaal bewaard worden (zie ook hoofdstuk 2 voor de uitwerking hiervan).

Voorts is e-mail een mooi voorbeeld van een 'vluchtig' digitaal communicatiemiddel. Als de gemiddelde medewerker het verzoek van de ICT-afdeling krijgt zijn postbus te legen omdat de limiet is bereikt, dan doet hij dat trouw met een druk op de knop zonder zich te realiseren dat bepaalde e-mail echt bewaard moet blijven.

Een ander voorbeeld betreft het terugvinden van informatie, zoals bij de vraag hoeveel werklozen een uitvoeringsorganisatie de afgelopen decennia aan het werk heeft geholpen? Deze vraag zal niet kunnen worden beantwoord, als de informatiehuishouding niet op orde is, of niet goed ontsloten is. Dit onderwerp stond overigens centraal op het symposium dat Taskforce Digitale Duurzaamheid samen met de dienst Arbeidsvoorziening in november 2002 organiseerde. Goed bewaren (ook langdurig), terugvinden en hergebruiken van digitale gegevens zijn kortom de sleutelwoorden.

Digitale dienstverlening door de overheid is in opbouw. Men kan zich afvragen of een digitale vergunning die een gemeente heeft afgegeven, na vijf jaar en drie conversies naar nieuwere software, nog precies dezelfde betekenis heeft?

Kortom, de hierboven geschetste voorbeelden grijpen direct in op het functioneren van de overheid. De continuïteit van de bedrijfsvoering, de externe verantwoording van de overheid en de toekomstige generaties die onderzoeken hoe de overheid heeft gefunctioneerd: dit alles kan alleen als er een goede, betrouwbare bewaarmethode voor digitale informatie is.

1.4 *Digitaal bewaren en de wet*

De overheid heeft deels het belang van digitale bewaring onderkend, en de bestaande wetgeving op bepaalde punten aangepast. Hier volgt een beknopt overzicht van deze wetten en richtlijnen.

De Archiefwet 1995

Elk document dat een functie vervult bij de taakuitoefening is in beginsel een archiefstuk of archiefbescheid. Formeel, of niet formeel, alle vormen van elektronische post van de overheid hebben in potentie betrekking op haar handelen en komen dus voor bewaring in aanmerking. Ook e-mail kan dus een archiefstuk zijn.

De Regeling geordende en toegankelijke staat archiefbescheiden (2002).

De Regeling geordende en toegankelijke staat archiefbescheiden is een nadere uitwerking van artikel 12 uit het Archiefbesluit 1995. De Regeling geeft als belangrijkste eisen dat archiefbescheiden authentiek moeten zijn en dat archiefbescheiden binnen redelijke tijd vindbaar en leesbaar moeten zijn. Voor digitale archiefbescheiden, waaronder e-mail, gelden extra eisen. Het gaat dan onder meer om het vastleggen van metagegevens over inhoud, vorm en structuur van een document, over technische gegevens, conversie, migratie en opslagformaten.

De Wet Openbaarheid van Bestuur (WOB) (1998)

Als archiefbescheiden van overheidsorganisaties zijn overgebracht naar een archiefbewaarplaats, zijn deze op grond van de Archiefwet 1995 in principe openbaar.

Zolang archieven nog bij overheidsorganisaties zijn ondergebracht, is de openbaarheid anders geregeld. In die gevallen geldt namelijk de WOB. De WOB geeft iedereen het recht om bij een overheidsorgaan informatie op te vragen. Daarbij wordt net als bij de Archiefwet geen onderscheid gemaakt naar het type informatiedrager van het document, of het nu op papier is, of digitaal.

Wet Bescherming Persoonsgegevens (2001)

Ook de Wet Bescherming Persoonsgegevens is aangescherpt wat het gebruik betreft van persoonsgegevens. Wat geldt voor deze gegevens op papier, is ook van toepassing op de digitale versie.

Richtsnoer E-mail gebruik ten behoeve van de rijksoverheid (2001)²

Het richtsnoer E-mailgebruik is geen wet of regeling. Een werkgroep van verschillende departementen heeft basisregels voor het omgaan met e-mail opgesteld. Dit was nodig omdat de communicatie via e-mail steeds meer status krijgt. Burgers en maatschappelijke organisaties die via e-mail met de overheid willen communiceren, moeten hiertoe in de gelegenheid worden gesteld. Aan een e-mail van de overheid kan de ontvanger rechten ontleen die overeenkomen met de rechten van berichten op papier, ook al denken veel overheden dat zij dit met een disclaimer kunnen ondervangen.

Samenvattend kan worden gesteld dat bewustwording van organisaties en de medewerkers een voorwaarde is om informatie goed te kunnen bewaren, vooral in het digitale tijdperk. Op het gebied van wetgeving is het een en ander al in gang gezet. De vraag is nu of de techniek niet een eenvoudige oplossing kan bieden voor het goed bewaren voor vandaag en morgen.

1.5 Een technische oplossing voorhanden?

Over de hele wereld zijn ICT-experts en wetenschappers bezig met het zoeken naar het antwoord op de vraag hoe digitale informatie het best kan worden bewaard. Er zijn verschillende strategieën bekend die goede perspectieven lijken te bieden om op een verantwoorde en duurzame manier met de digitale neerslag van het overheidshandelen om te kunnen blijven gaan. In hoofdstuk 4 wordt uitgebreid op deze strategieën ingegaan.

Het probleem van dit moment is dat er geen pasklare oplossing is voor overheidsorganisaties die werkelijk aan hun digitale geheugen willen gaan bouwen. Welke bewaarstrategie een organisatie moet kiezen, welke voorzieningen er moeten worden aangeschaft, dat zijn vragen waar nu geen antwoord op is. De meeste strategieën zijn in de praktijk ook nog niet eerder getoetst.

Om voor deze situatie oplossingen te onderzoeken hebben het ministerie van Binnenlandse Zaken en Koninkrijksrelaties en het ministerie van Onderwijs, Cultuur en Wetenschappen, in casu de Rijksarchiefdienst, besloten een 'testbed' in te richten om aan de hand van experimenteel onderzoek kennis en ervaring op te doen met het duurzaam bewaren van verschillende digitale documenten: Testbed Digitale Bewaring.

Testbed Digitale Bewaring, gestart in 2000, voert experimenten uit aan de hand van oplossingsgerichte onderzoeksvragen, om zo de beste bewaarstrategie of combinatie van strategieën te kunnen vaststellen. Testbed richt zich daarbij op drie verschillende, vooral theoretische, methoden voor het langdurig bewaren van digitale informatie, namelijk migratie, XML en emulatie. Van deze methoden zal niet alleen de effectiviteit

² *Richtsnoer E-mailgebruik t.b.v. de Rijksoverheid* /Ministerie van BZK; werkgroep e-mailgebruik, Den Haag 2001.

worden beoordeeld, maar ook de beperkingen, kosten en gebruiksmogelijkheden. Testbed houdt hierbij rekening met de wettelijke en beleidsmatige context zoals die hierboven is geschetst.

Het team van Testbed Digitale Bewaring bestaat uit een internationale mix van deskundigen op het gebied van archiveren, ICT, informatie management en communicatie.

1.6 De opdracht voor Testbed Digitale Bewaring

Het Testbedteam is met de opdracht van de departementen aan de slag gegaan. Om de wetenschappelijke theorieën te kunnen beoordelen is een unieke laboratoriumachtige omgeving gebouwd, gebruikmakend van een zelf ontworpen en gebouwd systeem, waarin alle onderzoeksgegevens zijn ondergebracht. De experimenten en tests die worden uitgevoerd zijn volledig reproduceerbaar en wetenschappelijk verantwoord. De aanbevelingen zijn voor iedereen toegankelijk op de website www.digitaleduurzaamheid.nl

Het project Testbed zal de volgende producten en diensten opleveren:

- o Kennis en inzicht in technische oplossingen voor de langetermijnbewaring van digitale documenten;
- o Adviezen hoe om te gaan met huidige digitale records;
- o Onderbouwde strategie(ën) voor de langdurige bewaring van vier typen digitale documenten, te weten tekstdocumenten, spreadsheets, e-mail en databases.
- o Functionele eisen voor een bewaarsysteem voor digitale documenten: de functionele specificaties voor het bouwen van een bewaarfunctie binnen een archiefsysteem.
- o Kostenmodellen van de verschillende bewaarstrategieën: Wat zijn de kostenindicatoren bij het implementeren van een bepaalde bewaarstrategie?
- o Beslissingsmodel bewaarstrategieën. Een hulpmiddel waarmee kan worden vastgesteld wat, gegeven een bepaald documenttype, de meest geschikte bewaarstrategie is.
- o Voorstellen voor aanpassing van de huidige wet- en regelgeving.

In dit deel van de serie *Van digitale vluchtigheid naar digitaal houvast* wordt met name ingegaan op de eerste drie bovengenoemde punten.

2. *Digitale documenten en authenticiteit*

Wat maakt digitale documenten nu zo bijzonder? In dit hoofdstuk wordt ingegaan op de eigenschappen en kenmerken van digitale informatie. Ook het kernbegrip 'authenticiteit' komt aan bod. Want de essentie is dat een document gegarandeerd authentiek moet zijn: het eenmaal bewaarde document mag niet worden gewijzigd.

2.1 Definitie van een digitaal document

Digitale documenten zijn niet simpelweg het 21ste-eeuwse equivalent van traditionele papieren documenten. Ze hebben andere eigenschappen, kenmerken en toepassingen. Echter zowel digitale als papieren documenten moeten aan dezelfde wettelijke eisen voldoen. Dit vergt in de praktijk een andere aanpak.

Digitale documenten zijn geen tastbare objecten, zoals een boek of een tijdschrift, maar een combinatie van hardware, software en computerbestanden. Deze combinatie is noodzakelijk om de documenten te kunnen gebruiken, of waar te nemen. Binnen het kader van Testbed is specifiek gekeken naar tekstuele documenten, databases, e-mailberichten en spreadsheets. Ook multimedia documenten, digitale video en geluid kunnen digitale documenten genoemd worden, maar deze vallen buiten de scope van dit onderzoek.

Een belangrijk verschil met papieren documenten is het grotere verlies van informatie dat zich al tijdens het gebruik van digitale documenten kan voordoen, of naderhand, als de documenten bewaard zijn. Harde schijven en computers worden immers regelmatig vervangen en er is een lage drempel om computerbestanden te vernietigen. Met één druk op de knop 'delete' kan een document verdwijnen, zonder dat er een spoor van achterblijft.

Om het probleem van technologische veroudering te analyseren, en geschikte bewaarstrategieën te toetsen, onderscheidt Testbed drie aspecten van digitale documenten:

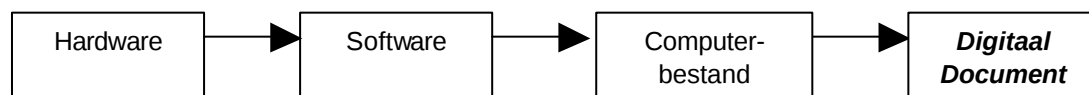
- o het begrip 'digitaal document' als een combinatie van hardware, software en computerbestanden;
- o het begrip 'authenticiteit' bij digitale documenten;
- o metadata ter veiligstelling van authenticiteit bij digitale documenten.

Deze aspecten worden in de onderstaande paragrafen verder uitgewerkt.

Het digitale document als combinatie van hardware, software en computerbestanden
In het papieren tijdperk was het begrip 'document' eenvoudig. Het document als bewijs van een transactie werd vastgelegd op een fysieke eenheid zoals perkament of papier, mogelijk in de vorm van een charter, bon, brief, memo of foto. Het bewijsstuk werd volgens vaststaande en vooraf bepaalde procedures bewaard. Wie het nodig had, nam contact op met het archief en vroeg om het document.

In het digitale tijdperk wordt een document niet op een dergelijke manier vastgelegd. Digitale documenten moeten technologisch worden verwerkt voordat de gebruiker ze kan lezen en voor het gewenste doel kan toepassen. Juist deze afhankelijkheid van hardware en software noopt ons anders te gaan denken over de manier waarop we documenten maken en gebruiken.

Onderstaand schema toont de componenten die noodzakelijk zijn om het digitale document weer te geven en te gebruiken³:



Een digitaal document wordt gemaakt met behulp van een bepaalde combinatie van hardware en software en opgeslagen in gecodeerde vorm, het computerbestand. Dit computerbestand bestaat uit een reeks nullen en enen en is ruwe data in de puurste vorm. Deze reeks nullen en enen wordt door een bepaald programma gelezen en op een manier geïnterpreteerd die vaak uniek is voor dat programma. Het resultaat van die interpretatie wordt vervolgens weergegeven op het scherm en die weergave wordt het digitale document genoemd.

In de meeste gevallen kan het computerbestand alleen correct worden gelezen via de eerdergenoemde combinatie van hardware en software. Als het digitale document in een andere computeromgeving wordt weergegeven dan die waarin het oorspronkelijk is gemaakt, kan het digitale document er heel anders uitzien en zich ook anders gedragen. Als die overgang naar een andere computeromgeving niet wordt beheerst, kan de authenticiteit van het digitale document worden aangetast.

2.2 Authenticiteit als kernbegrip

Authenticiteit is een kernbegrip bij het bewaren van documenten. Authenticiteit betekent dat een document is wat het zegt te zijn. Het mag niet ongeoorloofd veranderd of gecorrumpeerd worden. Een besluit dat bijvoorbeeld door het parlement is genomen, is een papieren document, waarop datum en namen van de partijen zijn weergegeven. Dit maakt het document tot een waardevolle bron van kennis. Niets aan dit Kamerstuk mag veranderd worden als het eenmaal gemaakt is. Mochten er toch wijzigingen in aangebracht worden, dan zal dit in de regel eenvoudig te zien zijn.

Van een digitaal document is het minder gemakkelijk vast te stellen of het authentiek is. De problemen die dat kan opleveren mogen niet onderschat worden. In september 2001 bijvoorbeeld verkeerde het CDA in een interne crisis. Een beleidsmedewerker van de CDA-fractie in de Tweede Kamer speelde een cruciale rol door een digitaal rapport zo te bewerken dat het leek of opinie-onderzoek uitwees dat fractieleider De Hoop Scheffer een zwak imago zou hebben. Dit document werd doorgespeeld aan een actualiteitenrubriek. Tegen de tijd dat ontdekt werd dat dit document niet authentiek was, was de schade al niet meer te herstellen, en traden zowel Van Rijn, de voorzitter, als De Hoop Scheffer af. Het kostte de CDA-fractie de nodige moeite om de dader te achterhalen. Een extern automatiseringsbedrijf heeft alle personal computers moeten onderzoeken om de dader op te sporen, hetgeen uiteindelijk gelukt is.

Volgens de definitie van Testbed en Depot 2000 is authenticiteit 'de weergave van een document geheel conform de oorspronkelijke vastgelegde versie en de beoogde functie van het document'. Bij authenticiteit staan twee begrippen centraal:

³

InterPARES Authenticity Task Force Final Report,
http://www.interpares.org/book/interpares_book_d_part1.pdf

Verificatie: dat het document is wat het zegt te zijn. Verificatie stelt ons in staat te bevestigen dat een document, digitaal of niet, is wat wij denken dat het is en dat het door de aangegeven personen is gemaakt. Authenticiteit van documenten kan gewaarborgd worden door de oorspronkelijke context van die documenten vast te leggen en te bewaren en door een ononderbroken beheer ('unbroken custody').

Integriteit: het document is intact en niet zodanig veranderd of gecorrumpeerd dat de betekenis ervan niet meer duidelijk is. Met integriteit wordt het digitale document als geheel bedoeld. Dit houdt in dat het document ongewijzigd en intact moet blijven. Wijzigingen zijn tot op zekere hoogte aanvaardbaar, zolang de oorspronkelijke betekenis of functie van het document er niet door wordt aangetast. Een voorbeeld hiervan is de eerder genoemde website van de provincie Friesland met kaarten waarop de positie van risicovolle bedrijven in kleur is aangegeven. Kleuren hebben op die kaarten een belangrijke betekenis en dienen dus bewaard te blijven in de oorspronkelijke staat. Een conversie van dit document naar een hogere versie die kleurverschil geeft (rood wordt groen bijvoorbeeld), tast de integriteit van het document aan.

In wezen maakt het geen verschil of een document een digitale vorm heeft, of een fysieke. Het probleem dat een digitaal document met zich meebrengt is echter dat door de veranderende technologie niet alle aspecten van een document precies zo bewaard kunnen worden als toen het werd gemaakt. Dit betekent echter niet dat het duurzaam bewaren van authentieke digitale documenten onmogelijk is.

2.3 *Digitale documenten, digitale kenmerken*

In het papieren tijdperk vormen de kenmerken van een document fysiek een eenheid. Context, inhoud, structuur en uiterlijk maken een document authentiek. Als één eigenschap wordt veranderd, heeft dat invloed op de andere. Zo wordt de structuur van het document, bijvoorbeeld de indeling in hoofdstukken, gerepresenteerd in het uiterlijk. Het uiterlijk van het document, bijvoorbeeld een complete publicatie met tabbladen, bevat op zijn beurt weer de totale inhoud van het document, met daarin bovendien besloten veel verwijzingen naar de context, zoals de naam van de auteur, of de datum van publicatie. Al deze aspecten van het document, dus context, inhoud, structuur en uiterlijk, staan vast en kunnen na publicatie van het document niet meer worden gewijzigd.

Digitale documenten zijn anders. Ze hebben weliswaar nog steeds de vier hierboven genoemde kenmerken, maar kunnen ook nog een ander kenmerk hebben: gedrag⁴. In tegenstelling tot papieren documenten zijn bij digitale documenten de kenmerken echter niet zo onwrikbaar met elkaar verbonden. Zij zijn sterk afhankelijk van de wijze waarop de software het computerbestand interpreteert. Dat maakt dat ze veel sterker onderhevig kunnen zijn aan ongewilde wijziging. Het bewaken van die kenmerken en de samenhang ervan vereist dus extra maatregelen.

In de Nederlandse wet- en regelgeving wordt gesproken over context, inhoud, structuur en vorm. Het kenmerk gedrag, dat zo wezenlijk kan zijn voor digitale documenten, wordt niet genoemd. Daarnaast wordt in de huidige regelgeving onder het begrip vorm verstaan: "de uiterlijke verschijning waarin de structuur en opmaak zichtbaar zijn"⁵.

⁴ Rothenberg, Jeff & Bikson, Tora: *Carrying Authentic, Understandable and Usable Records Through Time*, Den Haag, 1999.

⁵ Zie artikel 1, lid 1, sub o van de Ministeriële regeling Geordende en toegankelijke staat archiefbescheiden.

Voor zijn onderzoek heeft Testbed het kenmerk vorm uit elkaar getrokken en worden structuur en uiterlijk als afzonderlijke kenmerken van een digitaal document onderscheiden. Hieronder worden de vijf kenmerken van digitale documenten nader toegelicht.

Context⁶

Met context wordt de omgeving bedoeld waarin het digitale document is gemaakt. Om het document betekenis te geven, is een bepaalde hoeveelheid informatie over de context nodig. Deze informatie betreft alleen het document, los van het medium en omvat dan ook niet de technische omgeving waarin het document wordt gemaakt en gebruikt. Het gaat hierbij onder meer om informatie over het bedrijfsproces en het overheidsorgaan in welk kader het digitale document is ontvangen of opgesteld. Daarnaast moet de relatie met andere documenten, onder andere uit hetzelfde bedrijfsproces, worden vastgelegd en bewaard. Het gaat dan bijvoorbeeld om dossiers. Dit wordt ook wel aangeduid met het begrip externe structuur.

Inhoud

De inhoud is de platte tekst, ongeacht welke structuur, kleur, positie of lettertype het heeft. Bij e-mail omvat de inhoud de specifieke tekst die in het berichtvenster is getypt, de zogeheten body. Deze begint gewoonlijk met een aanhef en eindigt met een afsluiting (groet, naam etc).

Structuur

Met de structuur van een digitaal document wordt de structuur bedoeld zoals deze oorspronkelijk op het scherm is gemaakt en weergegeven. Het gaat hier om de logische hiërarchie van, en de relaties tussen, de delen van een document. De structuurelementen van een e-mail zijn bijvoorbeeld de headers, de berichttekst (de inhoud) en eventuele bijlagen. De structuurelementen van een rapport (een tekstdocument) daarentegen kunnen worden gevormd door een voorblad, een inhoudsopgave, hoofdstukken (opgesplitst in paragrafen en alinea's) en een bibliografie en/of appendix.

Het is belangrijk dat deze structuurelementen correct worden geïdentificeerd en dat de e-mail of het rapport in de juiste volgorde wordt weergegeven. Bovendien is het belangrijk te weten of er andere wezenlijke structuurkenmerken zijn, bijvoorbeeld de aanwezigheid van voet- en eindnoten in een document. Als deze structuur tengevolge van een migratie verloren gaat, kan de tekst verkeerd worden weergegeven.

Uiterlijk

Met het uiterlijk van het digitale document wordt de uiteindelijke presentatie bedoeld, hoe het digitale document er uitziet als het op het scherm verschijnt. Het uiterlijk omvat bijvoorbeeld aspecten als het lettertype en lettergrootte. Ook pagina- en sectie-einden en de grootte van marges kunnen het uiterlijk van het digitale document beïnvloeden. Daarnaast kan kleur een uiterlijk kenmerk zijn; dit kan, zoals hierboven al werd aangegeven, soms van invloed zijn op de betekenis van een document.

Gedrag

Het gedrag van een digitaal document is het moeilijkst te bewaren. Met het gedrag worden de interactieve kenmerken van een document bedoeld; datgene wat ons in staat stelt het document te bewerken en te gebruiken, zodat nieuwe of aanvullende inhoud ontstaat. Het gedrag kan betrekking hebben op de omgeving waarin het digitale document wordt 'gehost' of geplaatst. Als het gedrag echter wezenlijk deel uitmaakt van het digitale document zelf en dus moet worden bewaard, komt het meestal voort uit de inhoud van het document. Voorbeelden zijn het opnemen van

⁶

Een uitdijend heelal? Context van archiefbescheiden, H. Hofman, Stichting Archiefpublicaties, Jaarboek 2000.

e-mailadres in een document (bijv. remco.verdegem@ictu.nl) of een verwijzing naar een website in de vorm van een URL zoals www.ictu.nl

De vijf hierboven genoemde kenmerken context, inhoud, structuur, uiterlijk en gedrag maken het digitale document bijzonder. Opgemerkt moet worden dat de kenmerken niet voor alle typen digitale documenten even belangrijk zijn. Zo mag men ervan uitgaan dat uiterlijk minder belangrijk is voor e-mailberichten dan voor tekstdocumenten. De kenmerken spelen een belangrijke rol bij het toetsen van de verschillende bewaarstrategieën dat in hoofdstuk 4 aan de orde komt.

Metadata

Metadata zijn gegevens over gegevens. We voegen metadata toe aan een digitaal document om aanvullende informatie vast te leggen over de vijf eerder genoemde kenmerken van het document, zodat onder meer kan worden nagegaan of het document is wat het 'zegt' dat het is. Tevens maken metadata het mogelijk een bepaalde digitaal document terug te vinden en te gebruiken. Hierbij kan gedacht worden aan gegevens als de auteur van het document, het onderwerp, het bedrijfsproces waarin het document tot stand is gekomen en de datum waarop het document tot stand is gekomen. Maar ook in het kader van het registreren van uitgevoerde bewaaracties zijn metadata belangrijk.

Er zijn verschillende soorten metadata te onderscheiden, bijvoorbeeld:

- o Archivering metadata ('record keeping' metadata). Deze groep metadata richt zich met name op contextuele gegevens die het digitale document betekenis geven, maar kunnen ook essentiële kenmerken van het digitale document bevatten, bijvoorbeeld de aanwezigheid van een hyperlink naar een bepaalde website (gedragkenmerk);
- o Technische metadata. Deze metadata beschrijven de oorspronkelijke technologische omgeving, waarin het digitale document is ontstaan en gebruikt (hard- en softwareomgeving) alsmede de nieuwe technische omgeving na bijvoorbeeld een migratie;
- o Bewaar metadata. Deze groep van metadata registreren de uitgevoerde bewaaracties om een digitaal document leesbaar en toegankelijk te houden (bijvoorbeeld een uitgevoerde migratie of omzetting naar XML), alsmede eventuele wijzigingen in het digitale document die het gevolg zijn van de uitgevoerde bewaaractiviteit.

Met de metadata kunnen we ons een beeld vormen van het digitale document zonder dat we het bewuste document hoeven weer te geven. Metadata maken deel uit van het digitale document; ze vergezellen een digitaal document gedurende zijn levenscyclus en bevatten informatie over de creatie van het digitale document en over uitgevoerde bewaaracties. Metadata zijn dus van cruciaal belang.

Op basis van de aanwezige metadata kunnen immers de juiste bewaarmaatregelen worden genomen en tevens kan worden nagegaan of wezenlijke elementen van het digitale document na bijvoorbeeld een migratie nog dezelfde zijn en of het document niet is aangetast. Metadata zijn dus deel van het bewijs dat een document authentiek is.

3. *E-mail bewaren in authentieke staat*

E-mail heeft razendsnel een plaats veroverd als snel communicatiemiddel, zowel binnen als buiten de overheid. Het uitgangspunt is dat ook formele e-mail in zijn oorspronkelijke vorm bewaard moet worden. Om dit te kunnen doen, is het nodig het wezenlijke van e-mail te beschrijven en authenticiteitseisen te formuleren.

3.1 *De vlucht van e-mail: problemen bij het bewaren*

In 1971 werd de eerste e-mail verstuurd door ingenieur Ray Tomlinson, werkzaam bij het Amerikaanse computerbedrijf BBN Technologies. Weinigen dachten op dat moment dat e-mail zo'n vlucht zou nemen. De intrede van de personal computer, de pc, zorgde ervoor dat in het midden van de jaren tachtig het computergebruik enorm toenam. In de jaren negentig raakte e-mail snel ingeburgerd als een compleet nieuw medium. E-mail wordt nu veelvuldig gebruikt voor zowel informele als formele berichtgeving binnen de Nederlandse overheid. Anno 2003 bestaan er wereldwijd meer dan een half miljard elektronische postbussen, en is e-mail niet meer weg te denken uit het werkproces.

Mensen in een organisatie maken, ontvangen, gebruiken en verzenden documenten om hun taken te kunnen uitvoeren; de organisatie moet die documenten bewaren en toegankelijk houden omdat ze later nog eens nodig kunnen zijn: voor verantwoording, voor bedrijfsvoering of als kennisbron. De invoering van e-mail systemen verandert de manier waarop mensen binnen en buiten organisaties communiceren en heeft daardoor invloed op besluitvormingsprocessen. Digitale communicatie verloopt sneller en is veel toegankelijker.

De voordelen van de digitale informatie-uitwisseling zijn enorm. Echter de problemen rondom het goed, duurzaam bewaren van digitale documenten zijn ook van toepassing op e-mail. E-mail is een vluchtig medium, dat voor velen geen formele status heeft. Rondom het bewaren van e-mail zijn verschillende factoren te noemen, die elkaar over en weer beïnvloeden. Zij hebben te maken met de organisatie (-cultuur), wettelijke en technische aspecten.

Organisatie en organisatiecultuur

E-mail is een relatief nieuwe toepassing bij de overheid, en de ontwikkelingen staan niet stil. Er is daarom nog weinig ervaring opgedaan met het langdurig bewaren van e-mailberichten. E-mail is bij uitstek een individueel en daardoor moeilijk te controleren communicatiemiddel. Het valt veelal buiten de normale geregistreerde papieren post. Medewerkers bepalen vaak zelf of er wel of niet bewaard wordt en bewaren dan wel vernietigen hun e-mailberichten naar eigen inzicht en willekeur. Ze zien elektronische post ten onrechte als een deel van hun persoonlijk werkdomain. Er bestaan wel richtlijnen voor het gebruik en organisatie van e-mail⁷, maar het is de vraag of deze ook juist toegepast worden. Als e-mailberichten in de praktijk bewaard worden, doet men dat niet op de juiste manier: men print ze namelijk uit. Daarmee

⁷

Zie voor meer informatie over e-mailbeleid: *Archivering van elektronische post. Methoden, meningen en alternatieven*, P. Horsman, Amsterdam 1999 en *Richt snoer E-mailgebruik t.b.v. de Rijksoverheid* /Ministerie van BZK; werkgroep E-mailgebruik, Den Haag 2001.

gaat een deel van de (context) informatie verloren en ook de toegankelijkheid verslechtert. Bovendien is het sowieso niet mogelijk multimediale attachments uit te printen.

Wettelijke aspecten

De bestaande wetgeving geeft wel een kader voor het bewaren van e-mail en andere digitale documenten, zoals de *Regeling Geordende en toegankelijke staat archiefbescheiden*. In hoofdstuk 1 kwamen deze en andere wetten al aan de orde. Probleem is dat het zeer lastig is de regelingen te vertalen naar praktische handvatten. Jurisprudentie over het gebruik van e-mail als bewijsmateriaal is er bijvoorbeeld niet of nauwelijks. Dat hier een schemergebied is, blijkt uit het feit dat veel organisaties disclaimers in hun e-mailbericht opnemen.

Technisch vragen

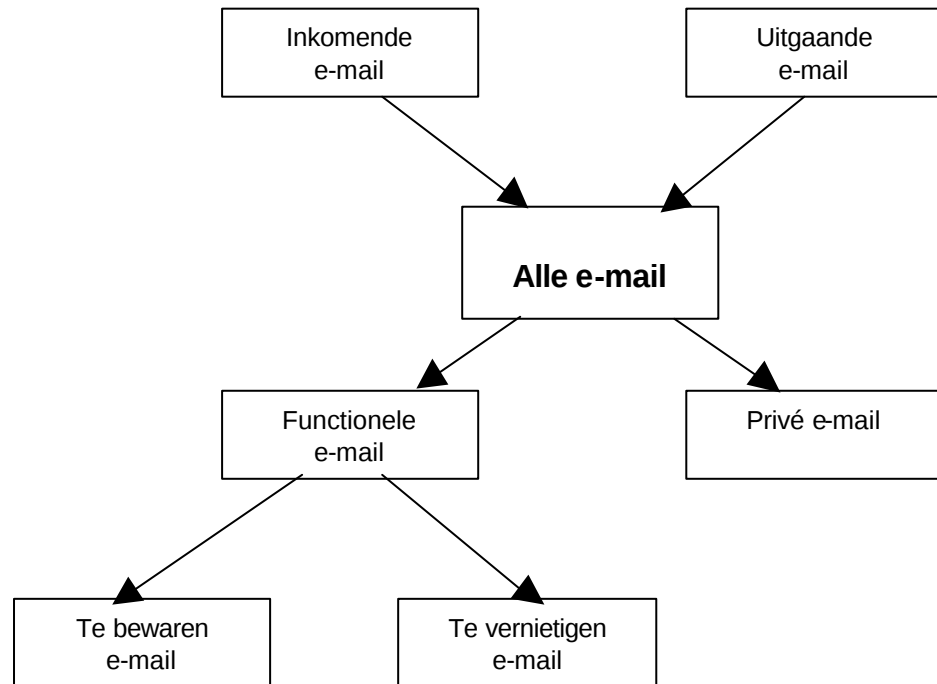
Hard- en software applicaties verouderen snel, waardoor digitale bestanden niet langer toegankelijk zijn. Er is, nationaal en internationaal, weinig praktisch onderzoek gedaan naar technische manieren voor het bewaren van e-mail. Hier en daar zijn er wel projecten, zoals het DAVID project van het stadsarchief te Antwerpen waar onderzoek plaatsvindt, maar dit gebeurt niet op grote schaal. Er is niet alleen het probleem van het bewaren van e-mail zelf, maar ook van de attachments. Moeten deze in hun oorspronkelijke formaat worden bewaard of juist niet? Worden deze bij het e-mailbericht of juist apart opgeslagen?

Als gevolg van de complexiteit van deze problemen en ontwikkelingen worden e-mailberichten niet of nauwelijks bewaard. De oplossing ligt in een aanpak waarin alle genoemde aspecten aan bod komen. Alleen een technische of juridische oplossing is niet voldoende: het gaat ook om bewustwording van het belang van goed bewaren. Om tot een praktische aanpak te komen is het noodzakelijk eerst het eigene van e-mail te beschrijven en authenticiteitseisen te formuleren: waar moet een goed bewaarde e-mail aan voldoen.

3.2 De status van e-mail

Niet alle e-mailberichten behoeven te worden bewaard, zoveel moge duidelijk zijn. De keuze welke e-mailberichten wel of niet voor bewaring in aanmerking komen, hangt onder andere af van de taken van een organisatie waarbinnen het e-mailbericht is aangemaakt of ontvangen. Dit is beschreven in een Rapport Institutioneel Onderzoek (RIO). Het Basis Selectiedocument (BSD), dat gebaseerd is op dit RIO, biedt de grondslag voor het vernietigen dan wel voor blijvende bewaring overbrengen van de neerslag van de overheid.

Zonder hier nu al te diep op deze kwestie in te gaan, kan de volgende indeling⁸ behulpzaam zijn bij het beantwoorden van de vraag welke e-mail bewaard moet worden.



Inkomende of uitgaande e-mail

Dit onderscheid is van een andere aard dan de hieronder genoemde indelingen, maar desondanks wel relevant voor de omgangsregels voor e-mail. Binnen deze categorie kan men ook nog interne en externe e-mail onderscheiden. Hiermee wordt het onderscheid aangegeven tussen de uitwisseling van elektronische berichten binnen een organisatie en de berichten die van of naar een externe organisatie komen of gaan.

Functionele e-mail versus privé e-mail

De e-mail die een medewerker uit hoofde van zijn/haar functie ontvangt of verzendt, is formele e-mail. De e-mail die een medewerker als privé-persoon ontvangt of verzendt en die niet is gerelateerd aan het feit dat de medewerker een functie binnen de organisatie bekleedt, wordt geclassificeerd als privé e-mail.

Te bewaren e-mail versus te vernietigen e-mail

Indien een e-mailbericht functioneel is, dient vastgesteld te worden of deze in aanmerking komt voor bewaring. Hiervoor gelden in principe dezelfde criteria als voor 'normale' papieren post.

Bij de huidige gang van zaken is het zo dat e-mailberichten in eerste instantie door de gebruikte e-mailapplicatie zelf vastgelegd worden in het leverancierseigen formaat, bijvoorbeeld *.msg voor Outlook. Dit is de manier waarop de meeste mensen hun e-mail nu (tijdelijk) bewaren. Als tussenoplossing, of bij gebrek aan beter, printen sommigen hun e-mail ook uit, hetgeen zoals al eerder betoogd, informatieverlies oplevert en de voordelen van e-mail onbenut laat.

⁸ Richtsnoer e-mail gebruik, p. 4.

3.3 Kenmerken van e-mail

Het begrip elektronische post (het equivalent van het Engelse begrip e-mail) dekt twee ladingen: enerzijds het e-mailsysteem dat langs elektronische weg berichten transporteert en anderzijds de e-mailberichten zelf. Dat is in feite analoog aan de gewone post, waaronder beurtelings wordt verstaan de dienstverlening en de brieven zelf. Een e-mailsysteem bestaat uit programmatuur, transportmedium, (zoals netwerkvoorzieningen) en computers. Het stelt mensen in staat a-synchroon berichten uit te wisselen van en naar elektronische postbussen. Voor het duurzaam bewaren van e-mail gaat het in de eerste plaats om de inhoud, de e-mailberichten en hun context. Waar verder in deze uitgave wordt gesproken over e-mail wordt bedoeld op de inhoud, het e-mailbericht zelf.

E-mail bestaat dus uit berichten die door middel van een e-mail systeem zijn verzonden of ontvangen. Evenals bij papieren post kunnen e-mailberichten een veelheid van verschijningsvormen hebben. Een bericht kan als een eenvoudige mededeling zijn opgesteld in de editor van het e-mail programma, maar kan ook bestaan uit een complex digitaal document, met bewegende beelden of geluid gekoppeld aan het e-mailbericht. Een dergelijk attachment kan zijn gemaakt met een veelheid aan programma's, bijvoorbeeld een tekstverwerkingsprogramma, een rekenprogramma of een grafisch programma.⁹

Een e-mailbericht wordt door (eind)gebruikers meestal ervaren als één geheel, een elektronisch bericht waarmee informatie in digitale vorm wordt overgedragen tussen computers. Strikt genomen bestaat een e-mailbericht uit twee componenten: een kop (message header) en de inhoud van het bericht (message body). Een message header bevat informatie over het bericht. Afzender, geadresseerde, onderwerp, datum en vele andere gegevens kan men in de header aantreffen. Het bericht zelf, inclusief eventuele bijlagen kan data bevatten in iedere denkbare vorm. Dit kan eenvoudige ASCII tekst zijn, maar ook afbeeldingen, tekstverwerker bestanden, multimedia bestanden, uitvoerbare programmabestanden of HTML komen steeds meer voor.

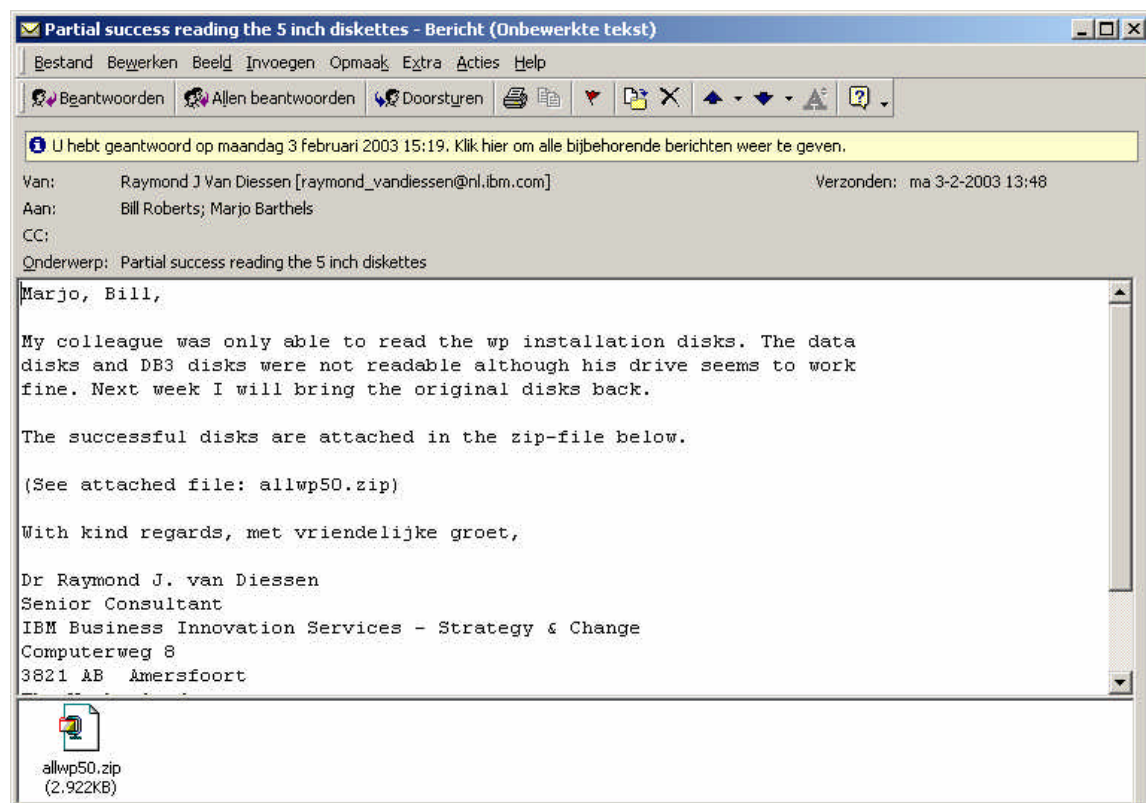
In de beleving van de gebruiker zou het e-mailbericht ook kunnen bestaan uit drie onderdelen. De header, de body en het attachment. Dit heeft alles te maken met de manier waarop e-mailberichten worden weergegeven in e-mail toepassingen, dus datgene wat je ziet op het beeldscherm. De representatie van een e-mailbericht en van datgene dat daadwerkelijk aan bits en bytes binnenkomt wil nog wel eens verschillen per e-mailprogramma. Het transmissiebestand (de bits en bytes die binnenkomen) is uitdrukkelijk ontwikkeld om interoperabiliteit te garanderen tussen de verschillende hard- en software platforms. Welk e-mailprogramma men ook gebruikt, het e-mailbericht zal altijd leesbaar zijn. Dit principe ligt ten grondslag aan het succes van e-mail als communicatiemedium, maar het betekent wel dat hetzelfde e-mailbericht er in Hotmail anders zal uitzien dan in bijvoorbeeld Outlook2000 van Microsoft.

3.4 E-mail en het transmissiebestand

Het e-mailbericht zoals zichtbaar op het scherm, is slechts één representatie van het transmissiebestand, het feitelijk verzonden bestand. De betreffende e-mailapplicatie leest dit transmissiebestand in en verwerkt deze tot hetgeen zichtbaar wordt op het scherm.

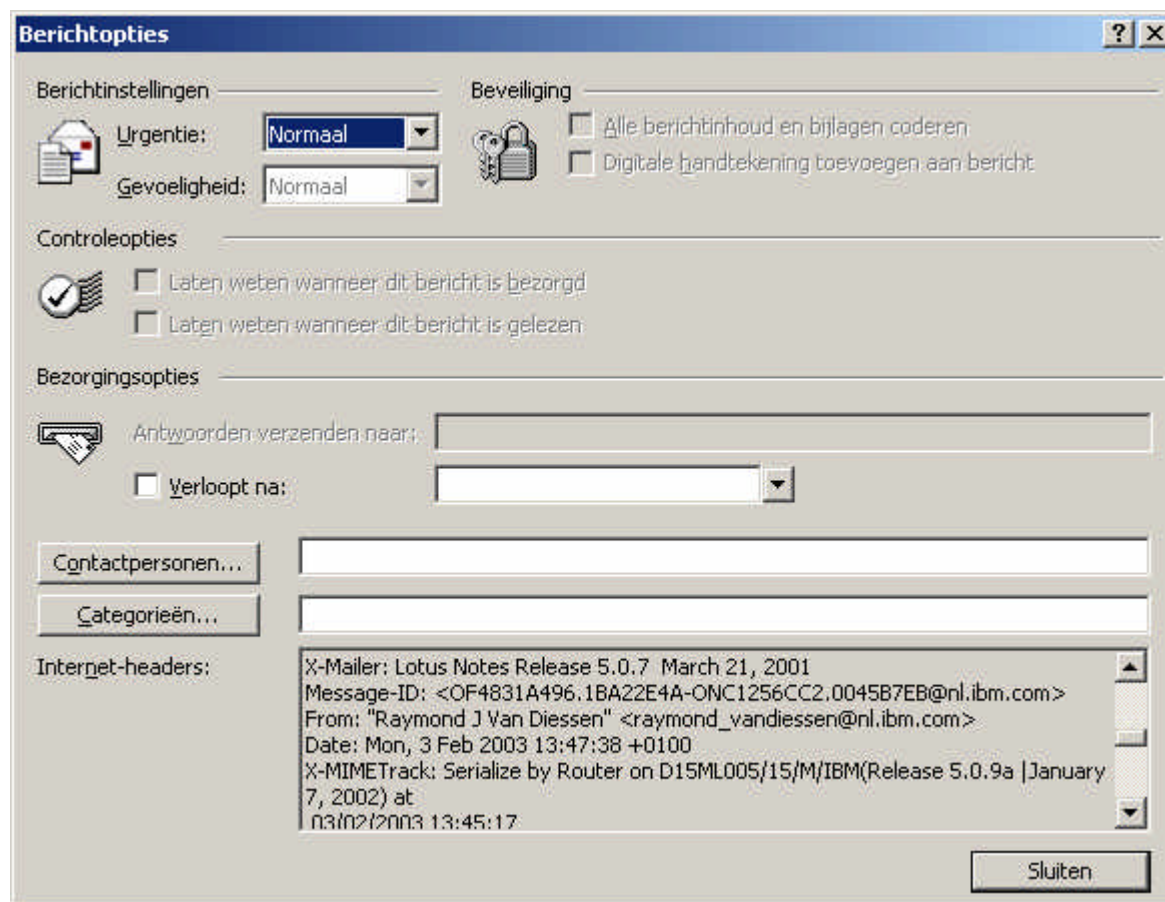
⁹

Ibid, p. 7.



Afbeelding 1. Voorbeeld van een e-mailbericht zoals de gemiddelde gebruiker deze op zijn scherm ziet.

Het e-mailbericht op het scherm is geenszins representatief voor de manier waarop het wordt verzonden. Het bestand dat verstuurd wordt, bestaat uit platte tekst (zie appendix D3 voor een voorbeeld). Het bestand wordt door de e-mailapplicatie achter de schermen gegenereerd als de gebruiker een e-mailbericht aanmaakt. Dit transmissiebestand is de primaire bron van inhoud en metadata voor alle e-mailberichten. In sommige e-mailapplicaties, bijvoorbeeld in Outlook, kan in 'Berichtopties' een beperkt gedeelte van dit transmissiebestand (deel van de header informatie) worden getoond.



Afbeelding 2. Voorbeeld van deel transmissiebestand (deel header informatie) zoals dat in Outlook wordt getoond.

De structuur van transmissiebestanden is in wezen bij zowel inkomende als uitgaande e-mailberichten hetzelfde, ongeacht hoe en met welke applicatie ze zijn gemaakt. Het enige verschil tussen inkomende en uitgaande transmissiebestanden is dat inkomende bestanden bij ontvangst extra headers bevatten. In deze headers staan de datum en het tijdstip waarop het bericht bij verzending door elke server is gegaan. Uitgaande e-mailberichten hebben deze headers niet, ofwel omdat ze nog niet zijn verzonden, ofwel omdat ze door de verzender zijn geselecteerd uit het venster 'Verzonden Items'. Hoe dan ook: uitgaande berichten zijn gezien vanuit perspectief van de verzender niet door servers zijn gegaan (natuurlijk wel voor de ontvangende partij). Voor een compleet overzicht van een Transmissie Bestand zie appendix D.

Evenals bij het e-mailbericht zelf, bestaat het transmissiebestand uit een header, body en bijlagen. Doordat deze onderdelen de totale opmaak van de e-mail bepalen, zal elk e-mailbericht er enigszins anders uitzien. De basiscomponenten van de e-mail zijn echter steeds hetzelfde. Hieronder worden de drie onderdelen van het transmissiebestand verder uitgediept.

De header

Een transmissiebestand begint met headerinformatie. Deze omvat veel meer dan de meeste e-mailapplicaties gewoonlijk weergeven. De basisinformatie in de header bestaat uit: datum en tijdstip waarop het bericht door de servers is ontvangen (alleen bij inkomende e-mail);

- o MIME-versie;
- o bericht-ID (uniek nummer);
- o discussieonderwerp;
- o datum discussie-index;
- o onderwerp;
- o van;
- o aan;
- o cc.

Aanvullende headerinformatie kan bestaan uit berichtmarkeringen, de naam van de applicatie waarmee het bericht is gemaakt, de prioriteit en de urgentie van de berichten, etc. De hoofdheader-sectie is een belangrijke bron van metadata over het bericht, de inhoud en het gebruik ervan (wie stuurde en/of ontving dit bericht en wanneer; wat was het onderwerp). De onderdelen van het bericht, zoals de berichttekst en eventuele bijlagen volgen de hoofdheader. Elk onderdeel heeft verder een eigen header. Deze header is belangrijk voor het beheer en bewaren, omdat hiermee de delen van het transmissiebestand nauwkeurig opnieuw zijn samen te stellen. De tweede, kleinere header geeft de volgende informatie:

- o content type (bijvoorbeeld text/plain; multipart/mixed);
- o tekenset (bijvoorbeeld ISO 8859-15);
- o Content-Transfer-Encoding (bijvoorbeeld Quoted Printable or Base64).

De body

De body is de tekst van de boodschap die de gebruiker in het berichtvenster intypt. Normaliter begint de tekst met een aanhef en eindigt hij met een slotformule. De berichttekst is doorgaans het eerste inhoudsdeel dat in het transmissiebestand wordt gedefinieerd. De header 'Content type' definieert eerst het type en geeft aan of het e-mailbericht als onbewerkte tekst is verzonden dan wel is opgemaakt.

Berichten in onbewerkte tekstindeling hebben één tekstdeel met als content type 'text/plain'. Berichten met een meer complexere opmaak kunnen uit meerdere tekstdelen bestaan. Zo wordt bij berichten die in HTML zijn opgesteld de berichttekst vaak in twee vormen weergegeven: de eerste vorm is de onbewerkte tekst, de tweede de HTML vorm. Hierdoor kunnen gebruikers met e-mailapplicaties die HTML niet ondersteunen toch de onbewerkte tekstinhoud zien, echter niet de totale opmaak. Berichten van dit type worden gedefinieerd als 'content type = multipart/alternative' en bevatten twee weergaven van de body.

Rich Text-berichten, zoals die in Outlook worden gemaakt, kunnen op soortgelijke wijze worden verzonden, maar de opmaak en eventuele bijlagen zijn vaak met elkaar verbonden in een uniek bestand met een *.dat- extensie. Deze gecodeerde versie van de berichttekst kan alleen in Outlook worden gedecodeerd. Dat kan problematisch zijn voor gebruikers die Outlook RTF-berichten ontvangen met een andere applicatie, zoals Eudora of zelfs Outlook Express.

Complexere, opgemaakte berichten kunnen uit meerdere tekstgedeelten of bijvoorbeeld ingevoegde foto's bestaan. De bijbehorende bestandsdelen moeten dan door de e-mailapplicatie opnieuw worden samengesteld voor een complete weergave van de body. Dit soort berichten worden gewoonlijk gedefinieerd als 'content type' = 'multipart/related'. Bij deze berichten identificeert de eerste header 'Content type' het totale verzendbestand als content type 'multipart'. Hiermee wordt aan de applicatie die het verzendbestand ontvangt, doorgegeven dat sommige onderdelen van het bericht moeten worden samengesteld voor een complete weergave van de body. Onder de eerste header 'Content type' staan de onderdelen, die elk een afzonderlijke header hebben. Die headers geven aan welke positie het onderdeel inneemt in het opnieuw samen te stellen digitale document, of het een ingevoegde afbeelding is, een ingevoegd bestand, een ingevoegd item, of een achtergrondafbeelding. Deze

informatie is heel erg belangrijk voor het goed kunnen reproduceren van het e-mailbericht. Als deze ontbreekt of verloren gaat, is de authenticiteit van het digitale document in gevaar.

Bijlagen

Het derde onderdeel van het e-mailbericht bevat bijlagen alsmede metadata over die bijlagen. De bijlagen kunnen van vrijwel elk bestandstype zijn, bijvoorbeeld een Word document, een PowerPoint presentatie, of een afbeelding in bitmap formaat. De enige beperking bij het meesturen van bijlagen is de grootte van de mailbox van de ontvanger. Te grote bijlagen worden dan geweigerd. De header 'content type' laat zien met welke applicatie het ingevoegde bestand is gemaakt en welke codering is gebruikt. Binaire bijlagen (zijnde geen kale [ASCII] tekstbestanden) worden vóór verzending gecodeerd hetzij in Quoted Printable hetzij in Base64, conform de MIME (Multipurpose Internet Mail Extensions) standaard. Dit is de standaard voor het verzenden van niet-tekstuele bestanden. De e-mailapplicatie moet dergelijke inkomende bijlagen dan ook eerst decoderen, voordat het digitale document volledig kan worden weergegeven.

3.5 Authenticiteitseisen voor e-mail

Zoals in hoofdstuk 2 al aan de orde kwam is het begrip authenticiteit van groot belang bij het bewaren van informatie, of dit nu op papier is, of digitaal. Voor elk type digitaal document, zoals e-mail, databases of tekstdocumenten kunnen echter andere authenticiteitseisen gelden. Deze eisen spelen een cruciale rol bij de keuze van een bewaarstrategie.

Testbed heeft uitgebreid geëxperimenteerd met e-mail en met methoden voor het waarborgen van de authenticiteit ervan. Op basis daarvan heeft Testbed een aantal authenticiteitseisen voor e-mail vastgesteld, waarin niet alleen de aspecten van een e-mailbericht worden gespecificeerd, maar ook bepaald wordt wat minimaal bewaard moet worden om ze goed te kunnen reproduceren. De hieronder geformuleerde eisen zijn afgestemd op de kenmerken of attributen van een digitaal document: context, inhoud, structuur, uiterlijk en gedrag. Daarnaast kunnen er aanvullende authenticiteitseisen vanuit het werkproces worden geformuleerd. Bijvoorbeeld het behoud van een bepaalde kleur op een digitale plattegrond, omdat de kleur een specifieke betekenis heeft die niet verloren mag gaan.

Context

Bij e-mailberichten staan vrijwel alle gegevens over de context in de berichtheaders. Bij sommige applicaties kunnen de headergegevens worden afgedrukt, bij andere niet. De headerinformatie bevat behalve metadata (verzender, ontvanger, onderwerp) ook technische aspecten van het bestand.

Testbed heeft de volgende minimale set authenticiteitseisen voor de context van een e-mail geïdentificeerd:

Essentiële headerelementen zijn:

- o het e-mailadres, de volledige naam en de organisatie van de afzender;
- o het e-mailadres, de volledige naam en de organisatie van alle geadresseerden;
- o in chronologische volgorde de datum en het tijdstip waarop het uitgaande bericht is verzonden;
- o in chronologische volgorde de datum en het tijdstip waarop het inkomende bericht is ontvangen, alsmede de datum en tijdstip waarop dit bericht door de afzender is verstuurd;
- o het onderwerp van het bericht;

- o de instellingen voor beveiliging en/of vertrouwelijkheid;
- o de bestandsnaam en het bestandsformaat van eventuele bijlagen (alle berichten).

Vrijwel alle bovenstaande gegevens bevinden zich al in het transmissiebestand, hoewel niet alle e-mailapplicaties deze informatie in zijn geheel laten zien.

Het bijhouden van een 'bewaar logboek', welke minimaal de volgende informatie bevat:

- o de gegevens over de oorspronkelijke en huidige bestandsformaten;
- o informatie nodig voor de interpretatie van het huidige bestandsformaat;
- o informatie over de strategie die is toegepast om het digitale document te wijzigen, zoals datum, tijdstip en de verantwoordelijke persoon of personen.

Het vaststellen van de organisatorische context, zoals:

- o het dossier dat bij het digitale document hoort (classificatie- of dossiercode);
- o de naam van de organisatie die het digitale document heeft gemaakt;
- o het bedrijfsproces waartoe het behoort.

Hierbij dient te worden opgemerkt dat de organisatorische context vooral wordt vastgelegd in een RMA of DMS. Dit valt overigens buiten de scope van het Testbed onderzoek.

Inhoud

De inhoud van een e-mail wordt vaak aangevuld met een bijlage van een willekeurig documenttype en is opgemaakt in een willekeurig bestandsformaat. De inhoud van een bijlage dient eveneens te worden bewaard.

Testbed heeft de volgende minimale set authenticiteitseisen voor de inhoud van een e-mailbericht geïdentificeerd:

De feitelijke inhoud moet altijd leesbaar zijn.

De inhoud van een gereproduceerde e-mail (dit is de e-mail na migratie of emulatie) dient even duidelijk en leesbaar te worden weergegeven als in het oorspronkelijke digitale document. Hierbij hoeft het uiterlijk van de inhoud niet identiek te zijn, maar de feitelijke inhoud moet natuurlijk wel hetzelfde blijven.

Ook aanvullingen op de inhoud moeten bewaard worden.

Eventuele aanvullende bestanden, zoals foto's of afbeeldingen, die in het oorspronkelijke bericht zijn ingevoegd, moeten eveneens worden bewaard, evenals de koppelingen en de hyperlinks.

Structuur

De structuur van een e-mailbericht betreft de indeling van een bericht, de interpretatie of de volgorde tussen de onderdelen van de e-mail. Elke willekeurige verandering van de structuur van de inhoud kan de interpretatie van het digitale document beïnvloeden. Bij e-mailberichten blijft de structuur gewoonlijk beperkt tot headerinformatie, inhoudsgegevens, informatie over de opmaak, bijlagen en ingevoegde items, bijvoorbeeld een digitaal visitekaartje.

Testbed heeft de volgende minimale set authenticiteitseisen voor de structuur van een e-mail geïdentificeerd:

De structuur van de inhoud van het oorspronkelijke bericht moet getrouw worden weergegeven

De onderdelen van het bericht moeten in de juiste logische volgorde opnieuw worden samengesteld. Bijlagen en ingevoegde items moeten duidelijk te onderscheiden zijn en deelheaders moeten naar het juiste deel verwijzen.

Uiterlijk

Het uiterlijk van een e-mailbericht is in de regel niet echt zwaarwegend als authenticiteitseis. Het uiterlijk van een bericht is immers afhankelijk van de e-mailapplicatie waarmee het wordt weergegeven. Verschillende applicaties geven berichten anders weer en ook de persoonlijke instellingen van de gebruiker kunnen het uiterlijk van de e-mail beïnvloeden. Bijgevolg is het uiterlijk van het bericht al van meet af aan veranderlijk.

Testbed heeft de volgende minimale set authenticiteitseisen voor het uiterlijk van een e-mail geïdentificeerd:

Het uiterlijk van de bewaarde e-mail mag afwijken van het origineel.

Het uiterlijk van het oorspronkelijke bericht en dat van de bewaarde versie hoeft niet identiek aan elkaar te zijn, maar het nieuwe uiterlijk mag de oorspronkelijke betekenis van het digitale document geenszins beïnvloeden.

De positie van een bijlage in een e-mail is niet van belang; de vermelding ervan wel.

Een aanduiding van een bijlage in de inhoud van het oorspronkelijke bericht moet hetzelfde worden weergegeven in de bewaarde versie. De positie van de bijlage is niet van belang en mag dus afwijken.

Gedrag

E-mailberichten hebben gewoonlijk geen complexe gedragskenmerken. Het gedrag dat vaak met e-mail wordt geassocieerd, is ten eerste de mogelijkheid tot het openen van bijlagen. Ten tweede is er de mogelijkheid tot het beantwoorden of doorsturen van het bericht. Dit tweede gedragskenmerk hoort echter niet bij het bericht zelf, maar bij de applicatie waarmee het bericht is gemaakt. Het gedrag dat gewoonlijk het meest voorkomt bij e-mailberichten is in de vorm van hyperlinks naar bijvoorbeeld websites of andere documenten.

Testbed heeft de volgende minimale set authenticiteitseisen voor het gedrag van een e-mail geïdentificeerd:

De mogelijkheid tot het openen van ingevoegde bijlagen in een geschikte applicatie moet bewaard blijven

De mogelijkheid tot het openen van ingevoegde bijlagen in een geschikte applicatie moet bewaard blijven maar niet de mogelijkheid tot het verzenden, doorsturen of wijzigen van het bericht.

Hyperlinks naar websites dienen te worden bewaard.

Hyperlinks naar websites dienen te worden bewaard, inclusief de URL. Voor de authenticiteit van de e-mail is het in het algemeen niet noodzakelijk om de inhoud van een website waar de e-mail naar verwijst, te bewaren. Het is afhankelijk van het belang van de gekoppelde website voor de organisatie of deze bewaard moet worden.

Verwijzingen naar andere documenten

Ook verwijzingen naar andere documenten dienen bewaard te worden. De titel van het gekoppelde document moet worden vastgelegd in de metadata van het bericht.

Evenals bij gekoppelde websites is het bewaren van de inhoud van het gekoppelde document afhankelijk van het belang dat de organisatie eraan hecht.

3.6 De digitale handtekening

Elektronische communicatie binnen de overheid, maar ook tussen overheid enerzijds en burger en bedrijfsleven anderzijds zal steeds vaker van digitale handtekeningen worden voorzien.

Deze ontwikkeling zal worden versterkt door acties op wetgevingsgebied (Wet elektronische handtekeningen, Wet elektronisch bestuurlijk verkeer) en ook door de ontwikkeling van de benodigde infrastructurele stelsels waarvan de PKI-overheid binnen de overheid een belangrijke representant is.

Naarmate het gebruik van de digitale handtekening toeneemt, komt ook de vraag omtrent de bewaring van die handtekening prominenter naar voren. Wat is het beleid inzake de bewaring van de digitale handtekening?

Testbed Digitale Bewaring heeft een eerste verkenning gedaan van het onderwerp. Om de analyse te vervolmaken en beleidskeuzen vast te stellen is vervolgonderzoek nodig.

Hieronder is een aantal bevindingen van deze verkenning weergegeven.

Sommige gegevens die aan de basis liggen van digitale handtekeningen en die in hoge mate bepalend zijn voor het vertrouwen dat in een digitale handtekening kan worden gesteld, berusten bij de certificatedienstverlener (welke in de Wet elektronische handtekeningen wordt geïntroduceerd), oftewel een vertrouwde derde partij (Trusted Third Party). Het betreft hierbij vooral de gegevens waarmee de certificatie kan worden bewezen (gegevens over geraadpleegde identiteitsdocumenten, aanvraagformulieren en getekende voorwaarden) en historische gegevens over ingetrokken certificaten. Deze gegevens kunnen ook van groot belang zijn in het geval van een dispuut over de authenticiteit en geldigheid van een digitale handtekening.

Voor deze belangrijke gegevens die bij de certificatedienstverlener berusten, geldt een minimum bewaartermijn van zeven jaar na het verstrijken van de geldigheid van een certificaat (Wet Elektronische Handtekeningen). De wetgever staat bij het bepalen van deze minimum bewaartermijn het gebruik voor ogen in privaatrechtelijke verhoudingen (e-business), terwijl het partijen (de gebruiker van de digitale handtekening en de certificatedienstverlener) vrij staat om een langere bewaartermijn overeen te komen waar de toepassing dit wenselijk of noodzakelijk maakt.

Testbed Digitale Bewaring adviseert, totdat nader onderzoek heeft plaatsgevonden, om alle gegevens over de digitale handtekening en de identiteit van de ondertekenaar, gegevens omtrent de verificatie van die handtekening en het bijbehorende certificaat integraal te bewaren in/bij de metadata van het digitale document op het moment van verificatie in het werkproces.

3.7 Samenvatting

Voor het vaststellen van de authenticiteit van digitale documenten zijn de integriteit en de verificatie cruciaal:

Ten eerste moeten die kenmerken van het digitale document bewaard blijven die in de authenticiteitseisen zijn vermeld, zodat de integriteit van het digitale document is

gewaarborgd. Dit wordt grotendeels bewerkstelligd door een strategie te ontwikkelen waarbij de belangrijke aspecten van de inhoud, de structuur, het uiterlijk en het gedrag van het digitale document kunnen worden bewaard. Deze eisen hebben betrekking op de kenmerken van een digitaal document als e-mail.

Op de tweede plaats draait het om verificatie. In de metadata wordt de context vastgelegd waarin het digitale document is gemaakt en gebruikt en worden bovendien eventuele wijzigingen opgenomen die ten gevolge van beheer- en bewaaractiviteiten in het digitale document zijn aangebracht. Hierdoor kan aangetoond of geverifieerd worden in hoeverre het digitale document nog authentiek is, ook al is het e-mailbericht niet meer exact hetzelfde als het oorspronkelijke.

4. Drie bewaarstrategieën getoetst

De meest voor de hand liggende strategieën om digitale informatie langdurig te kunnen bewaren zijn migratie, XML en emulatie. Deze wereldwijd veel bestudeerde methoden worden hier kort besproken en beoordeeld op bruikbaarheid voor het bewaren van e-mail.

4.1 Inleiding

Migratie, XML en emulatie zijn de bewaarstrategieën die het meest in aanmerking komen om digitale documenten goed te kunnen bewaren. Elke bewaarstrategie bestaat uit verschillende subcategorieën, die in dit hoofdstuk eveneens aan de orde zullen komen. Tevens wordt, waar mogelijk, beschreven hoe elke strategie kan worden geïmplementeerd. Daarnaast zullen de voor- en nadelen van elke strategie worden beoordeeld in het licht van de specifieke eisen die gesteld worden aan het duurzaam bewaren van e-mail, zoals eerder beschreven in hoofdstuk 3. Op basis van deze afwegingen zal worden bepaald wat de meest geschikte strategie is voor het duurzaam bewaren van e-mail.

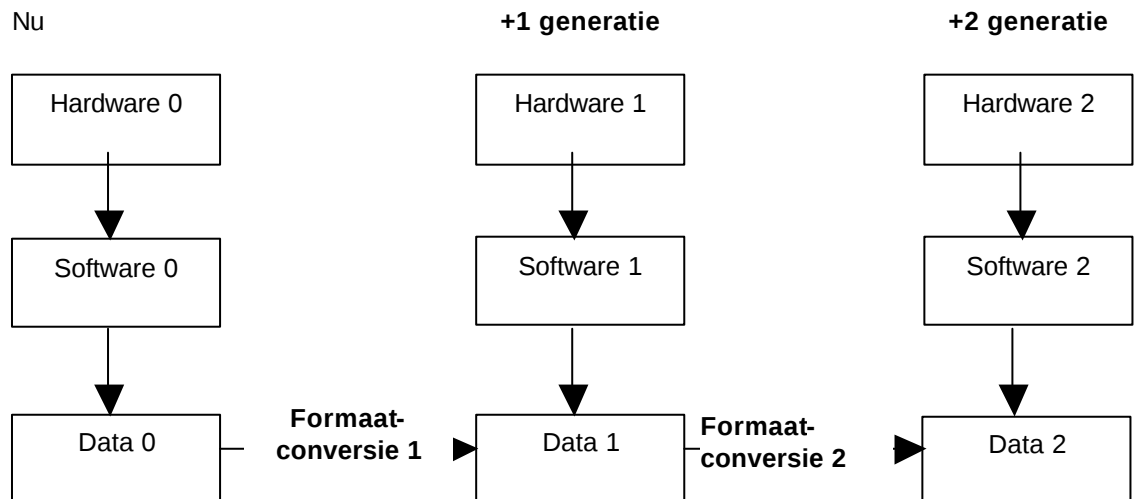
4.2 Migratie als bewaarstrategie

Testbed Digitale Bewaring hanteert voor migratie de volgende definitie: "Het overzetten van bestanden van de ene hardware-/softwareomgeving naar een andere".

Migratie is een gebruikelijke manier om digitale veroudering aan te pakken. Bestanden die in een oud bestandsformaat zijn gemaakt, worden overgezet naar een nieuw formaat dat wel draait op de moderne computerplatforms. Zo kan een tekstdocument van Word 95 naar Word 2002 of van WordPerfect 5.1 naar Adobe's PDF 1.4 (Portable Document Format) worden overgezet.

Aan elke migratie dient wel het nodige onderzoek vooraf te gaan. Het formaat van het doelbestand moet immers compatibel zijn met het bronformaat, zodat alle belangrijke eigenschappen van het digitale document ook in de omgezette vorm worden weergegeven en de authenticiteit en integriteit van het digitale document is gewaarborgd.

Het volgende schema toont de relaties tussen de hardware, software en data bij het toepassen van migratie:



Afbeelding 3. Basisschema migratie

Testbed heeft onderzoek gedaan naar en/of geëxperimenteerd met de volgende vormen van migratie:

- o Achterwaartse compatibiliteit;
- o Interoperabiliteit;
- o Conversie naar standaard formaten.

Bij de keuze van de gewenste vorm moet een organisatie rekening houden met ten eerste de eisen ten aanzien van de authenticiteit van de digitale documenten waarmee wordt gewerkt. Ook de periode gedurende welke het digitale document moet worden bewaard, is bepalend: gaat het om twee, tien of twintig jaar?

4.2.1 Achterwaartse compatibiliteit

Achterwaartse compatibiliteit maakt het mogelijk om een bestand dat in een oudere versie van een applicatie is gemaakt, in een nieuwe versie van die applicatie te interpreteren en correct weer te geven. Softwareleveranciers garanderen dat nieuwe versies van hun software compatibel zijn met de vorige. Een voorbeeld is het gebruik van Word 2002 voor het lezen van bestanden die met Word 95 zijn gemaakt en in het bestandsformaat van Word 95 zijn opgeslagen.

Het heeft de voorkeur om de oude bestanden na een achterwaartse compatibiliteitsmigratie op te slaan in het nieuwe bestandsformaat, aangezien software in de regel slechts een beperkt aantal generaties oudere bestandsformaten ondersteunt. Deze strategie kan worden gebruikt voor niet al te oude en in gebruik zijnde bestanden, maar is minder geschikt voor langdurige bewaring, aangezien na enkele generaties problemen zullen optreden. Want hoewel leveranciers er belang bij hebben om te zorgen voor achterwaartse compatibiliteit - aangezien de gebruiker dan makkelijk kan upgraden naar een nieuwe versie - kan het vanwege de eigenschappen van de nieuwe versie technisch moeilijk worden oude, deels in onbruik geraakte eigenschappen van de oude versie volledig te blijven ondersteunen. Een nadeel is dus dat slechts een beperkt aantal generaties compatibel zijn met

elkaar en zelfs dan al kunnen sommige eigenschappen van het digitale document door de nieuwe softwareversie anders worden geïnterpreteerd. Dit kan leiden tot aantasting van de authenticiteit en integriteit van het digitale document. Geen enkel programma heeft het eeuwige leven.

Een ander nadeel van achterwaartse compatibiliteit als bewaarstrategie is dat het digitale document blijft opgeslagen in het leverancierseigen bestandsformaat (bijvoorbeeld *.msg voor e-mailberichten in Outlook), waarmee de ongewenste afhankelijkheid van in dit geval de software in stand wordt gehouden.

Een laatste nadeel is dat een migratie naar een hogere versie iedere paar jaar moet worden herhaald. Experimenten bij het Testbed hebben uitgewezen dat bij iedere migratieslag het risico aanwezig is dat bestanden een wijziging, hoe klein ook, ondergaan, waardoor de authenticiteit en integriteit van het digitale document kunnen worden aangetast.

Is achterwaartse compatibiliteit geschikt voor het bewaren van e-mail?

Gezien de nadelen van achterwaartse compatibiliteit als bewaarstrategie (opslag in het leverancierseigen formaat, moet iedere paar jaar worden herhaald, risico van aantasting authenticiteit en integriteit van het digitale document) is de conclusie dat achterwaartse compatibiliteit als bewaarstrategie een benadering is die functioneel kan zijn voor de korte- of middellange termijn (zolang het bestand nog wordt gebruikt), maar geen reële optie is voor het aanpakken van digitale veroudering van e-mailberichten op de lange termijn.

4.2.2 Interoperabiliteit

Bij interoperabiliteit in technische zin wordt het probleem van digitale veroudering aangepakt door er voor te zorgen dat bestanden en digitale documenten niet langer of in mindere mate afhankelijk zijn van een bepaalde combinatie van hard- en software. Interoperabiliteit houdt in dat een bestand van het ene platform naar het andere kan worden overgezet en dan nog steeds op dezelfde of een soortgelijke manier kan worden weergegeven. Bij de eenvoudigste implementatie zijn er twee vormen mogelijk:

- o bestand afhankelijk van de applicatie, maar niet van het besturingssysteem;
- o bestand onafhankelijk van zowel het besturingssysteem als de applicatie.

Bij de eerste optie is onder andere een scenario mogelijk waarbij applicaties onder verschillende besturingssystemen kunnen draaien. Softwareleveranciers beseffen dat niet iedereen hetzelfde computerplatform heeft en geven daarom verschillende applicatieversies uit voor verschillende besturingssystemen, bijvoorbeeld versies die onder Windows, Linux of Solaris draaien. Vaak moeten fabrikanten echter voor elk besturingssysteem verschillende applicatieversies uitgeven, waardoor de interoperabiliteit beperkt kan zijn. Er zijn nog steeds maatregelen nodig om verlies van informatie, authenticiteit en functionaliteit te voorkomen. Ook hier is nog steeds een nadeel dat bestanden in het leverancierseigen formaat zijn opgeslagen en dat het overzetten naar een volgende versie van de applicatie noodzakelijk is.

Bij de tweede optie is een scenario mogelijk waarbij bestanden onafhankelijk zijn van zowel de applicatie als het besturingssysteem. Een voorbeeld hiervan is de situatie waarin het gekozen bestandsformaat een breed geaccepteerde standaard is, bijvoorbeeld XML. Er bestaan ook andere manieren voor uitgebreide interoperabiliteit, maar door hun complexiteit is de duurzaamheid

vaak veel moeilijker te garanderen. Bovendien richten we ons meer op het duurzaamheidsaspect van het proces dan op de interoperabiliteit. Desondanks verdient een strategie waarbij zowel interoperabiliteit als duurzaamheid mogelijk zijn de voorkeur boven een strategie op basis van interoperabiliteit alleen. Een strategie die interoperabiliteit en duurzaamheid combineert (toepassing van XML) wordt uitvoerig in paragraaf 4.3 besproken. Dit type interoperabiliteit is hetzelfde als de methode van conversie naar standaarden, die hieronder wordt besproken.

Een complexere vorm van interoperabiliteit vereist het gebruik van een interim conversieprogramma. Daarbij worden bestanden vanuit het leverancierseigen formaat, bijvoorbeeld MS Word, omgezet naar een uitwisselingsformaat, zoals ASCII (American Standard Code for Information Interchange) of RTF (Rich Text Format), dat weer kan worden gelezen door een ander programma, zoals WordPerfect. Bij een dergelijke aanpak is het risico groot dat essentiële kenmerken van het digitale document verloren gaan, met name wanneer een digitaal document een complexe opmaak of een multimediale inhoud heeft.

Is interoperabiliteit geschikt voor het bewaren van e-mail?

Totale, probleemloze interoperabiliteit van complexe digitale documenten tussen en binnen platforms zonder het gebruik van een conversieprogramma is zeldzaam. Leverancierseigen applicaties produceren bestanden waarvoor de desbetreffende versie van de applicatie nodig is. Pas dan kan het digitale document goed worden weergegeven. Het is onwaarschijnlijk dat bestanden die met een applicatie onder een bepaald besturingssysteem zijn gemaakt accuraat en op authentieke wijze kunnen worden gelezen door een versie van die applicatie die voor een ander besturingssysteem is ontworpen.

Het is belangrijk te benadrukken dat hoewel het e-mail transmissiebestand uitdrukkelijk voor interoperabiliteit is ontwikkeld, het e-mailbericht na verwerking door een e-mailapplicatie gewoonlijk wordt opgeslagen in een leverancierseigen formaat zoals *.msg voor Outlook berichten of *.vew voor Novell Groupwise berichten. Als het bericht wordt opgeslagen als *.txt of *.html bestand worden niet alle essentiële transmissiegegevens opgeslagen die noodzakelijk zijn voor langdurige bewaring.

Zoals eerder besproken kan bij interoperabiliteit het bestand worden geconverteerd naar een uitwisselingsformaat dat door andere applicaties kan worden gelezen, vaak ook op andere platforms. Het grootste probleem bij deze optie is dat de bestanden eerst moeten worden omgezet naar een uitwisselingsformaat zoals ASCII. Hierdoor kunnen niet alle aspecten van het digitale document worden opgenomen. Als de bestanden vervolgens door een ander programma worden gelezen, zullen bepaalde eigenschappen en aspecten van de digitale documenten zodanig worden geïnterpreteerd dat de authenticiteit en integriteit worden aangetast. Bovendien worden bij deze strategie de problemen van leverancierseigen bestandsformaten niet noodzakelijkerwijs aangepakt.

Gezien de bovengenoemde nadelen (verlies van essentiële eigenschappen van het e-mailbericht; opslag in het leverancierseigen formaat) is interoperabiliteit niet geschikt voor het duurzaam bewaren van e-mail. Desondanks kan de strategie wel worden gebruikt in combinatie met de methode van conversie naar standaarden, waarbij interoperabiliteit vaak impliciet het gevolg is van de conversie. Deze optie wordt hieronder besproken.

4.2.3 Conversie naar standaarden

Conversie naar standaarden betreft migratie vanuit een vaak gesloten leverancierseigen bestandsformaat naar een formaat met een brede ondersteuningsbasis die ruime toepassing vindt bij computerplatforms. Het voordeel is dat digitale documenten niet langer afhankelijk zijn van de oorspronkelijke hard- en software waarmee zij zijn gemaakt en die een bedreiging vormt voor hun duurzaamheid. Hierbij zijn óf 'de jure'- óf 'de facto'-standaarden mogelijk. 'De jure'-standaarden komen tot stand via een formeel proces door een formeel geaccrediteerd standaardisatieorgaan (ISO, enz). Een de jure standaard komt ook altijd tot stand in een open proces, aangezien consensus en beschikbaarheid de belangrijkste motieven zijn voor formele standaardisatieorganisaties. 'De facto'-standaarden zijn breed geïmplementeerd; er is een kritische massa die van de specificatie gebruik maakt. De facto standaarden kunnen door consortia worden ontwikkeld d.m.v. open processen, maar ze kunnen ook d.m.v. gesloten processen tot stand komen (leverancierseigen standaarden).¹⁰

Voor de duurzame bewaring van digitale documenten zijn twee vormen van conversie naar standaarden bruikbaar:

- o conversie naar de gepubliceerde, zij het leverancierseigen standaard, zoals PDF (een de facto standaard);
- o conversie naar de open, door de gemeenschap ontwikkelde standaarden, zoals XML.

Conversie naar een leverancierseigen standaard heeft niet de voorkeur, want hoewel de specificatie gratis verkrijgbaar is en er tools mee kunnen worden ontworpen, moeten de basissoftware en licentiekosten nog steeds worden betaald en beheert de leverancier de laatste uitgave van elke nieuwe versie. De fabrikant kan de ontwikkeling van de software te eniger tijd beëindigen en is niet verplicht om ervoor te zorgen dat een derde die ontwikkeling voortzet. Dit betekent dat als de fabrikant failliet gaat, de standaard mogelijk verdwijnt. Om die reden heeft conversie naar open standaarden de voorkeur boven een standaard die eigendom is van één bedrijf.

Ofschoon ook bij open standaarden nog altijd software voor verwerking en conversie moet worden ontwikkeld en/of worden aangeschaft, zijn licentiekosten en dergelijke niet nodig. Bij de ontwikkeling van open standaarden zijn zeer veel personen betrokken en de specificaties zijn gratis verkrijgbaar.

Is conversie naar standaarden geschikt voor het bewaren van e-mail?

De meeste e-mailapplicaties bieden (nog) geen mogelijkheid voor het rechtstreeks off line opslaan van e-mailberichten in een niet leverancierseigen formaat, waarbij het complete digitale document toegankelijk blijft. Om die reden moet worden uitgeweken naar andere formaten. Conversie naar standaarden zal bij een niet-leverancierseigen bestandsformaat, zoals XML, resulteren in zowel achterwaartse compatibiliteit als interoperabiliteit. In dat geval zijn achterwaartse compatibiliteit en interoperabiliteit de voordelen van deze strategie en niet zelf de strategie.

In de eerder genoemde ministeriële 'Regeling geordende en toegankelijke staat archiefbescheiden' worden meerdere standaarden vermeld voor het duurzaam bewaren van digitale documenten¹¹. Voor tekstdocumenten en afbeeldingen zijn dat onder andere PDF, XML, SGML en TIFF. Er wordt echter

¹⁰ Zie: XML: de mogelijkheden en valkuilen voor de overheid; W. Thomas, 19 september 2002.

¹¹ Regeling geordende en toegankelijke staat archiefbescheiden, februari 2002.

geen specifiek formaat vermeld voor het bewaren van e-mailberichten. In het kader van deze standaarden kunnen e-mailberichten worden beschouwd als een type tekstdocument dat alfanumerieke gegevens en mogelijk afbeeldingen bevat. Uiteraard moet de standaard zorgvuldig worden gekozen en geschikt zijn voor het weergeven van het betreffende digitale document. Testbed heeft in dit kader PDF, RTF en XML onderzocht.

De omzetting van e-mailberichten naar PDF formaat is niet zinvol. Het merendeel van de e-mail headerinformatie (waaronder context informatie) gaat verloren, met als gevolg dat de authenticiteit van het e-mailbericht niet langer kan worden gewaarborgd. Alleen de gegevens die tijdens de weergave door de e-mailapplicatie worden getoond, kunnen worden afgedrukt naar PDF. Bijlagen worden niet geopend en de bijbehorende pictogrammen zoals die in het oorspronkelijke bericht voorkwamen, verschijnen in het PDF bestand gewoon in de vorm van een afbeelding. PDF is een formaat dat primair bedoeld is voor het publiceren van gefixeerde documenten, waarbij de identieke vorm van het digitale document moet worden bewaard. Dit is voor e-mailberichten geen vereiste aangezien e-mailberichten geen vast uiterlijk hebben. Die is immers afhankelijk van de gebruikte e-mailapplicatie.

Behalve PDF heeft Testbed ook RTF (Rich Text Format) onder de loep genomen. RTF is evenmin geschikt om e-mail duurzaam te bewaren. Ook dan gaat essentiële headerinformatie verloren. Daar komt bij dat zowel PDF als RTF geen open standaard zijn, waarmee de afhankelijkheid van hard- en of softwareleveranciers in stand wordt gehouden, hetgeen niet wenselijk is vanuit de optie van digitale duurzaamheid.

Tenslotte heeft Testbed geëxperimenteerd met de toepassing van XML. De resultaten zijn positief: In tegenstelling tot PDF en RTF is XML wel in staat alle headerinformatie van het e-mailbericht vast te houden. Voorts is XML een voorbeeld van een platform- en programmaonafhankelijke, open standaard. Meer specifieke voor- en nadelen van het gebruik van XML als bewaarstrategie voor digitale documenten worden besproken in de volgende paragraaf.

4.3 XML als bewaarstrategie

Testbed Digitale Bewaring heeft ook XML onderzocht als strategie voor het duurzaam bewaren van digitale documenten. XML staat voor Extensible Markup Language. Het is een opmaaktaal voor het verrijken van data met informatie over structuur en betekenis die tevens kan worden toegepast als bestandsformaat. Het is een open standaard die is gedefinieerd door het World Wide Web Consortium, een non-profitorganisatie die interoperabele technologieën ontwikkelt, zoals specificaties, richtlijnen, software en tools, zodat het internet optimaal kan worden gebruikt¹². XML is platformonafhankelijk en kan direct door de mens worden gelezen. Om deze redenen kan XML worden ingezet voor digitale bewaring en interoperabiliteit.

Afhankelijk van de implementatiewijze is bij de XML-strategie overlapping mogelijk met de andere, hierboven beschreven strategieën. Zo kan de conversie van bestanden naar XML-formaat worden beschouwd als een aparte migratietechniek (zie Conversie naar standaarden, hierboven). XML kan ook worden gebruikt in combinatie met de emulatiestrategie en de UVC-variant, of worden toegepast als bestandsformaat en koppelingstool als onderdeel van een objectgeoriënteerde

¹²

Zie <http://www.w3.org>.

oplossing. In de volgende paragrafen wordt eerst beschreven hoe met XML het probleem van technologische veroudering bij digitale documenten kan worden aangepakt. Daarna wordt ingegaan op alternatieve en aanvullende toepassingen van XML voor digitale bewaring.

4.3.1 Wrapper en kapstok

Wrapper en kapstok zijn termen die worden gebruikt voor het beschrijven van één basismanier waarop XML kan worden geïmplementeerd voor digitale bewaring. Bij deze strategie blijft het bestand (of een gedeelte daarvan) gewoonlijk in het oorspronkelijke formaat en wordt een wrapper, die de inhoud beschrijft, over de buitenkant geplaatst. De inhoud van de wrapper, de structuur ervan en de manier waarop de strategie wordt geïmplementeerd, zijn afhankelijk van de eisen van de betrokken instelling.

Zo kan in één wrapper een groot aantal verwante objecten worden bewaard of ingekapseld. In zo'n wrapper kunnen dan worden opgeslagen: een exemplaar van het bestand in het oorspronkelijke formaat (het 'masterbestand'), een geconverteerde versie van het bestand in een gewenst bestandsformaat, zoals een PDF bestand (het toegankelijke digitale document) en metadata voor het beschrijven van deze objecten en de geschiedenis ervan. Alle objecten kunnen samen met een XML wrapper worden opgeslagen die niet alleen de inhoud beschrijft, maar ook de toegankelijkheids- en bewaareisen. Hierdoor beschikken instellingen op één veilige locatie over:

- o het masterbestand, het authentieke digitale origineel voor toekomstige bewaaractiviteiten;
- o een geconverteerde versie waartoe onderzoekers toegang hebben met behulp van moderne software en
- o metadata die het digitale document beschrijven. Metadata met betrekking tot de toegankelijkheid en het bewaren kunnen worden geüpdatet door de wrapper te voorzien van aanvullende gegevens. Als dat nodig is, kunnen ook extra digitale objecten worden toegevoegd.

De wrapperstrategie is op een aantal manieren getest en geïmplementeerd. Het VERS project in Australië¹³ kapselt de documenten, context en authenticiteitinformatie in één object. XML wordt gebruikt om de ingekapselde digitale documenten zodanig te coderen dat de geschiedenis van de digitale document eenvoudig kan worden bijgewerkt door een nieuwe laag metadata over de vorige te leggen (het zogeheten ui-model).

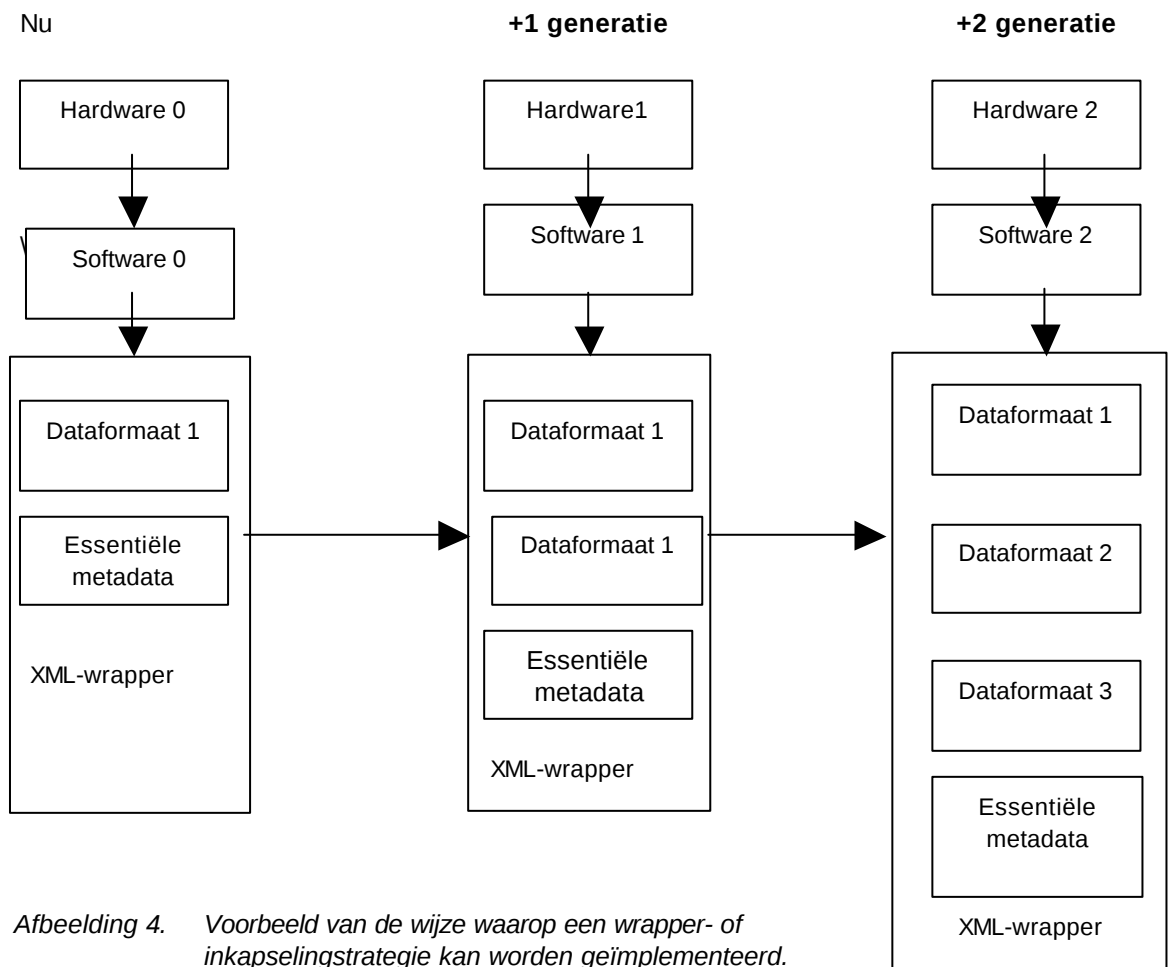
Het e-Archiving project¹⁴ gaat op vergelijkbare wijze te werk bij items die wetenschappelijke data bevatten, maar gebruikt de term containers, waarbij de keuze tussen het gebruik van conversie en emulatie afhangt van kostenoverwegingen en verouderingsfactoren.

¹³

<http://www.prov.vic.gov.au/vers/welcome.htm>

¹⁴

<http://www.library.tudelft.nl/e-archive/>



Afbeelding 4. Voorbeeld van de wijze waarop een wrapper- of inkapselingstrategie kan worden geïmplementeerd.

Opmerking bij deze afbeelding

Veroudering van hardware/software wordt voorkomen doordat XML platformonafhankelijk is. Als de bestanden in de wrapper dreigen te verouderen, wordt dat aangegeven via de XML wrapper en kan door instellingen met behulp van moderne applicaties een nieuwe versie van het digitale document worden gemaakt. Dit is een voorbeeld. Een dergelijke strategie kan ook op andere manieren worden geïmplementeerd.

Organisaties kunnen er ook de voorkeur aan geven slechts één bestand in XML te verpakken en aanvullende bewaarmetadaten aan de wrapper toe te voegen als deze wordt gemaakt. Dit is mogelijk als het bronbestand relatief eenvoudig is en makkelijk kan worden 'getagged' (verrijkt). Bestanden in complexe of leverancierseigen formaten zijn wellicht meer geschikt voor de uitgebreide inkapselingstrategie zoals die hierboven is afgebeeld.

Bij de zogenoemde kapstokstrategie wordt XML op een ietwat andere manier gebruikt, namelijk als een boomstructuur waarin het digitale document en de bijbehorende componenten kunnen worden ondergebracht. De complexiteit van de digitale documenten en de eisen of voorkeuren van de instelling zijn bepalend voor de wijze waarop de strategie wordt geïmplementeerd. Uiteraard is er nog steeds geschikte software nodig om de bestanden te lezen. Hoewel XML niet afhankelijk is van een bepaald platform, moet de taal nog steeds door software worden verwerkt. Deze software kan al dan niet platformonafhankelijk zijn en moet daarom ook worden gecontroleerd op tekenen van naderende veroudering.

XML is geschikt voor een inkapseling- of 'wrapper'- strategie omdat XML een beschrijvende, hard- en software onafhankelijke, voor mens en machine leesbare meta-taal is, die het mogelijk maakt meerdere bestanden bij elkaar te houden en te voorzien van noodzakelijke metadata, waarmee digitale veroudering kan worden tegengegaan. Het oorspronkelijke bestand, of althans het grootste gedeelte ervan, blijft intact en kan samen met andere representaties van het digitale document worden opgeslagen. Deze beschrijving van wrapper en kapstok is geen uitputtende en definitieve opsomming van de manieren waarop XML kan worden geïmplementeerd.

Zijn wrapper en kapstok geschikt voor het bewaren van e-mail?

Door wrappers of kapstok bij een XML-strategie te gebruiken kunnen verwante bestanden probleemloos aan elkaar worden gekoppeld via één digitaal documentobject. Ook kan met dit model bij volgende bewaaractiviteiten extra bestanden aan het digitale documentobject worden toegevoegd (zie afbeelding 11 in het volgende hoofdstuk). De verwante bestanden worden gekoppeld met XML mark-up en bij het opslaan zijn deze relaties expliciet aangegeven.

E-mailberichten bestaan in de regel uit meerdere bestanden of componenten, zoals headerinformatie (de adressering), de body van het bericht en eventuele attachments (bijlagen). Deze opbouw van een e-mail maakt de toepassing van de wrapper of kapstok benadering tot een bijzonder geschikte strategie, waar Testbed goede resultaten mee heeft geboekt:

Wrapper benadering

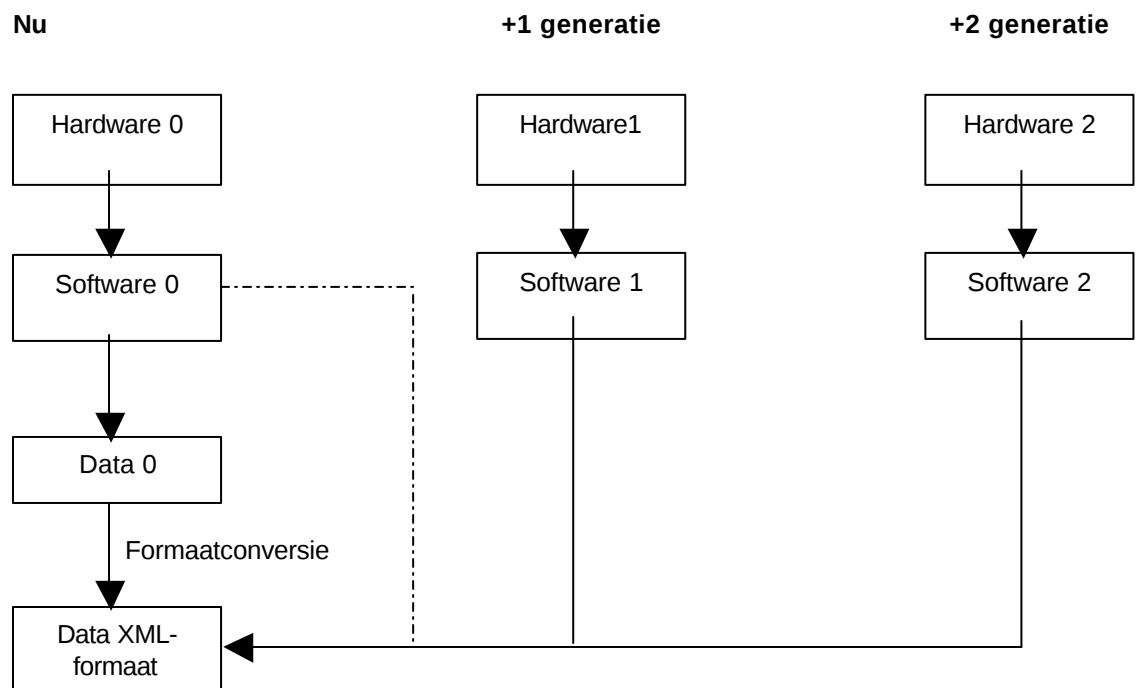
Een groot voordeel van de wrapper benadering is het beheer van het XML bestand. Er is immers slechts één bestand dat alle overige componenten omsluit. Het nadeel van deze benadering is dat eventuele attachments in gecodeerde vorm (middels base64) in het XML bestand zijn opgeslagen. Immers evenals het e-mail transmissiebestand staan XML bestanden alleen tekst toe en geen binaire bestanden. Een binair bestand bestaat uit een verzameling bits die meer dan alleen platte tekst bevatten. Er staan extra codes in die alleen door een specifiek systeem of programma zijn te gebruiken, zoals Word, Excel of PowerPoint. Consequentie is dat de decodeeralgoritmes eveneens moeten worden bewaard, om in een later stadium de attachments te kunnen decoderen (om te zetten van hun gecodeerde vorm naar hun oorspronkelijke binaire formaat)

Kapstok benadering

De kapstok benadering kent dit nadeel niet: attachments worden in hun oorspronkelijke bestandsformaat opgeslagen. Dit biedt de mogelijkheid om per (groep van) attachment(s) de meest geschikte bewaarstrategie toe te passen. Deze benadering heeft de voorkeur van Testbed. Een gedetailleerde beschrijving van de implementatie van deze benadering is te vinden in hoofdstuk 5 'De aanbevolen e-mail benadering'.

4.3.2 XML als bestandsformaat

Een alternatief voor wrapping is het rechtstreeks converteren van de bestanden naar XML of ze rechtstreeks in XML te genereren, waarbij XML als bestandsformaat wordt gebruikt. Deze techniek is verwant aan de methode van conversie naar standaarden, die onder Migratie is beschreven, maar in dit geval wordt naar één specifieke standaard geconverteerd, namelijk XML. Aangezien XML onafhankelijk is van hard- en software is het duurzamer dan veel van de huidige leverancierseigen bestandsformaten. Het aantal benodigde conversies zal aanzienlijk afnemen en daarmee de kans op aantasting van de authenticiteit van het digitale document.



Afbeelding 5. Conversie naar XML vereist minder omzettingen dan migratie

XML is een geschikt bestandsformaat voor het registreren van metadata en het weergeven van de vijf eerder beschreven kenmerken van een digitaal document, te weten 'inhoud', 'context', 'structuur', 'uiterlijk' en 'gedrag': *Inhoud en context* kunnen worden weergegeven in XML, zodat de inhoud op een duidelijke en voor de mens leesbare manier kan worden beschreven.

XML is een goed gestructureerde verrijkingstaal waarmee de *structuur* van het digitale document eenvoudig kan worden weergegeven. Bovendien kan via een XML-schema of DTD expliciet de structuur van het digitale document worden gespecificeerd.

Het XML bestand op zich bevat geen instructies met betrekking tot het *uiterlijk* van het digitale document. Via een XSL style sheet (XSL, Extensible Style Sheet Language) wordt aangegeven hoe het digitale document er uit moet zien als deze wordt weergegeven. Ten slotte kan ook het *gedrag* van het digitale document worden weergegeven met XML. Eenvoudig gedrag zoals hyperlinks en e-mail adressen worden weergegeven door middel van geschikte 'tagging'. Complexer gedrag is moeilijker weer te geven met behulp van XML en moet apart worden bekeken.

Door deze mogelijkheden is XML in veel gevallen een geschikte keuze als doelformaat voor duurzame bewaring van digitale documenten.

De conversie naar XML kan op verschillende manieren worden geïmplementeerd en de specifieke kenmerken van de implementatie zijn grotendeels afhankelijk van het digitale documenttype. Er zijn conversietools nodig om de bestanden om te zetten van het oorspronkelijke formaat naar XML. Er zijn verschillende soorten commerciële conversietools verkrijgbaar die geschikt kunnen zijn, maar de kwaliteit van de output wisselt sterk, zodat de tool zorgvuldig moet worden gekozen om ervoor te zorgen dat de output voldoet aan de eisen ten aanzien van duurzame bewaring. Dergelijke eisen kunnen voor eenvoudige digitale documenten bijvoorbeeld een geschikte DTD zijn en voor complexere digitale documenten de mogelijkheid om de digitale componenten van het digitale document gekoppeld te houden, zodat ze opnieuw kunnen worden samengesteld en daarbij voldoende overeenkomst vertonen met het origineel.

Is XML als bestandsformaat geschikt voor het bewaren van e-mail?

XML kan worden gebruikt als bestandsformaat en is een goede keuze voor het duurzaam bewaren van e-mail: XML is een open standaard, interoperabel en is in staat is vrijwel alle vijf kenmerken van een digitaal document weer te geven. Doordat e-mail en XML een aantal gemeenschappelijke kenmerken hebben, leent XML zich bij uitstek voor het duurzaam opslaan van e-mailberichten. E-mailberichten moeten immers voldoen aan een technische standaard, het MIME formaat, om op verschillende platforms interoperabel te kunnen zijn. Dit message formaat heeft een duidelijke opmaak en definieert de componenten van een e-mail transmissiebestand¹⁵. Dit formaat wordt beheerd door een non-profitorganisatie, de Internet Engineering Task Force¹⁶, en is goed gedefinieerd, goed gestructureerd en tekstgebaseerd. Conversie van dit formaat naar XML is om die reden redelijk eenvoudig.

XML is niet alleen een goed doelformaat voor het achteraf converteren van e-mailberichten, maar het is tevens geschikt voor het direct in XML genereren van nieuwe e-mailberichten. Het gebruik van XML voor het maken van e-mailberichten heeft wel een beperking. XML is een niet-binair

¹⁵ De standaard die tegenwoordig bij e-mail wordt gebruikt, is RFC 2822, waarbij de MIME-extensies zijn gespecificeerd in RFC 2045 – 2049. De RFC2822-specificatie zelf is te vinden op <http://www.ietf.org/rfc/rfc2822.txt?number=2822>. Zie ook <http://www.nacs.uci.edu/indiv/ehood/MIME/2045/rfc2045.html> voor meer informatie over standaard- (en niet-standaard-) MIME-headers en de continue ontwikkeling ervan.

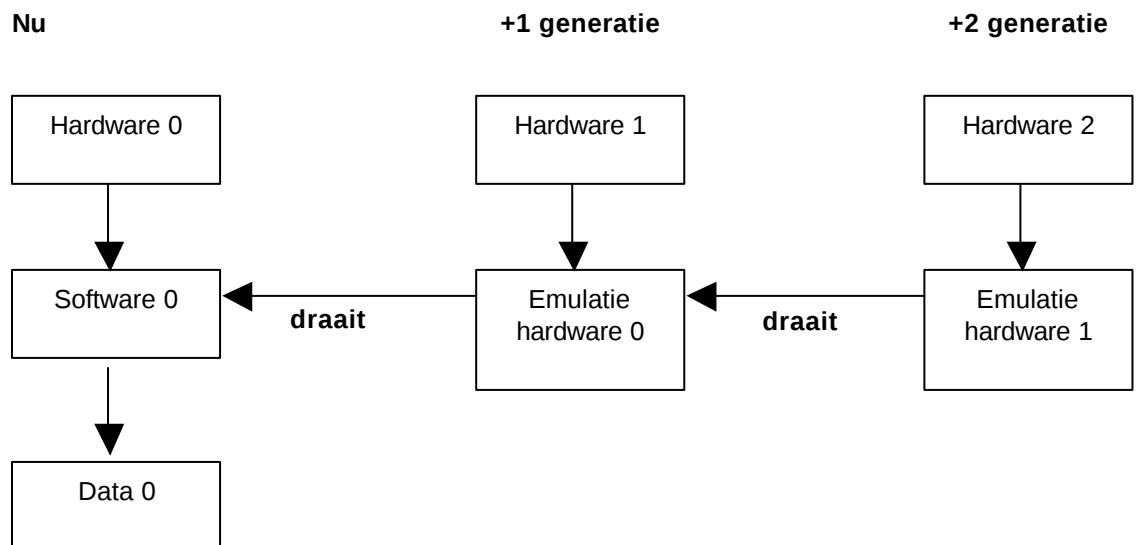
¹⁶ Zie <http://www.ietf.org/>

bestandsformaat en dat betekent dat deze aanpak zich niet leent voor het duurzaam bewaren van binaire attachments of bijlagen (zoals Word documenten, Excel spreadsheets en PowerPoint presentaties). Deze kunnen alleen in een naar tekst omgezette vorm, dus gecodeerd, in XML worden opgeslagen, (zie ook de paragraaf Zijn Wrapper en kapstok geschikt voor het bewaren van e-mail?).

4.4 Emulatie als bewaarstrategie

Emulatie is voor het bewaren van digitale documenten een technisch complexere strategie dan migratie. Evenals bij migratie zijn er verschillende manieren waarop een emulatiestrategie kan worden geïmplementeerd. De complexiteit daarvan verschilt al naargelang het vereiste emulatietype en emulatie niveau. Bij emulatie wordt de oude, oorspronkelijke computeromgeving opnieuw gecreëerd op een nieuwe, moderne computer.

Het volgende schema toont de relaties tussen de hardware, software en data bij een emulatietoepassing:



Afbeelding 6. Basisschema emulatie

De theorie achter emulatie als bewaarstrategie voor het duurzaam bewaren van digitale documenten is gelegen in de eisen die men stelt aan digitale documenten. Velen zijn van mening dat de authenticiteit en integriteit van een digitaal document op den duur alleen kunnen worden gegarandeerd door dat digitale document toegankelijk te houden in de oorspronkelijke omgeving, dat wil zeggen met het oorspronkelijke besturingssysteem en programma. Het bestand wordt dan niet beschadigd door conversie of omzetting. Het oorspronkelijke gedrag en uiterlijk (de 'look and feel') van de omgeving en het digitale document blijven behouden. Hiervoor moet dan wel een emulator worden gebouwd die de oude software- of hardwareomgeving (inclusief drivers voor randapparatuur) emuleert op een nieuwe, moderne computer.

Anders dan bij migratie wordt bij emulatie het bestandsformaat en de bitstream niet gewijzigd en is er geen conversie of omzetting nodig. Dit betekent dat het digitale

document in principe altijd op dezelfde wijze zal worden weergegeven, ongeacht de tijd die verstreken is en ongeacht het toekomstige computerplatform waarop de emulatie software zal draaien.

In de loop der tijd zijn er diverse vormen van emulatie ontwikkeld:

1. hardware-emulatie;
2. software-emulatie;
3. UVC-benadering (UVC, Universele Virtuele Computer)¹⁷:
 - o datapreservatie;
 - o programmapreservatie.

Hieronder wordt elke optie uitvoeriger beschreven.

4.4.1. Hardware-emulatie

Hardware-emulatie is een van de oudste en bekendste concepten op het gebied van computeremulatie. Bij hardware-emulatie, voor de vele volhardende fans van de Commodore 64 of de BBC Micro een begrip, wordt een oude en mogelijk in onbruik geraakte hardwareomgeving opnieuw gecreëerd op een moderne computer. Daartoe wordt een programma geschreven waarmee de structuur van een oude computer op een nieuw computerplatform wordt nagebootst. Een voorbeeld is het draaien van een Commodore 64-emulator op een moderne Pentium 4 computer. De gebruiker kan dan in de nieuwe omgeving alles doen wat ook mogelijk was in de oude. De Pentium 4, die zich gedraagt als een Commodore, voert applicaties uit, begrijpt opdrachten en leest bestanden op dezelfde wijze als de Commodore 64 dat deed. De bestanden worden dus hetzelfde weergegeven en kunnen hetzelfde worden gebruikt als toen ze voor het eerst werden gemaakt en toegepast in de oorspronkelijke omgeving.

Via hardware-emulatie kan de gehele computeromgeving opnieuw worden gecreëerd, zodat de het gedrag en het uiterlijk (look and feel) van een digitaal document behouden blijven. Dit kan nodig zijn voor uiterst complexe digitale documenttypen, waarvan functionaliteit en gedrag belangrijke gebruikaspecten waren. Hierbij kan worden gedacht aan complexe multimedia digitale documenten of CAD objecten. Voor dergelijke digitale documenten is hardware-emulatie wellicht de enige manier om te waarborgen dat de authenticiteit en integriteit behouden kunnen blijven. Het grote voordeel van hardware-emulatie is dat het oorspronkelijke digitale document geen migratie of omzetting behoeft te ondergaan. De keerzijde is dat hardware-emulatie geen eenvoudige opgave is. Het is arbeidsintensief en daarmee kostbaar om de volledige oorspronkelijke computeromgeving te emuleren.

4.4.2. Software-emulatie

Bij software-emulatie wordt het probleem op soortgelijke wijze aangepakt, maar dan op een hoger abstractieniveau. Alleen de software die op de oude hardware draaide hoeft opnieuw te worden gecreëerd, niet de hardwareomgeving zelf. De digitale documenten worden dus geopend op een moderne computer, maar met behulp van speciale software die de oorspronkelijke softwareomgeving nabootst. Zo zou op een computer waarop gewoonlijk Unix draait een softwareapplicatie kunnen draaien waarmee het Windows besturingssysteem wordt geëmuleerd. Applicaties die onder

¹⁷

Het concept van een Virtuele Machine/Computer bestaat al vele jaren; het concept van een Universele Virtuele Machine/Computer is specifiek en is afkomstig van Raymond Lorie van het IBM Almaden Research Centre.

Windows draaien, kunnen dan via de Windows emulator op de normale manier worden gebruikt.

Ook hier geldt dat het voordeel is dat het oorspronkelijke digitale document geen migratie of conversie ondergaat, maar intact blijft. Het nadeel is nog steeds dat emulatie arbeidsintensief en daarmee een dure bewaarstrategie is.

Voor zowel hardware- als software-emulatie moet een emulatiespecificatie worden gemaakt voor een bepaald aspect van de oorspronkelijke computeromgeving. Deze specificatie moet dan samen met de digitale objecten worden opgeslagen, zodat deze kunnen worden weergegeven. In principe is één emulatorspecificatie voldoende voor een aantal digitale documenten, mits voor deze digitale documenten hetzelfde type hardware of software werd gebruikt. De emulatiespecificatie moet echter uiterst nauwkeurig zijn en velen zijn van mening dat deze strategie om die reden te complex en te foutgevoelig is. Mogelijk wordt de specificatie pas volledig getest als deze nodig is voor de daadwerkelijke emulatie. Tegen die tijd kan de oorspronkelijke applicatie waarmee de digitale documenten werden gemaakt al niet meer bestaan. Als er bij de emulatiespecificatie een probleem optreedt, terwijl de applicatie niet meer bestaat, kan de omgeving niet met zekerheid worden geëmuleerd en kunnen de authenticiteit en integriteit van de digitale documenten in het geding zijn.

Zijn hardware- en/of software-emulatie geschikt voor het bewaren van e-mail?

Hardware- en software-emulatie zijn voor e-mailberichten een te zwaar middel, omdat het uiterlijk (look and feel) van een e-mailbericht van ondergeschikt belang is, zoals reeds aangegeven in de authenticiteitseisen van e-mail in hoofdstuk 3. Hoewel e-mail interoperabel is, is deze eigenschap alleen van toepassing op het transmissiebestand, niet op de uiteindelijke versie van het digitale document zoals die door het e-mail programma wordt verwerkt en weergegeven. Het transmissiebestand moet door een willekeurig aantal e-mailapplicaties kunnen worden gelezen, omdat de zender nooit zeker weet welke e-mailapplicatie de ontvanger gebruikt. Mede vanwege die interoperabiliteit is e-mail zo'n geweldig succes: het doet er niet toe welke e-mailapplicatie wordt gebruikt, aangezien alle applicaties het transmissiebestand kunnen interpreteren. Daar houdt de interoperabiliteit echter op. E-mailapplicaties converteren het transmissiebestand namelijk naar een ander formaat (in de regel het leverancierseigen formaat van de e-mailapplicatie), zodat het digitale document op het scherm kan worden weergegeven. Afhankelijk van de instellingen, mogelijkheden en interface zullen e-mailapplicaties een e-mailbericht op een andere wijze decoderen. Zo kunnen specifieke eigenschappen van de ene e-mailapplicatie (zoals bepaalde markeringen) bij een andere applicatie niet standaard aanwezig zijn, zodat de ontvanger bepaalde informatie in het transmissiebestand niet te zien krijgt. Het gedrag en uiterlijk van een verzonden e-mailbericht zullen dus ook niet altijd hetzelfde zijn als het ontvangen e-mailbericht. Er is dan ook geen reden om een arbeidsintensieve en daarmee dure bewaarstrategie als emulatie, die bij uitstek geschikt is voor behoud van het oorspronkelijke gedrag en uiterlijk van een digitaal document, in te zetten voor het duurzaam bewaren van e-mailberichten.

4.4.3. De Universele Virtuele Computer-strategie (UVC)

Emulatie waarbij gebruik wordt gemaakt van de UVC, wijkt enigszins af van het oorspronkelijke emulatieconcept. Er moet weliswaar nog steeds een emulator worden geschreven, maar in dit geval voor een niet-bestaande, virtuele computer, de UVC genaamd (Universal Virtual Computer). De UVC is een computer met een zo eenvoudige architectuur en basale instructieset dat

elke softwareontwikkelaar in de toekomst in staat is een emulator voor de UVC te schrijven. De UVC wordt vervolgens gebruikt om een programma (UVC dataformaat decoder) te draaien dat het originele digitale document als input neemt en als output een Logische Data Beschrijving (LDB) van de data oplevert. Deze logische data beschrijving is opgebouwd uit tags die additionele informatie geven over de inhoud van het digitale document. De additionele semantische informatie is zo opgezet dat men in de toekomst in staat is de logische data beschrijving te interpreteren zonder additionele hulpmiddelen. Vervolgens kan de logische data beschrijving worden verwerkt door een te bouwen viewer, waarmee het oorspronkelijk digitale document op het scherm wordt getoond.

De Universele Virtuele Computer bewaarstrategie (UVC) is een variant op de emulatiestrategie en moet in elke toekomstige computeromgeving kunnen worden toegepast. De strategie leunt slechts gedeeltelijk op emulatie en omvat enkele aspecten van de migratiestrategie. Met de UVC worden oorspronkelijke databestanden omgezet in een Logische Data Beschrijving (LDB) via een programma dat in de UVC programmeertaal is geschreven. Deze LDB is een zelfstandig, zelfbeschrijvend en duidelijk gestructureerd dataformaat dat alle informatie bevat die nodig is om het digitale document in de toekomst opnieuw te kunnen samenstellen.

Een voordeel van de UVC benadering ten opzichte van de klassieke emulatie benadering is dat de UVC data formaat decoder direct kan worden getest. Vastgesteld kan worden of de decoder aan de eisen voldoet. Zo ja, dan kan deze decoder worden bewaard totdat hij in de toekomst nodig is om een eveneens bewaard databestand opnieuw te representeren op een dan gangbare computer.

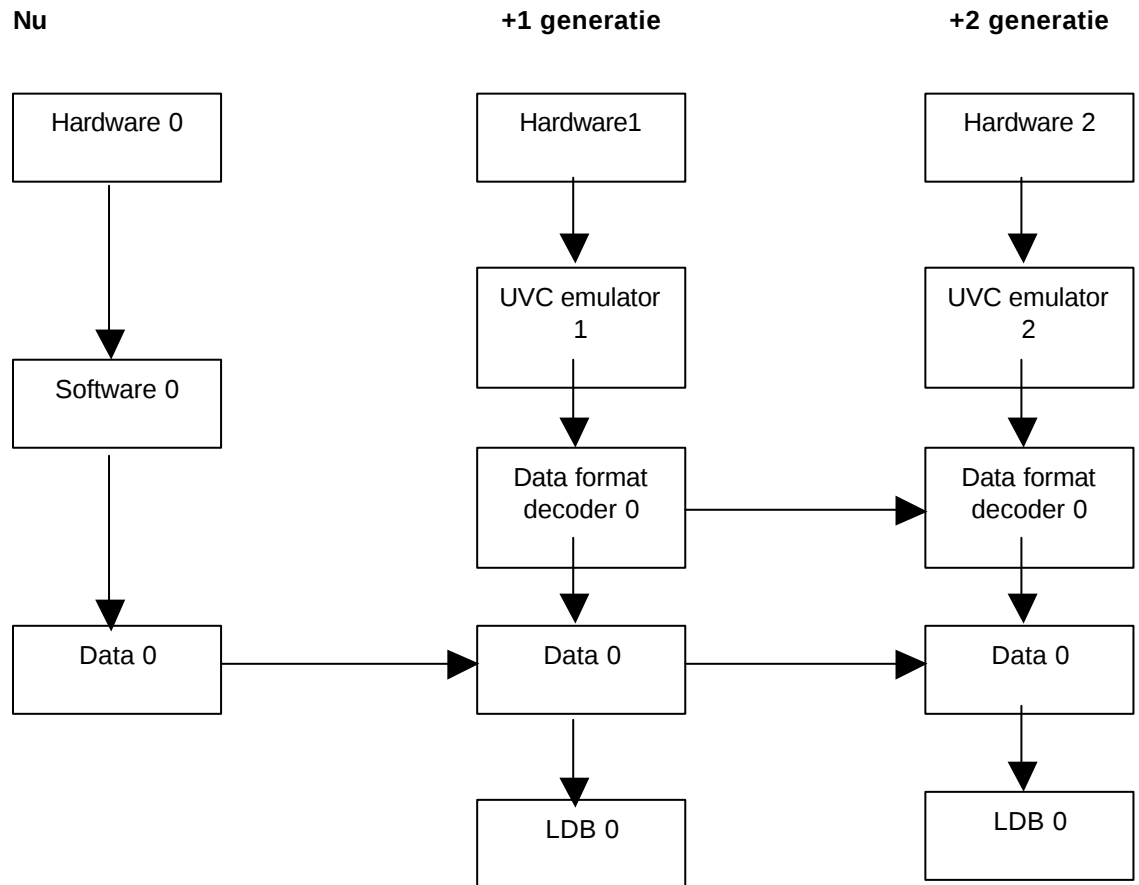
IBM ontwikkelt de UVC momenteel als alternatieve strategie voor duurzame bewaring. In de Koninklijke Bibliotheek in Nederland is in 2002 al succesvol een Proof of Concept uitgevoerd met publicaties, in het bijzonder met PDF bestanden.

Wat betreft de toepassing van de UVC zijn er twee implementatievormen: data preservatie en programma preservatie.

UVC datapreservatie

'Datapreservatie' is de eerste en eenvoudigste implementatievorm van de UVC-strategie. Hierbij worden de data – het oorspronkelijke bestand in het oorspronkelijke formaat – opgeslagen met een programma dat de data uit de bitstream kan halen en deze gegevens op eenvoudige en technologisch onafhankelijke wijze beschrijft, zodat de gegevens door een viewer kunnen worden verwerkt.

Het oorspronkelijke bestand – bijvoorbeeld een JPEG bestand – wordt samen met het specifieke UVC data formaat decoder programma voor JPEG opgeslagen. In de toekomst wordt dit UVC JPEG programma uitgevoerd op de UVC emulator. Het UVC JPEG programma leest de bitstream van het oorspronkelijke bestand en levert een LDB als output (Logische Data Beschrijving). De LDB wordt op een toekomstig computerplatform weergegeven met behulp van een viewer die in de toekomst kan worden ontwikkeld op basis van de LDB.



Afbeelding 7. Schema van de Universal Virtual Computer

Bij deze strategie wordt de oorspronkelijke bitstroom niet gewijzigd en wordt het nieuwe bestand (de LDB), gemaakt bij het uitvoeren van het UVC data formaat decoder programma, niet opgeslagen. De LDB wordt getoond via een viewer. Het formaat en de structuur van de Logische Data Beschrijving zijn zo duidelijk dat er op eenvoudige wijze een viewer voor moet kunnen worden ontworpen en geschreven. Als dat nodig is, kunnen er nieuwe viewers worden ontwikkeld voor toekomstige computerplatforms.

Momenteel is er voor elk type LDB een aparte viewer nodig. Dit betekent dat in theorie mogelijk honderden viewers zouden kunnen worden gebruikt. In de praktijk zal het aantal verschillende formaten dat wordt geaccepteerd door de archiefbewaarplaatsen, worden beperkt op basis van de regeling Geordende en toegankelijke staat archiefbescheiden. In de volgende fase van de UVC ontwikkeling zullen klassen van objecten worden gevormd die volgens dezelfde logica functioneren. Een dergelijke klasse van objecten (bijvoorbeeld de verschillende bestandsformaten voor plaatjes) levert één LDB, waarmee nog maar één viewer hoeft te worden ontwikkeld. Wel blijft het noodzakelijk voor deze bestandsformaten een eigen UVC dataformaat decodeerprogramma te ontwikkelen.

Deze strategie, waarbij het oorspronkelijke digitale document wordt bewaard, is geschikt voor digitale documenten waarbij de functionaliteit van de applicatie losstaat van het digitale document en niet meer nodig is om het document te bewerken. Desondanks kunnen, afhankelijk van het digitale document, enkele belangrijke aspecten van het digitale object afhankelijk zijn geweest van eigenschappen van de oorspronkelijke applicatie. Als dat het geval is, moeten deze eveneens in de Logische Data Beschrijving worden opgenomen of beschreven.

Een nadeel van de UVC emulatie benadering is dat er UVC data formaat decoder programma's geschreven moeten worden voor elk bestandstype (om de logische databeschrijving te genereren). Voor elke nieuwe generatie hardware die zozeer afwijkt van vorige generaties dat de oude UVC emulator er niet meer betrouwbaar op kan draaien, moet een nieuwe emulator worden geschreven. De verantwoordelijkheid hiervoor berust bij de marktpartijen, momenteel IBM. IBM heeft echter het voornemen van de UVC een open standaard te maken.

Momenteel bestaan er alleen UVCformaat decodeer programma's voor PDF documenten en Excel spreadsheets. Gezien de grote verscheidenheid aan bestandsformaten en digitale documenttypen moeten er zo snel mogelijk grote aantallen decodeer programma's worden ontwikkeld, wil de UVC de komende jaren een haalbare en bruikbare strategie zijn voor het duurzaam bewaren van digitale documenttypen¹⁸. Het uiteindelijke succes van de UVC strategie is mede afhankelijk van de mate waarin deze strategie door de software- en computerbranche wordt geaccepteerd. Softwareleveranciers zouden voor hun software een UVC data formaat decoder moeten ontwikkelen dat op basis van het oorspronkelijke bestand een logische data beschrijving kan maken. Op dat moment zal de UVC strategie een hoge vlucht kunnen nemen.

Is UVC datapreservatie geschikt voor het bewaren van e-mail?

Bij UVC-emulatie wordt de omgeving gedeeltelijk geëmuleerd en het bestand gedeeltelijk geconverteerd. Data preservatie met de UVC is veelbelovend voor duurzame bewaring, maar verkeert nog in de ontwerpfase en kan daarom de eerste jaren nog niet in de praktijk worden gebruikt.

Het inzetten van de UVC-datapreservatiestrategie voor het bewaren van e-mail is conceptueel gezien aantrekkelijk. Het vertalen van het transmissiebestand naar een logische databeschrijving, die overkomsten vertoont met XML, en die vervolgens op een gewenst moment in de tijd aan de hand van een te ontwikkelen viewer op het scherm kan worden gerepresenteerd, kan een bruikbare strategie zijn. Het uitvoerbestand is leesbaar voor mens en machine en is niet langer afhankelijk van de oorspronkelijke e-mail programmatuur. Wel ligt hier het risico van afhankelijkheid van de UVC van IBM op de loer. IBM heeft echter het voornemen geuit om de UVC als Open Source Software te willen inzetten. Het grote praktische probleem is echter dat er nog geen UVC programma voor e-mail is ontwikkeld en getest voor gebruik in de praktijk.

¹⁸

Een andere benadering, Migration On Request, is onlangs geponeerd door het CAMiLEON project in Groot-Brittannië. Dit voorstel heeft veel gemeen met de UVC-strategie, maar de nadruk ligt op het conversieprogramma, dat in C is geschreven en naar verwachting de tand des tijds zal doorstaan. Hierdoor is er geen Universele Virtuele Computer nodig, waar het UVC-conversieprogramma afhankelijk van is. Zie <http://www.si.umich.edu/CAMiLEON/reports/migreq.pdf> voor meer informatie.

UVC: programmapreservatie

'Programmapreservatie' werkt op dezelfde manier als datapreservatie, maar heeft meer overeenkomsten met de volledige emulatiestrategie. Bij toepassing van de UVC-strategie voor programmapreservatie draait de UVC emulator wederom op een toekomstige computer, maar in plaats van een UVC data formaat decodeer programma waarmee de databestanden worden gelezen en geconverteerd naar een LDB wordt voor het lezen en openen van die bestanden een software emulator uitgevoerd die de vereiste applicatie nabootst. De bestanden worden dan geopend in hun 'oorspronkelijke' softwareomgeving (dat wil zeggen een geëmuleerde versie van deze oorspronkelijke omgeving) met als voordeel dat eventuele gedragskenmerken die deel uitmaakten van de applicatie beschikbaar zijn.

Is UVC programmapreservatie geschikt voor het bewaren van e-mail?

Deze benadering bevindt zich nog in een conceptuele fase en zal zich in de praktijk nog moeten bewijzen. Echter, gezien de interoperabiliteit van het transmissiebestand en de 'onafhankelijkheid' van e-mail van het gebruikte e-mailprogramma is deze implementatie van de UVC evenmin als de UVC datapreservatie niet direct de meest voor de hand liggende optie voor het langdurig bewaren van e-mailberichten.

4.5 Conclusie

Op basis van de voor- en nadelen van de verschillende bewaarstrategieën voor het duurzaam bewaren van e-mailberichten, gecombineerd met de resultaten van het experimenteel onderzoek van Testbed Digitale Bewaring, is de conclusie dat op dit moment de toepassing van XML de meest geschikte strategie is voor het duurzaam bewaren van e-mailberichten.

Migratie van e-mail naar een volgende versie (achterwaartse compatibiliteit) van de applicatie heeft als nadeel dat het e-mailbericht in het leverancierseigen formaat blijft opgeslagen, hetgeen niet wenselijk is vanuit het oogpunt van digitale duurzaamheid. Het is juist zaak onafhankelijk te worden van de veroudering van de hard- en software, waarin het digitale document is gecreëerd. Migratie moet voorts regelmatig worden herhaald en dat terwijl iedere migratieslag het risico met zich meebrengt dat het digitale document een wijziging ondergaat, die de authenticiteit en integriteit van het digitale document aantasten.

Bij toepassing van interoperabiliteit, als tweede vorm van migratie, blijft de e-mail opgeslagen in het leverancierseigen formaat, hetgeen niet wenselijk is. Bij de omzetting naar bijvoorbeeld ASCII of RTF is het risico aanwezig dat belangrijke kenmerken van het digitale document verloren gaan.

De conversie naar standaarden ten slotte (de derde vorm van migratie) biedt perspectief afhankelijk van de gekozen standaard, waarbij een open, niet-leverancierseigen standaard de voorkeur heeft. Een omzetting naar PDF, een voorbeeld van een leverancierseigen standaard, wordt niet aanbevolen. Enerzijds om de ongewenste afhankelijkheid van de eigenaar van de standaard te voorkomen (in dit geval Adobe) en anderzijds omdat PDF niet in staat is alle essentiële (header-) gegevens van een e-mailbericht te registreren. De authenticiteit van het e-mailbericht is daarmee in het geding.

XML is evenals het e-mail transmissiebestand interoperabel en onafhankelijk van de gebruikte hard- en software. XML is tekst gebaseerd en is daarmee 'leesbaar' voor mens en computer. Het is voorts een open standaard, hetgeen inhoudt dat de standaard niet eenzijdig door een marktpartij kan worden gewijzigd. Tevens wordt bij XML een scheiding aangebracht tussen inhoud/structuur en uiterlijk, hetgeen de

duurzaamheid ten goede komt, omdat juist het uiterlijk voor veel afhankelijkheid van de software zorgt.

Emulatie als bewaarstrategie biedt zeker perspectief, maar is voor e-mailberichten een te zwaar en kostbaar middel. Emulatie is met name gewenst in die gevallen waarin de functionaliteit van de software, waarmee het digitale document in eerste instantie is gecreëerd, moet worden behouden: het oorspronkelijke uiterlijk (de look en feel) moet intact blijven. Bij e-mail is dit geenszins een vereiste aangezien ieder e-mailprogramma het transmissiebestand op eigen wijze verwerkt en representeert op het scherm.

In het volgende hoofdstuk wordt een specifieke strategie en implementatie beschreven voor het duurzaam bewaren van e-mailberichten door gebruik te maken van XML.

5 *Aanbevolen e-mail benadering*

In hoofdstuk 4 zijn verschillende bewaarstrategieën beschreven en afgezet tegen het documenttype e-mail, waarbij XML als het meest veelbelovend uit de bus kwam. In dit hoofdstuk wordt beschreven op welke wijze XML kan worden geïmplementeerd.

5.1 Inleiding

Op basis van experimenten uitgevoerd door Testbed is op dit moment de meest geschikte bewaarstrategie voor e-mail vastgesteld, te weten XML. In hoofdstuk 4 is uitvoerig uit de doeken gedaan wat de voordelen zijn van het gebruik van XML voor het duurzaam bewaren van e-mailberichten. In dit hoofdstuk wordt ingegaan op de vraag op welk wijze XML kan worden ingezet als bewaarstrategie voor e-mail.

Voor een beter inzicht in de door Testbed voorgestelde benadering en de wijze waarop de conversie naar XML plaatsvindt, is het noodzakelijk kennis te hebben van de eigenschappen van e-mail als computerbestand, het zogenoemde transmissiebestand. Dit is uitvoerig ter sprake gekomen in hoofdstuk 3 en wordt hier in verkorte vorm herhaald.

5.2 E-mail als transmissiebestand

Het transmissiebestand is de meest complete en definitieve bron voor inhoud en metadata van het oorspronkelijke e-mailbericht. Het bevat alle informatie die bij het uitvoeren van een transactie in tijd en ruimte is gecommuniceerd. Bij reeds verzonden en ontvangen berichten is het transmissiebestand het beste uitgangspunt voor een conversie naar XML. Het transmissiebestand, bedoeld om de interoperabiliteit te garanderen, is immers onafhankelijk van de applicatie waarmee het werd gemaakt. Een conversie vanuit het transmissiebestand naar XML heeft minder voeten in de aarde dan een conversie vanuit het leverancierseigen formaat (bijv. *.msg van Outlook), omdat het bestand dan niet eerst hoeft te worden teruggezet naar het oorspronkelijke formaat van het transmissiebestand. Bij e-mailberichten in het leverancierseigen formaat vindt tevens interactie plaats met de omgeving waarin ze worden weergegeven (en onttrekken daar bepaalde informatie aan). Zo nemen sommige e-mailapplicaties het volledige e-mail adres van een afzender over van het systeem en het netwerk waarop ze worden gehost. Als het e-mailbericht in het leverancierseigen formaat fysiek wordt overgebracht naar een ander systeem kan het e-mail adres van de afzender veranderingen ondergaan. Het adres bevat dan mogelijk het verkeerde domein, waardoor de context van het e-mailbericht wordt beïnvloed.

Het transmissiebestand bestaat uit verschillende onderdelen. Doordat de onderdelen de totale opmaak van de e-mail bepalen, zal elke e-mail er enigszins anders uitzien, afhankelijk van de inhoud van de onderdelen van het transmissiebestand. De basisprincipes van e-mail zijn echter steeds hetzelfde. Alle e-mailberichten bestaan uit headers en body en kunnen voorts bijlagen, afbeeldingen en andere items bevatten. Elk oorspronkelijk bestand, de relatie ervan met de andere bestanden, en elk onderdeel van het bericht worden duidelijk aangegeven.

Een belangrijk verschil tussen inkomende en uitgaande transmissiebestanden is dat inkomende bestanden 'ontvangst headers' bevatten en uitgaande bestanden niet. In de 'ontvangst header' staan de datum en het tijdstip waarop het bericht bij verzending door de mailserver(s) is gegaan.

5.3 Conversieprocedures

De conversie naar XML kan met meerdere tools worden uitgevoerd. In de handel verkrijgbare tools moeten wel worden beoordeeld op geschiktheid, aangezien de door deze tools gemaakte XML-bestanden van elkaar verschillen. Organisaties kunnen er de voorkeur aan geven een eigen conversietool te ontwikkelen en de verantwoordelijkheid hiervoor met verwante organisaties te delen. Testbed kan desgewenst nader adviseren over geschikte conversietools.

Er zijn twee verschillende scenario's mogelijk om naar XML te converteren: Post-use (achteraf in XML omzetten) en pre-use (direct in XML genereren). Het post-use-scenario is bedoeld voor bestaande e-mailberichten (zowel reeds verzonden als inkomende berichten) die gedurende een nog onbepaalde tijd moeten worden bewaard (deze berichten worden dus **achteraf** omgezet naar XML). Het pre-use-scenario kan worden toegepast bij nieuwe uitgaande e-mailberichten en is de eerste stap in de richting van het maken en duurzaam opslaan van formele e-mailberichten (de berichten worden **direct, aan de bron**, gegenereerd in XML).

Met de gekozen tool moet het overigens mogelijk zijn om aanvullende informatie, bijvoorbeeld 'dossier' of 'werkproces', die niet in het transmissiebestand aanwezig is, op te slaan in het XML-bestand. Deze voor beide scenario's noodzakelijke extra gegevens zijn metadata voor het bewaren van digitale documenten. Voor het pre-use-scenario is een meer geïntegreerde tool vereist, die verderop in een aparte paragraaf wordt besproken.

Er zijn drie categorieën e-mail die in aanmerking komen voor een conversie naar XML:

- o Inkomende e-mailberichten. E-mailberichten die worden ontvangen kunnen alleen achteraf worden omgezet naar XML, waarbij het transmissiebestand als bron voor de conversie wordt gebruikt.
- o Uitgaande e-mailberichten (bestaande). E-mailberichten die in het verleden zijn verstuurd en in aanmerking komen voor duurzame bewaring kunnen achteraf worden omgezet naar XML. Ook in dit geval dient het transmissiebestand als bron voor de conversie.
- o Uitgaande e-mailberichten (nieuwe). Nieuwe uitgaande e-mailberichten is de enige categorie die in aanmerking komt voor de pre-use benadering. Nieuwe uitgaande e-mail kan direct aan de bron worden gegenereerd in XML. Een conversie achteraf is dan niet nodig.

De eigenlijke XML-bestanden voor deze drie categorieën e-mailberichten verschillen nauwelijks van elkaar. Bij de ontvangen en bestaande en reeds verzonden e-mailberichten is het uitgangspunt voor de conversie naar XML het transmissiebestand. Het enige belangrijke verschil tussen deze twee soorten berichten is de ontvangheader, die in de vorige paragraaf is besproken. Beide soorten e-mailberichten kunnen op dezelfde manier naar XML worden geconverteerd. Verderop in deze paragraaf wordt hier nader op ingegaan¹⁹.

Voor de zogenoemde 'pre-use verzonden e-mailberichten' geldt een andere aanpak. Om het eindgebruikers van e-mail zo eenvoudig mogelijk te maken, is het aan te bevelen dat organisaties hun huidige e-mailapplicatie uitbreiden met de mogelijkheid om nieuwe uitgaande e-mailberichten direct te genereren in XML. Testbed heeft daartoe Microsoft Outlook als voorbeeld genomen om te laten zien hoe deze applicatie met een redelijk eenvoudige aanpassing geschikt kan worden gemaakt voor

¹⁹

Zie Appendix A voor informatie over de conversieprocessen van Testbed.

duurzame bewaring van e-mailberichten. Testbed heeft een add-in²⁰ ontwikkeld, waarbij de gebruiker in het geval van formele e-mail een aangepast venster te zien krijgt en waarbij hij verplicht wordt bepaalde metadata in te vullen.

The screenshot shows a window titled 'Formele E-mail' with a menu bar (Bestand, Bewerken, Extra, Beeld, Help) and a toolbar. A yellow warning box at the top states: 'Dit is een formele e-mail die zal worden opgeslagen. U dient alle metagegevens in te vullen. U kunt de e-mail controleren voordat deze verzonden wordt.' Below this are input fields for 'Aan...' (To), 'CC...', and 'Onderwerp' (Subject). The 'Aanhef' (Greeting) section has three icons and a large text area. At the bottom is the 'Afsluitingsfrase' (Closing phrase) field.

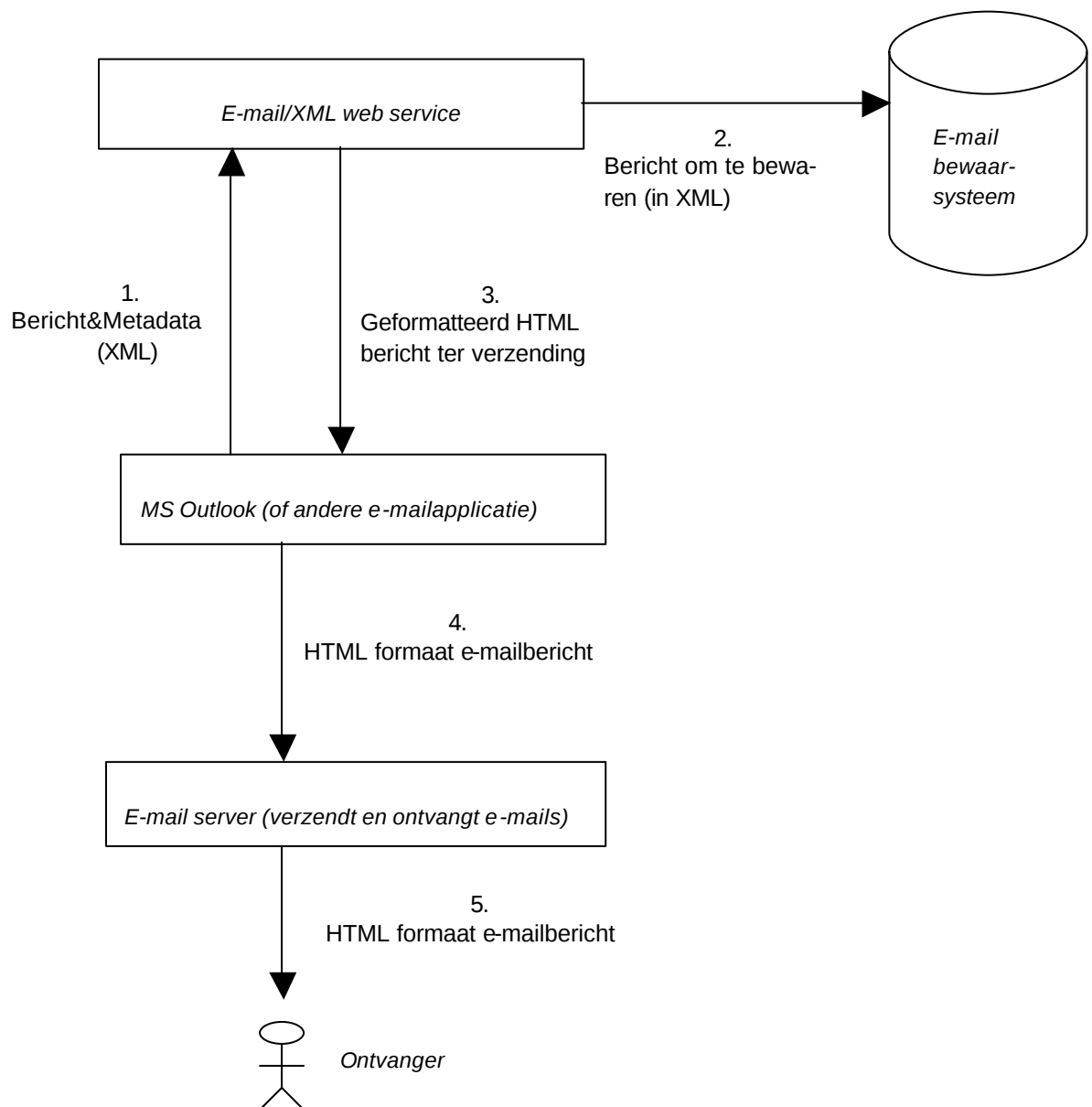
Afbeelding 8. Voorbeeld van een aangepast venster in Outlook.

²⁰ Zie Appendix A voor informatie over de 'E-mail naar XML-demo' van Testbed, die werd ontworpen als een add-in voor een bestaande e-mailapplicatie en waarmee e-mailberichten op de juiste wijze kunnen worden gemaakt in XML.

Afbeelding 9. Voorbeeld van verplicht in te vullen velden in een e-mailbericht.

Daarbij is er bewust voor gekozen het aantal verplicht in te vullen velden tot een minimum te beperken. In de door Testbed ontwikkelde demo gaat het om slechts twee velden, te weten dossier en werkproces. De persoonlijke gegevens behoeven slechts eenmalig te worden ingevuld en worden vervolgens door Outlook opgehaald.

De aangepaste versie van Outlook combineert de metadata en de inhoud van de e-mail in een XML bestand. Dit XMLbestand wordt op een aparte server opgeslagen, die controleert of de XML tegemoet komt aan het gedefinieerde XML-schema. Vervolgens wordt er een XSL stylesheet gebruikt om de omzetting van XML naar HTML te realiseren. Hierbij worden de inhoud van het e-mailbericht en de metadata verzameld, in een geformatteerde vorm weergegeven, waarbij de 'overall' lay-out van het e-mailbericht, de keuze van het lettertype, kleuren en een logo of afbeelding worden toegevoegd. Dit document wordt, nu in de vorm van een HTML document, teruggestuurd naar Outlook en kan vervolgens worden verstuurd.



Afbeelding 10. Omzetting naar XML met behulp van webservice.

Belangrijk is dat zoveel mogelijk wordt aangesloten bij de huidige werkwijze van de eindgebruikers om de acceptatie te vergroten.

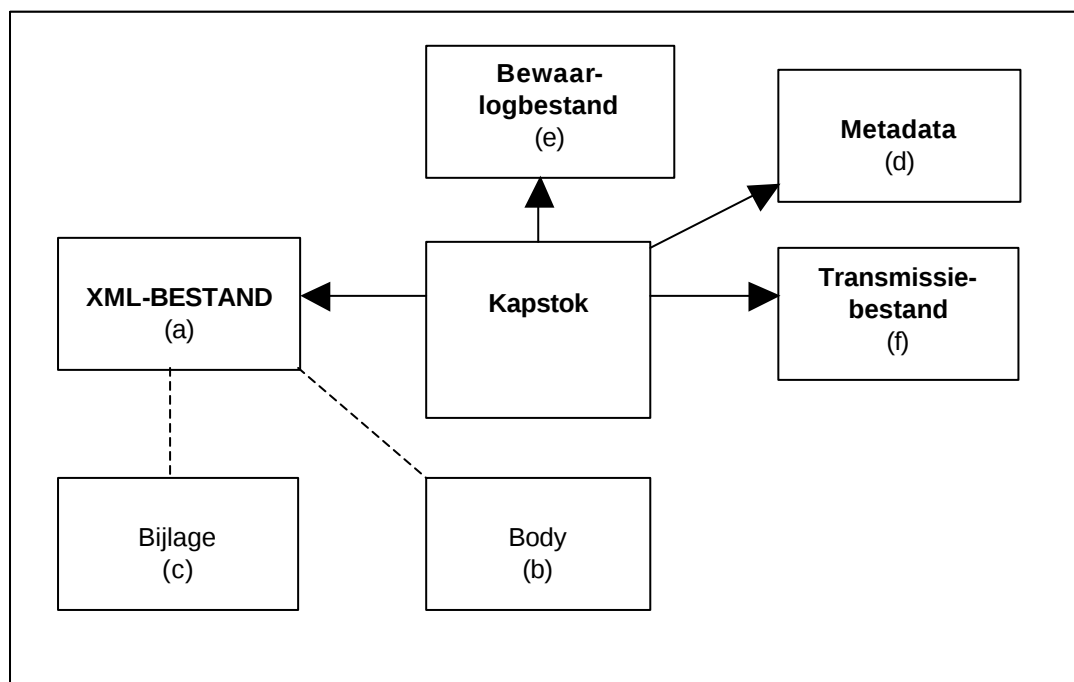
Aanvullende metadata moeten vóór de conversie worden verzameld en bij het bericht worden gevoegd. Het XML-bestand moet een basisset metadata bevatten (beschreven in het XML-schema). Waar nodig moeten koppelingen en verwijzingen worden ingevoegd (bijvoorbeeld in de bijbehorende body of bijlagen) en het XML-bestand moet rechtstreeks worden gekoppeld aan het bewaarobject (zie paragraaf 4.2). Bij deze strategie wordt XML als bestandsformaat gebruikt in combinatie met een wrapper- en kapstokbenadering.

5.4 Duurzaam bewaarde e-mail

Zoals eerder gezegd is XML een goed bestandsformaat in combinatie met de toepassing van een wrapper- of kapstokbenadering. Het naar XML omgezette bestand heeft een complexere structuur dan het transmissiebestand, maar het bevat ook beduidend meer informatie. Bij deze strategie wordt het XML-bestand gekoppeld aan een groter bewaarobject. Via dit bewaarobject blijven de koppelingen tussen de verschillende onderdelen van het e-mailbericht behouden.

Het bewaarobject bestaat uit minimaal vier componenten die via de kapstokconstructie met elkaar zijn verbonden:

- o XML-bestand;
- o Bewaarlogbestand;
- o Transmissiebestand;
- o (aanvullende) Metadata.



Afbeelding 11. Schematische weergave van het e-mail bewaarobject

De doorgetrokken lijnen in afbeelding 11 geven die componenten aan die volgens Testbed altijd aanwezig moeten zijn. De stippellijnen duiden de optionele componenten aan zoals de body die in HTML of RTF (en dus niet in XML) zijn opgemaakt, of eventuele bijlagen.

De vier hierboven genoemde componenten zijn nodig om het digitale document duurzaam te kunnen bewaren. In de volgende paragrafen worden alle componenten (a tot en met f) van het bewaarobject besproken.

Het XML-bestand (a)

Het XML-bestand is de belangrijkste component van het e-mail bewaarobject. Het bestand wordt gemaakt door de inhoud van het transmissiebestand te converteren of door gebruik te maken van de pre use benadering (waarmee e-mailberichten direct in XML worden aangemaakt). Het XML-bestand is overigens meer dan een van XML-tags voorziene versie van het transmissiebestand: de informatie die het bevat, is afkomstig van verschillende bronnen, bijvoorbeeld door een gebruiker toegevoegde metadata.

Het XML-bestand bevat een subset van de headers uit het transmissiebestand, waaronder alle bekende geadresseerden, de afzender, het onderwerp, de datum en de informatie van de header van het betreffende onderdeel. Een beperkte set aanvullende metadata wordt toegevoegd en 'getagged', evenals verwijzingen naar het transmissiebestand en eventuele bijlagen. De body of berichttekst kan worden weergegeven door een verwijzing naar een apart bestand of door rechtstreekse 'tagging'.

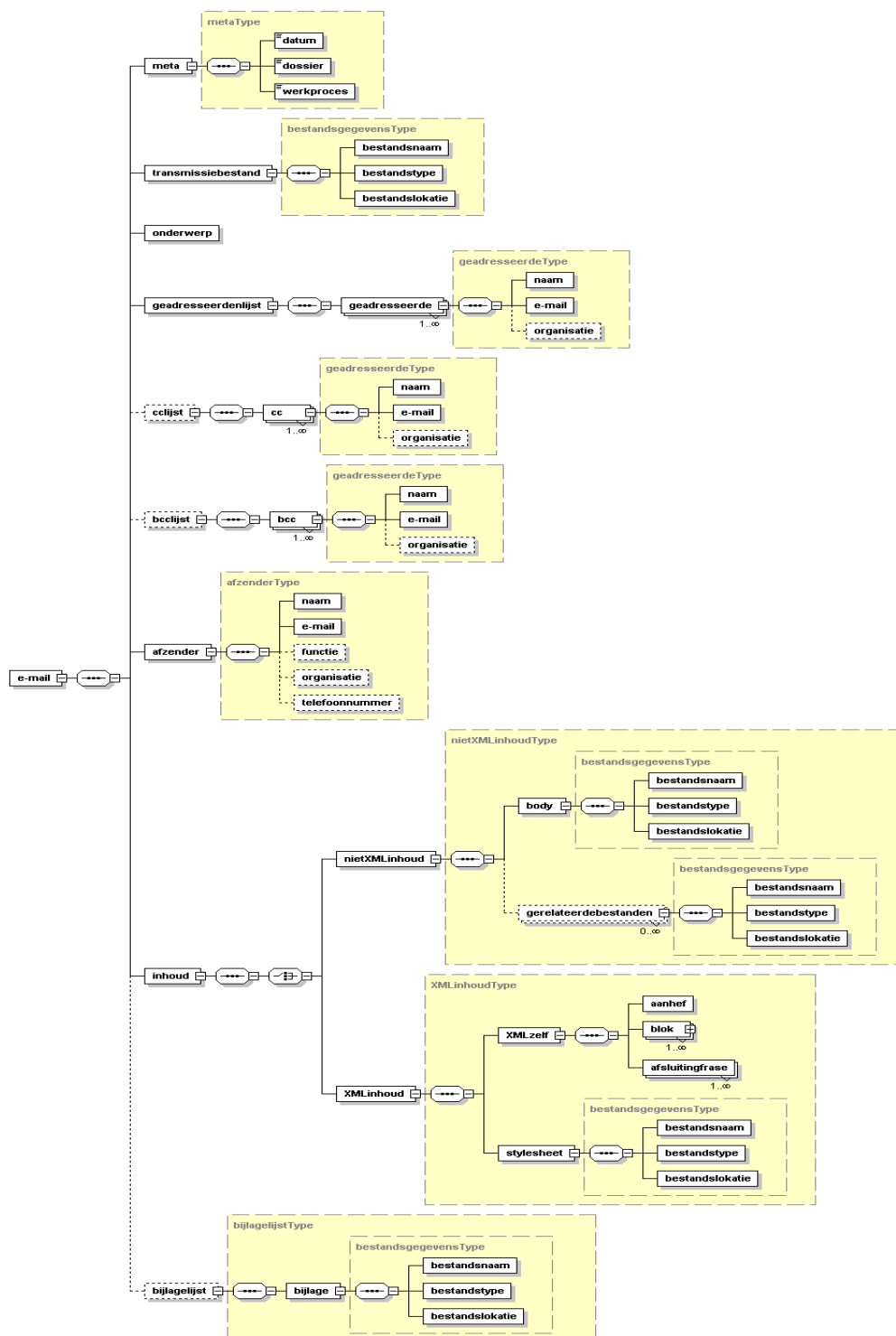
De aanvullende contextgegevens worden in het XML-bestand opgenomen in verband met de authenticiteitseisen. In de door Testbed ontwikkelde e-mail demo wordt uitgegaan van de volgende items:

- o conversiedatum (in het standaard MIME datum formaat);
- o dossier;
- o bedrijfsproces.

Het staat iedere organisatie vrij te bepalen welke metadata noodzakelijk zijn om de blijvende toegankelijkheid van de e-mailberichten te waarborgen.

In het geval van de pre-use benadering kan de opsteller van de e-mail gebruikmaken van een aangepaste applicatie, zodat deze informatie wordt opgenomen in elk nieuw e-mailbericht dat wordt gemaakt. Bij toepassing van de post-use benadering moet deze informatie tijdens het conversieproces alsnog worden toegevoegd aan het XML-bestand.

Onderstaande afbeelding is een schematische weergave van het XML-schema dat is gebruikt bij de ontwikkeling van de eerder besproken add-in voor Outlook in het kader van de Pre-use benadering. Bij de implementatie van het door Testbed ontwikkelde XML-schema wil een leeg BCC-veld overigens niet zeggen dat er niemand in dat veld is opgenomen. Uit het BCC-veld kan echter uitsluitend inhoud worden bewaard bij uitgaande berichten. Bij inkomende berichten staan er nooit gegevens in het BCC-veld. Dit komt doordat de BCC-inhoud wordt verwijderd wanneer het bericht de eerste keer door een server gaat. Vandaar de benaming: Blind Carbon Copy. Door de opzet van het BCC-veld is het niet mogelijk deze informatie bij inkomende berichten vast te leggen.



Afbeelding 12 Overzicht van een XML-bestand voor e-mail. Zie Appendix C voor het XML schema bij dit overzicht.

Body (b)

De body of berichttekst moet zo worden opgeslagen dat de inhoud intact blijft. Indien gebruik wordt gemaakt van de 'pre-use benadering', waarin een XML-versie van het e-mailbericht wordt gemaakt en een HTML-versie naar de ontvangers wordt verzonden, kan de inhoud van het bericht onmiddellijk worden 'getagged' in een XML-bestand. In dat geval moet een verwijzing worden opgenomen naar de stylesheet waarmee de HTML-versie van het bericht is gemaakt toen het werd verzonden.

Als het bericht vanuit het transmissiebestand wordt geconverteerd naar XML (post-use), moet een apart bestand worden gemaakt om het tekstgedeelte op te slaan. Als het een bericht in onbewerkte tekstindeling betreft, kan het eenvoudigweg worden opgeslagen als een 'plain text' bestand. Betreft het een meer complex HTML-document, dan moet de HTML-weergave van de berichttekst worden opgeslagen. Eventuele bijbehorende 'multipart/related' bestanden kunnen samen met de berichttekst worden opgeslagen, met duidelijke instructies voor het opnieuw samenstellen van het digitale document.

In het XML-bestand moeten de volgende verwijzingen naar elk van deze bestanden voorkomen:

- o Bestandsnaam: de naam van het bestand zoals deze in het transmissiebestand voorkomt, bijvoorbeeld 'body.txt'.
- o Bestandstype: dit beschrijft het bestandstype waarin de inhoud is opgeslagen, bijvoorbeeld HTML. Hiervoor kan een MIME-aanduiding worden gebruikt, bijvoorbeeld 'text/html' of 'app/msword'.
- o Bestandslocatie: deze bevat een unieke ID waarin wordt aangegeven waar het bestand zich bevindt. Deze locatie kan bijvoorbeeld de Windows mappen- en bestandsnaamsysteem, een database-ID of een URL zijn. De exacte ID is afhankelijk van de gekozen implementatie.

Op de hierboven beschreven manier kan de software de onderdelen van de body vrij eenvoudig vinden en deze opnieuw samenstellen.

Bijlagen (c)

Bijlagen moeten apart worden opgeslagen. Organisaties die gebruik maken van de pre-use benadering, moeten deze aangepaste e-mailapplicatie zodanig vormgeven dat een kopie van de bijlage bij het opgeslagen XML-bestand kan worden gevoegd. Bij berichten die vanuit het transmissiebestand worden geconverteerd (post-use), moeten bijlagen vanuit hun gecodeerde verzendvorm worden gedecodeerd en opgeslagen in hun 'native' formaat (Excel, MS Word, WordPerfect, enzovoort). Bijlagen worden in het XML-bestand weergegeven door aparte metadata en een verwijzing. Elke bijlage wordt gedefinieerd door de tag *<Bijlage>* en heeft de volgende kenmerken:

- o Bestandsnaam, bijvoorbeeld 'pic23004.JPG';
- o Bestandstype, bijvoorbeeld 'Image/JPEG';
- o Bestandslocatie: deze geeft de nieuwe opslaglocatie aan van het gekoppelde bestand, bijvoorbeeld 'C:\xmail\files\loc'.

Verwijzingen naar meerdere bijlagen worden voorafgegaan door de parent tag *<Bijlagelijst>*.

Voor zover bijlagen in hun 'native' formaat worden opgeslagen, kan de records manager voor dergelijke bestanden die maatregelen voor bewaring nemen die nodig zijn voor dat specifieke bestandsformaat. Immers voor elk digitaal documenttype gelden andere bewaareisen. Zo kan worden besloten om bepaalde bijlagen vanuit een leverancierseigen en gesloten formaat, zoals Microsoft Word over te zetten naar Adobe's PDF, een van de formaten die door de Regeling geordende en toegankelijke

staat zijn erkend voor de langetermijnbewaring van digitale archiefbescheiden. De bewaarobjectaanpak maakt het mogelijk bijlagen te benaderen in hun 'native' formaat. Originele bestandsformaten kunnen worden vervangen door andere. Door het nieuwe bestand toe te voegen aan het bewaarobject, en dit bewaarobject bij te werken, ontstaat een nieuwe verwijzing. Hierbij blijven de andere delen van het e-mailbericht ongemoeid. Bij deze methode moet de records manager ook het bewaarlogbestand (e) bijwerken met de nieuwe metadata.

Metadata (d)

Voor het duurzaam en authentiek bewaren van digitale documenten is een aanvullend metadatabestand vereist. De exacte invulling van dit metadatabestand wordt overgelaten aan de organisaties zelf. Veel instellingen registreren al metadata, of dat nu gebeurt in een Records Management Applicatie (RMA) of een Document Management Systeem (DMS).

Het is in dit verband wel van belang dat organisaties alle header informatie uit het transmissiebestand halen die niet in het XML-bestand zijn 'getagged' en deze als aanvullende contextuele metadata onderbrengt in het metadatabestand. Op deze items moet kunnen worden gezocht binnen de grenzen van een RMA of DMS. Op deze manier worden alle transmissiegegevens bewaard en doorzoekbaar, inclusief verwijzingen naar berichtdiscussies, 'flag' markeringsgegevens (die de ontvanger mogelijk niet heeft gezien) en informatie over de applicatie. Het XML-bestand bevat overigens een aanzienlijke hoeveelheid metadata over het oorspronkelijke gebruik en de oorspronkelijke context van het digitale document. Deze metadata kunnen eenvoudig worden verwerkt en in een bekende metadata infrastructuur worden opgenomen, bijvoorbeeld EAD (Encoded Archival Description).

Bewaarlogbestand (e)

Het bewaarlogbestand bevat alle informatie over de bewaaractiviteiten die bij het digitale document zijn ondernomen. Bovendien kan in het logbestand informatie over bewaar- en toegangseisen worden opgenomen.

Het logbestand dient twee doelen. Ten eerste kan ermee worden nagegaan of het digitale document na verloop van tijd nog authentiek en toegankelijk is. Ten tweede is het mogelijk dat een component van het e-mailbericht (vaak zal dit een bijlage zijn) eenvoudig en continu kan worden beheerd, zonder dat de authenticiteit van het digitale document in gevaar komt.

Het bewaarlogbestand wordt gemaakt als het digitale document voor het eerst naar XML wordt geconverteerd. We schrijven hier geen formaat voor het logbestand voor, maar wijzen er alleen op dat het een formaat moet zijn waarmee het bewaarlogbestand gemakkelijk en continu kan worden bijgewerkt zonder dat reeds aanwezige gegevens worden overschreven. Een database kan hier geschikt voor zijn en ook het gebruik van XML zelf kan worden overwogen. De eerste inhoud in het logbestand moet bestaan uit de gegevens van het oorspronkelijke digitale document, zoals het was toen het werd ontvangen om te worden omgezet naar XML. Deze inhoud moet worden gevolgd door de gegevens van de conversie, waaronder de gebruikte conversietool, de datum en het tijdstip waarop de conversie plaatsvond en het nieuwe bestandsformaat.

Het bewaarlogbestand moet telkens worden bijgewerkt als er bewaaractiviteiten ten behoeve van het digitale document hebben plaatsgevonden. Het bewaarlogbestand moet bovendien informatie bevatten over eventuele wijzigingen die door bewaaractiviteiten in het digitale document zijn veroorzaakt. In Appendix B wordt de mogelijke inhoud van het logbestand besproken.

Transmissiebestand (f)

Het transmissiebestand is de meest complete en betrouwbare bron van het digitale document. Als het conversieproces mislukt of als om de een of andere reden twijfel bestaat over de authenticiteit van het omgezette digitale document, dan is het zaak om de omgezette versie te vergelijken met de oorspronkelijke. Hoewel dit misschien niet mogelijk is bij alle aspecten van het digitale document (de bijlagen bijvoorbeeld) moet dat wel kunnen bij de headers en mogelijk ook bij de berichttekst.

Wellicht kan het transmissiebestand in de toekomst worden gebruikt bij een nieuwe bewaarstrategie. Het mag worden verwacht dat er andere en betere strategieën komen voor het bewaren van e-mail, daarbij gebruikmakend van de nieuwe mogelijkheden die de technologie ons dan biedt. Het bewaren van een moederkopie van het bestand in het oorspronkelijke formaat kan daarom voordelen hebben die nu nog niet kunnen worden overzien of voorzien.

Bewaarstrategieën hebben niet het eeuwige leven en moeten na verloop van tijd worden beoordeeld en eventueel aangepast. De technologie staat immers niet stil. Op een bepaald moment zal de door Testbed aanbevolen bewaarstrategie voor e-mailberichten mogelijk niet meer de meeste geschikte zijn. Wanneer dat moment komt, is onzeker. Tot dat moment is de geschetste XML strategie voor e-mailberichten de beste strategie om ze duurzaam en in authentieke vorm te bewaren voor de lange termijn.

6 *Concrete acties*

De voorgaande hoofdstukken hebben het probleem van de veroudering van digitale documenten behandeld en de beste strategie voor het bewaren van e-mail voorgesteld. Nu is het zaak dat organisaties met deze informatie aan de slag gaan. Hoofdstuk 5 behandelde al de implementatie van de XML-strategie. De verschillende acties die de betrokkenen binnen een organisatie moeten ondernemen om dit succesvol in gang te zetten, zijn zo specifiek en verschillend van elkaar, dat dit een doelgroepsgewijze aanpak rechtvaardigt. Zo kan iedere medewerker snel zien welke acties hij of zij in gang moet zetten. De onderscheiden doelgroepen zijn:

- o algemene (lijn-)managers;
- o records managers;
- o ICT-ers en
- o eindgebruikers.

Iedere paragraaf is zo geschreven dat deze los van de complete publicatie gelezen kan worden.

6.1 Actieplan voor managers

Inleiding

In de publicatie *Van digitale vluchtigheid naar digitaal houvast. Het bewaren van e-mail* heeft u kunnen lezen wat de voordelen zijn van digitaal werken, maar ook welke specifieke problemen optreden bij het duurzaam bewaren van digitale documenten in het algemeen en van e-mailberichten in het bijzonder. Testbed Digitale Bewaring heeft bewaarstrategieën getoetst en afgezet tegen het documenttype 'e-mail'. De beste manier om e-mail te bewaren is op dit moment door XML te gebruiken. In de publicatie is eveneens uitvoerig besproken hoe de voorgestelde toepassing van XML kan worden geïmplementeerd.

Maar daarmee zijn we er nog niet. Binnen een organisatie zijn verschillende mensen betrokken bij het duurzaam bewaren van e-mailberichten: van de lijnmanagers in een organisatie tot en met de eindgebruikers die de beschikking hebben over de e-mail faciliteiten. De onderstaande concrete acties zijn specifiek gericht op:

- o algemene (lijn-)managers;
- o records managers;
- o ICT-ers en
- o eindgebruikers.

De genoemde vier actoren hebben een specifieke verantwoordelijkheid in deze. In dit laatste hoofdstuk wordt per doelgroep uiteengezet welke concrete stappen zij moeten zetten om het duurzaam bewaren van e-mailberichten tot een succes te maken. Aan de concrete stappen of acties gaat een beschrijving van de randvoorwaarden vooraf.

Randvoorwaarden

"U bent de aanjager van verbeteringen binnen uw organisatie. U heeft goede contacten met de werkvloer. U bent aanspreekbaar voor uw medewerkers. U bent bereid geld en tijd te steken in een communicatiemiddel, e-mail, dat de prestaties van uw organisatie verbetert". Het klinkt als een wervende folder voor een managementcursus. Toch zijn het *essentiële uitgangspunten* om e-mail, net als andere communicatiemiddelen, een verankerde plaats binnen uw organisatie te geven en de vruchten ervan te oogsten: toegankelijke, snel beschikbare en betrouwbare informatie.

Bewustwording bij alle medewerkers van uw organisatie dat e-mail een formeel document is, met alle consequenties van dien, is een voorwaarde om succesvol met dit medium te kunnen werken.

Het is verder belangrijk *snel actie* te ondernemen. De voorbeelden van gevallen waarbij het niet goed bewaren van elektronische post oorzaak was van grote problemen nemen in aantallen toe, omdat het e-mail gebruik de laatste jaren vervelend is.

Concrete acties voor managers

Formuleer e-mail beleid: hoe gaat uw organisatie om met e-mail? Hoe past het binnen het informatie en archiefbeleid binnen uw organisatie? Bij e-mail beleid gaat het erom naar dezelfde kwaliteitsprocedures voor e-mail te streven als voor papieren documenten (zie ook de NEN-ISO standaard 15489).

Een digitale overheid heeft veel voordelen bij het gebruik van e-mail in het werkproces. Ofschoon het medium dezelfde status heeft als papier, zijn veel mensen zich daarvan niet bewust. Veel organisaties gebruiken een disclaimer, omdat zij twijfelen aan de juridische status, maar in de actuele discussie neemt men steeds

vaker het standpunt in dat ook e-mailberichten juridisch bindend zijn. Testbed ontraadt daarom het gebruik van disclaimers: ook een digitale overheid moet betrouwbaar communiceren en e-mail is daar een onderdeel van, mits op de juiste manier gebruikt.

Ontwerp procedures voor en met uw medewerkers: hierin moet duidelijk staan wie waarvoor verantwoordelijk is, wie waarop kan worden aangesproken en welke mensen (functies) elkaar moeten informeren. In ieder geval moeten aan de orde komen:

- o Afspraken over het gebruik van e-mail (voor welke communicatie inzetten).
- o Afspraken over het beheren en bewaren van e-mail.

Gesprekspartners in deze zijn: de records managers binnen uw organisatie, ICT-managers en officemanagers.

Introduceer e-mailtemplates voor uw organisatie: Met behulp van e-mailtemplates in combinatie met een webservice worden alle relevante gegevens correct vastgelegd en kunnen zij vervolgens centraal worden opgeslagen in XML. Daarnaast kunnen e-mailtemplates uw e-mail een formeel karakter geven, bijvoorbeeld met het logo van uw organisatie. Zo is digitale informatie gemakkelijk en goedkoop te bewaren, ook voor de langere termijn.

E-mailtemplates zijn gemakkelijk te integreren in bestaande applicaties, zoals Outlook. Voor de uitvoering hiervan kan bijvoorbeeld een werkgroep met een records manager, ICT-er en communicatiemedewerker worden gevormd. Deze publicatie biedt reeds een (algemene) handleiding hoe e-mailtemplates, en de centrale opslag in XML gemaakt kunnen worden. Uiteraard is het ook mogelijk hiervoor ook een bedrijf in te schakelen.

Informeer alle medewerkers over e-mail beleid en procedures. Train alle medewerkers in het correcte gebruik van de e-mailtemplates.

Evalueer van tijd tot tijd het beleid en de procedures.

6.2 Actieplan voor records managers

In de publicatie *Van digitale vluchtigheid naar digitaal houvast. Het bewaren van e-mail* heeft u kunnen lezen wat de voordelen zijn van digitaal werken, maar ook welke specifieke problemen optreden bij het duurzaam bewaren van digitale documenten in het algemeen en van e-mailberichten in het bijzonder. Testbed Digitale Bewaring heeft bewaarstrategieën getoetst en afgezet tegen het documenttype 'e-mail'. De beste manier om e-mail te bewaren is op dit moment door XML te gebruiken. In de publicatie is eveneens uitvoerig besproken hoe de voorgestelde toepassing van XML kan worden geïmplementeerd.

Maar daarmee zijn we er nog niet. Binnen een organisatie zijn verschillende mensen betrokken bij het duurzaam bewaren van e-mailberichten: van de lijnmanagers in een organisatie tot en met de eindgebruikers die de beschikking hebben over de e-mail faciliteiten. De onderstaande concrete acties zijn specifiek gericht op:

- o algemene (lijn-)managers;
- o records managers;
- o ICT-ers en
- o eindgebruikers.

De genoemde vier actoren hebben een specifieke verantwoordelijkheid in deze. In dit laatste hoofdstuk wordt per doelgroep uiteengezet welke concrete stappen zij moeten zetten om het duurzaam bewaren van e-mailberichten tot een succes te maken. Aan de concrete stappen of acties gaat een beschrijving van de randvoorwaarden vooraf.

Randvoorwaarden

Als records manager bent u zich bewust van de variatie van problemen die moet worden opgelost voordat het beheren van e-mailberichten op dezelfde manier verloopt als het beheren van papieren documenten. De meeste mensen behandelen hun e-mail vooral als een snel en vluchtig medium; zij ervaren het meer als een telefoongesprek, dan als een officieel document. Wat moet u doen om uw collega's ervan te overtuigen dat ook e-mailberichten digitale archiefstukken zijn? Hoe kunt u hen overtuigen om hun e-mailberichten te ordenen en de belangrijke e-mailberichten te classificeren? En 'last but not least': hoe kunt u het management ervan overtuigen om gelden en middelen beschikbaar te stellen om het duurzaam bewaren en beheren van e-mailberichten mogelijk te maken? Dit is niet iets wat u als eenling kunt bewerkstelligen binnen de organisatie. Als records manager is het belangrijk om samenwerking te zoeken met het lijnmanagement, met de ICT-afdeling, en met de eindgebruikers.

Concrete acties

De beste manier om e-mail duurzaam te bewaren hangt af van een aantal factoren, zoals de organisatiecultuur, eisen die de omgeving aan de organisatie stelt, de politieke context, de stand van de technologie en de manier waarop de archieffunctie is vormgegeven. De concrete stappen die gezet moeten worden, betreffen:

- a. Analyse van de huidige situatie;
- b. Formuleren van het gewenste beleid en
- c. Opstellen van procedures.

Onderstaande beschrijving is gebaseerd op de uitgave *Archivering van Elektronische Post*.²¹

²¹ *Archivering van elektronische post. Methoden, meningen en alternatieven*, P. Horsman, Amsterdam 1999.

a. Analyse van de huidige situatie

Zorg ervoor dat het bewaren van e-mail prioriteit krijgt

Procedures hebben alleen kans van slagen indien ze zijn gebaseerd op expliciet uitgedragen beleid. Duidelijk moet zijn wat de organisatie wil met haar e-mail, welk belang zij eraan hecht en welke visie er is op de ontwikkelingen. Dit is vooral een zaak van het lijnmanagement, maar u als records manager moet daarin de rol van katalysator en aanjager vervullen. Voor u is het met name van belang om het bewaren van e-mail op de agenda te krijgen.

Haal de vereiste kennis en kunde in huis

Hoe duidelijk is het vigerende archiefbeleid over het bewaren van e-mailberichten? Voor de bepaling en uitvoering ervan is uw afdeling van belang. Denk eraan dat het duurzaam bewaren van digitale documenten andere kennis en vaardigheden vereist dan het bewaren van papieren documenten. Zorg ervoor dat u die kennis in huis haalt!

Zoek partners en belanghebbenden

Het maken van beleid is niet in de eerste plaats uw verantwoordelijk, maar u kunt wel een belangrijke rol spelen in het proces om de zaak op de agenda te krijgen. Daarbij is het van belang ook andere belanghebbenden op te sporen. Denk daarbij aan afdelingshoofden die bepaalde informatie nodig hebben voor hun bedrijfsvoering, de ICT-afdeling en het belang van alle e-mail gebruikers.

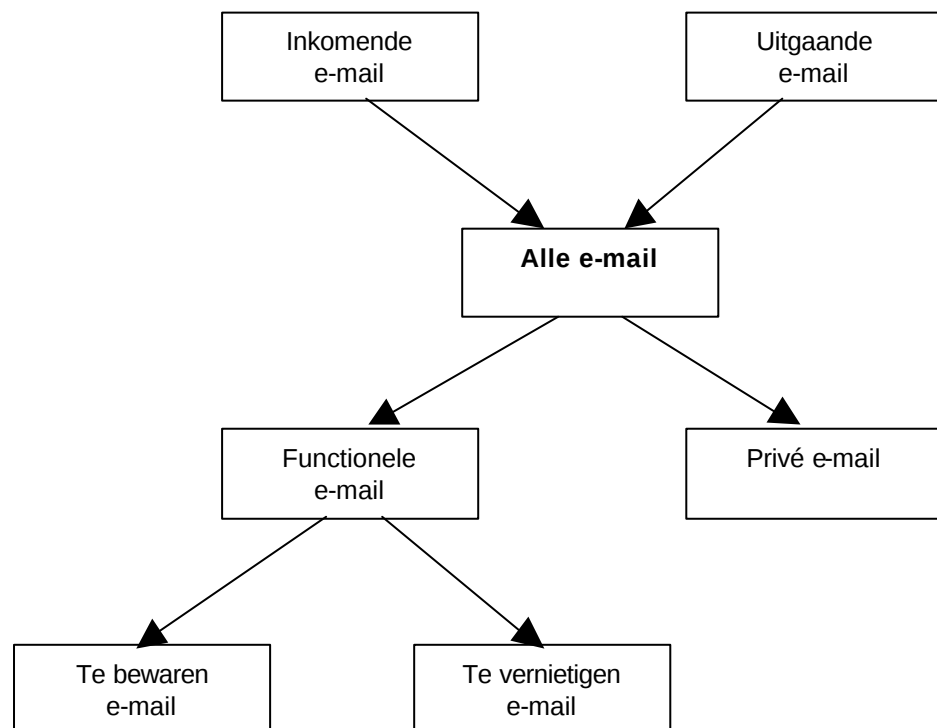
b. Formuleren van het gewenste beleid

Stel de selectiecriteria vast

Zelfs als uw organisatie zich er volledig van bewust is dat ook e-mailberichten bewaard moeten worden, blijft de vraag welke e-mailberichten voor bewaring in aanmerking komen. Zonder deze kwestie hier volledig uit te diepen, kan de volgende indeling²² behulpzaam zijn bij het beantwoorden van de vraag welke e-mail bewaard moet worden.

²²

Richtsnoer E-mailgebruik t.b.v. de Rijksoverheid /Ministerie van BZK; werkgroep E-mailgebruik, Den Haag 2001, p. 4.



Nog sterker dan bij papieren documenten is selectie direct aan het begin de meest geschikte strategie voor digitale documenten zoals e-mailberichten. Dit maakt het namelijk mogelijk om de te bewaren e-mail **direct aan de bron** om te zetten naar een formaat dat duurzamer is dan het leverancierseigen formaat waarin e-mailberichten normaliter worden opgeslagen door de gebruikte e-mailapplicatie.

Formuleer de selectiecriteria voor het bewaren van e-mail. Deze zijn in de regel al vastgesteld in een document structuurplan, of Basis Selectie Documenten (BSD's). Het gaat erom dat die criteria worden toegepast op e-mail. Dring erop aan dat deze selectie inderdaad aan de bron plaatsvindt.

De beslissing van het al dan niet bewaren van een e-mail ligt bij degene die het e-mailbericht verzendt of ontvangt. Als u erop wilt vertrouwen dat gebruikers de juiste e-mailberichten identificeren en bewaren, dan is het aan u om hen ervan bewust te maken dat formele e-mailberichten ook archiefstukken zijn en dat deze met de nodige zorgvuldigheid behandeld dienen te worden.

Behoudt de authenticiteit van e-mailberichten

De keuze voor de juiste wijze van opslag van e-mailberichten is van wezenlijk belang, aangezien het de authenticiteit ervan raakt. Het is duidelijk dat het afdrucken op papier afbreuk doet aan de authenticiteit van e-mailberichten. Veel contextuele informatie gaat dan verloren. Dit is dan ook een optie die vanuit uw achtergrond resoluut moet worden afgewezen.

In hoofdstuk 4 en 5 van deze publicatie kunt u lezen dat Testbed aanbeveelt om XML te gebruiken als opslagformaat. Gebruik die informatie om u, samen met andere disciplines in uw organisatie, sterk te maken voor deze oplossing.

Bepaal welke metadata noodzakelijk zijn

Van elk e-mailbericht is een aantal gegevens van belang om herkomst, bestemming, datum van ontvangst en verzending vast te stellen. Deze metagegevens zijn noodzakelijk om de authenticiteit en de functie van het document te bepalen. Bepaald moet worden welke metagegevens moeten worden geregistreerd²³.

Zorg ervoor dat in deze fase nauwkeurig wordt vastgelegd welke metadata van belang zijn om de informatie te kunnen (her) gebruiken en te interpreteren en die daarnaast nodig is voor uw organisatie om verantwoording af te kunnen leggen.

Bepaal de wijze van ordenen en klasseren

Het doel van ordening en vervolgens klasseren is het zichtbaar maken van de structuur, de samenhang tussen documenten en tussen documenten en processen waarin ze een rol speelden. Daarmee wordt de toegankelijkheid bevorderd en kan gestructureerd zoeken worden ondersteund. Dat betekent dat er een classificatieschema ontwikkeld moet worden dat gebaseerd is op taken of bedrijfsprocessen (zie ook NEN-ISO 15489). Betrek hierbij de ICT-afdeling voor het bepalen van zoekingen en samenhang tussen documenten. E-mailberichten moeten op dezelfde manier worden geklasseerd als papieren documenten.

Formuleer het beleid

Het doorlopen van de voorgaande stappen en de keuzen die daarin gemaakt zijn, moeten worden vastgelegd in een beleidsnotitie. Stel bij elke keuze vast wat haalbaar is en wat ideaal is. Deze notitie is de basis voor het vervolgtraject, die vooral is gericht op implementatie en waarin de feitelijke procedure geschreven moet worden.

c. Opstellen van procedures

Geef aan wie de feitelijke beslissing neemt om een e-mailbericht te bewaren

Voor papieren documenten is het doorgaans de DIV medewerker die beslist wat bewaard moet worden. Bij e-mail zal het doorgaans de eindgebruiker zijn die de beslissing neemt om een e-mailbericht al dan niet op te slaan, waarbij niet alleen het e-mailbericht moet worden opgeslagen, maar ook de bijbehorende contextuele metagegevens.

Beschrijf de wijze van klasseren en dossiervorming

Door toepassing van een classificatieschema (zoals hierboven beschreven) wordt een e-mailbericht toegekend aan een dossier. Indien het classificatieschema is gebaseerd op taken of activiteiten kan zo ook de relatie met het werkproces worden gelegd en vindt klassering plaats. In de door Testbed ontwikkelde e-mailtemplates wordt de eindgebruiker bij het verzenden en ontvangen van e-mail verplicht aan te geven tot welk dossier en werkproces een e-mailbericht behoort. Er zijn natuurlijk ook andere manieren om deze koppeling te leggen. Gezien het karakter van e-mail verdient het aanbeveling de koppeling decentraal te laten leggen door de individuele eindgebruiker op het moment dat e-mailberichten worden gecreëerd of ontvangen. Bedenk wel dat het handmatig toevoegen van te veel metagegevens aan e-mailberichten de acceptatie van de procedure door de eindgebruikers ondermijnt.

Regel de toegankelijkheid tot de opgeslagen e-mailberichten

De mogelijkheden van toegankelijkheid hangen nauw samen met de keuze van opslagformaat en de kwaliteit van de metadata. Indien de e-mailberichten op een centrale server worden opgeslagen, zoals in de e-mail demo van Testbed, kunnen de

²³

Zie voor de bepaling van metadata de eerder genoemde Regeling ex art. 12, of *Een uitdijend heelal? Context van archiefbescheiden*, H. Hofman, Stichting Archiefpublicaties, Jaarboek 2000.

e-mailberichten in principe algemeen toegankelijk worden gesteld. De vraag is of de organisatie dat wil; dit is aan het management om dit te bepalen. Met het organisatiebeleid als startpunt, kent het management autorisaties toe, eventueel gedelegeerd aan uw afdeling. De feitelijke uitvoering ligt vanzelfsprekend bij de ICT-afdeling.

Zorg ervoor dat e-mail in XML wordt bewaard

E-mailberichten die voor lange termijn bewaring in aanmerking komen moeten worden omgezet van het leverancierseigen formaat naar XML. Dit is een open standaard die zeer geschikt is voor het duurzaam bewaren van e-mailberichten. Een complicerende factor zijn de bijlagen of attachments, die vooralsnog in hun oorspronkelijke formaat moeten worden bewaard. Het hangt af van het specifieke bestandsformaat welke bewaarstrategie hier van toepassing is.

Zorg dat het beleid regelmatig wordt getoetst

De informatietechnologie verandert snel en dat geldt ook voor organisaties. De eisen die aan digitale archivering worden gesteld zijn eveneens in ontwikkeling. Dit betekent dat het beleid regelmatig getoetst en of aangepast moet worden.

Verwacht mag worden dat in de toekomst betere software beschikbaar komt voor het beheer van digitale documenten. Testbed pleit er daarom voor het oorspronkelijke databestand, in dit geval het transmissiebestand, eveneens te bewaren.

De komende jaren zijn een overgangperiode waarin uw afdeling zich kan voorbereiden op een nieuwe invulling van haar taak. De implementatie van de bewaring van e-mail biedt u daartoe een goed houvast.

6.3 Actieplan voor ICT-ers

Inleiding

In de publicatie *Van digitale vluchtigheid naar digitaal houvast. Het bewaren van e-mail* heeft u kunnen lezen wat de voordelen zijn van digitaal werken, maar ook welke specifieke problemen optreden bij het duurzaam bewaren van digitale documenten in het algemeen en van e-mailberichten in het bijzonder. Testbed Digitale Bewaring heeft bewaarstrategieën getoetst en afgezet tegen het documenttype 'e-mail'. De beste manier om e-mail te bewaren is op dit moment door XML te gebruiken. In de publicatie is eveneens uitvoerig besproken hoe de voorgestelde toepassing van XML kan worden geïmplementeerd.

Maar daarmee zijn we er nog niet. Binnen een organisatie zijn verschillende mensen betrokken bij het duurzaam bewaren van e-mailberichten: van de lijnmanagers in een organisatie tot en met de eindgebruikers die de beschikking hebben over de e-mail faciliteiten. De onderstaande concrete acties zijn specifiek gericht op:

- o algemene (lijn-)managers;
- o records managers;
- o ICT-ers en
- o eindgebruikers.

De genoemde vier actoren hebben een specifieke verantwoordelijkheid in deze. In dit laatste hoofdstuk wordt per doelgroep uiteengezet welke concrete stappen zij moeten zetten om het duurzaam bewaren van e-mailberichten tot een succes te maken. Aan de concrete stappen of acties gaat een beschrijving van de randvoorwaarden vooraf.

Randvoorwaarden

Het e-mail gebruik is de laatste jaren exponentieel toegenomen. Veel organisaties ontwikkelen beleid hoe om te gaan met e-mail. Tevens realiseert men zich meer en meer dat ook e-mailberichten bewaard moeten worden, bijvoorbeeld voor de bedrijfsvoering of om verantwoording te kunnen afleggen. Om dit in goede banen te leiden, moeten verschillende personen binnen de organisatie actie ondernemen. Uitgangspunten zijn dat er e-mailbeleid is geformuleerd, dat de records manager procedures heeft geformuleerd voor de selectie van e-mailberichten die in aanmerking komen voor (blijvende) bewaring. Daarnaast zijn de eindgebruikers getraind in het goed omgaan met de in uw organisatie gebruikte e-mailapplicatie.

De ICT-afdeling is onontbeerlijk om het op een goede manier bewaren van e-mailberichten mogelijk te maken. Hieronder worden de technische ICT kwesties beschreven die aan de orde zijn bij de implementatie van een e-mailbewaarstrategie. Het is echter niet mogelijk exact aan te geven hoe de voorgestelde bewaarstrategie moet worden geïmplementeerd. Deze is namelijk afhankelijk van de bestaande computeromgeving en de specifieke eisen van de betreffende organisatie, en die zijn per geval verschillend. We gaan echter wel in op de belangrijkste eisen en geven daarbij een overzicht van een mogelijke systeemarchitectuur.

Concrete acties

De concrete acties die u moet ondernemen betreffen:

- a. Algemene principes;
- b. Aanbevelingen inzake formaat en implementatiemogelijkheden en
- c. Praktische kwesties.

a. Algemene principes

Sla de te bewaren e-mailberichten op in een centraal beheerd systeem en niet op de computers of in de persoonlijke mappen van de afzonderlijke gebruikers. Hierdoor kan worden voorkomen dat e-mailberichten bedoeld of onbedoeld worden verwijderd

(bijvoorbeeld als de gebruiker de melding krijgt dat zijn of haar persoonlijke postbus vol is); de toegang tot de centraal bewaarde e-mailberichten kan worden gecontroleerd, zowel om de informatie beschikbaar te stellen voor diegenen die deze nodig hebben als om ongeoorloofde toegang te voorkomen. Via een centraal systeem kunnen ook de opslagmedia – meestal een combinatie van schijven en tapes – worden gecontroleerd en beheerd. Ook het maken van kopieën en back-ups valt hieronder. Bedenk wel dat er binnen de context van digitale duurzaamheid een wereld van verschil is tussen het bewaren van back-ups en het duurzaam bewaren van digitale documenten, waaronder e-mailberichten.

Leg metadata zoveel mogelijk automatisch vast

Veel van de metadata van een e-mailbericht bevinden zich in de e-mail headers: bijvoorbeeld van wie het bericht afkomstig is, aan wie het is gericht, het onderwerp en de datum. Zorg ervoor dat deze gegevens automatisch worden opgehaald door het e-mail bewaarsysteem, zodat de gebruiker minder werk heeft en de kans op fouten afneemt. U weet als geen ander dat het belangrijk is dat het bewaarsysteem, net als ieder ander systeem, gebruiksvriendelijk moet zijn, wil het breed geaccepteerd worden door degenen die ermee werken.

Er zijn echter metadata die door de gebruiker moeten worden ingevuld op het moment dat hij of zij aangeeft dat het e-mailbericht voor bewaring in aanmerking komt. Maak het hen zo eenvoudig mogelijk door templates te ontwikkelen met default waarden en drop down menu's waaruit de gebruiker de juiste waarde kan ophalen. Dit verhoogt de uniformiteit van de ingevoerde gegevens en verkleint de kans op fouten. Een belangrijk voorbeeld hiervan zijn de metadata over de ontvangers van een e-mailbericht. Het e-mail adres van iemand kan onvoldoende zijn om die persoon duidelijk te identificeren. Het is belangrijk om behalve het e-mailadres ook de volledige naam en functie en de gegevens over de organisatie op te slaan. Met het oog op een maximale gebruiksvriendelijkheid kan de e-mailapplicatie worden gekoppeld aan een geschikte database (zoals de map Contactpersonen of het adresboek in Outlook). De informatie hoeft dan maar één keer te worden ingevoerd.

Metadata over de classificatie en de context van een e-mail, bijvoorbeeld bij welk dossier het bericht hoort, moeten door het centrale bewaarsysteem worden gebruikt voor het ordenen van de opgeslagen berichten, vooral ter ondersteuning van een zoekfunctie.

Zorg ervoor dat het bewaarsysteem aan elk opgeslagen e-mailbericht een zogenoemd bewaarlogbestand (audit trail-informatie) toevoegt.

Een dergelijke logbestand moet metadata bevatten over de computeromgeving, zoals de versie van de gebruikte e-mailapplicatie, de versie van het gebruikte bewaarsysteem en een overzicht van eventuele bij de e-mail verrichte bewaaractiviteiten, zoals de datum en het tijdstip waarop conversie naar een bewaarformaat plaatsvond en de datum en het tijdstip waarop de e-mail naar het bewaarsysteem werd gestuurd. Zie Appendix C voor meer informatie over de aanbevolen inhoud van het Bewaarlogbestand.

Geef e-mail een uiterlijk in de huisstijl van de organisatie

Hoe gemakkelijk een e-mail op langetermijn kan worden bewaard, hangt in het algemeen sterk af van de manier waarop deze wordt gemaakt. Via een combinatie van richtlijnen en softwaretools kunnen gebruikers worden geholpen bij het maken van formele e-mailberichten. Hierbij kunnen ook tools worden gebruikt om op formele e-mailberichten een algemene huisstijl toe te passen, dat wil zeggen een vastgelegde structuur en uiterlijk wat betreft lettertypen, kleuren en logo's. Dit is bijvoorbeeld mogelijk door het e-mailbericht rechtstreeks in het XML-formaat te maken (een strategie die bij de e-maildemo van Testbed is onderzocht) met een gebruiksvriendelijke interface, zodat de gebruiker niets van XML hoeft te weten, en

vervolgens een centraal systeem een standaard-XSL-opmaakmodel te laten toepassen om het bericht in HTML-vorm te kunnen verzenden.

Sla e-mail niet op in leverancierseigen formaat maar in XML

In hoofdstuk 3 van deze publicatie zijn de authenticiteitseisen beschreven voor het duurzaam bewaren van authentieke e-mailberichten. Het bewaarsysteem moet aan deze eisen voldoen.

Het opslagformaat moet uiteraard alle belangrijke kenmerken van het bericht kunnen weergeven. Zoals beschreven in hoofdstuk 3 kunnen deze als volgt worden ingedeeld: inhoud, context, structuur, uiterlijk en gedrag. Deze kenmerken moeten zodanig worden opgeslagen dat deze zowel nu als in de toekomst zo eenvoudig en correct mogelijk worden weergegeven.

Wat we ten eerste afraden, is het opslaan van e-mailberichten in leverancierseigen formaten, aangezien die specifieke software vereisen voor de data-interpretatie. Hoewel andere documenttypen, zoals tekstdocumenten, moeilijker zonder de bijbehorende software kunnen, zijn e-mailberichten vrij eenvoudig in een 'leveranciers neutraal' opslagformaat te bewaren. Dit komt omdat er een goed gedefinieerde standaard is ontwikkeld door de Internet Engineering Task Force (IETF RFC2822 en verwante extensies, in het bijzonder MIME) voor het formaat waarin e-mailberichten moeten worden verzonden. Het opslaan van berichten in dit RFC2822+MIME formaat zou één optie kunnen zijn voor het duurzaam bewaren van e-mailberichten. Wij bevelen dit formaat dan ook slechts aan als één onderdeel van een bewaarstrategie (zie hoofdstuk 5). Op zichzelf is dit formaat echter niet voldoende en daarom bevelen we een op XML gebaseerde strategie aan, die hieronder wordt beschreven.

Omgaan met bijlagen

Bijlagen in e-mailberichten bemoeilijken het probleem van duurzaam digitaal bewaren in hoge mate. Aangezien in principe elk bestandstype in een e-mailbericht kan worden ingevoegd en samen met dat e-mailbericht kan worden verzonden, moeten in een bewaarsysteem voor e-mail dus veel verschillende ingevoegde bestandstypen kunnen worden bewaard. In de andere delen van deze serie wordt ingegaan op de bewaarstrategie voor die bestandstypen. Hier wordt alleen de manier beschreven waarop het e-mail bewaarsysteem moet worden ingericht opdat zo eenvoudig mogelijk met bijlagen of attachments kan worden omgaan.

b. Aanbevolen formaat en implementatiemogelijkheden

De door Testbed aanbevolen strategie voor het bewaren van e-mailberichten wordt gedetailleerd beschreven in hoofdstuk 5. Hieronder wordt een korte samenvatting gegeven, gevolgd door enkele opmerkingen over de mogelijkheden voor het implementeren van deze strategie.

De belangrijkste onderdelen van een duurzaam bewaard e-mailbericht zijn:

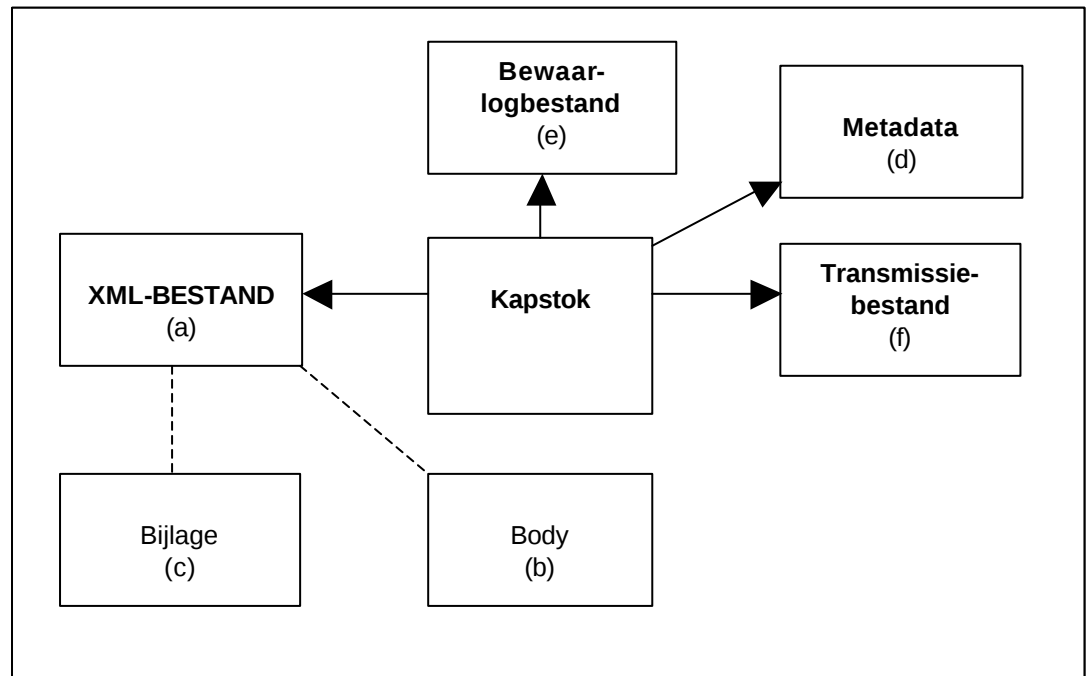
- o de berichttekst (body), indien mogelijk in meer dan één vorm (bijvoorbeeld zowel in onbewerkte tekstindeling als in HTML);
- o bijlagen/attachments; afbeeldingen of andere objecten (vaak aanwezig in e-mailberichten opgemaakt in HTML) die bij het e-mailbericht horen.
- o Metadata over context.

Om e-mailberichten ook op de lange termijn authentiek te kunnen bewaren, is het noodzakelijk dat deze onderdelen in een zogenoemd e-mail bewaarobject aanwezig zijn, samen met de volgende aanvullende componenten:

- o een bewaarlogbestand (in feite een audit trail over de handelingen die bij het e-mailbericht zijn uitgevoerd, en de technische gegevens die nodig zijn voor het bewaren);

- o het oorspronkelijke transmissiebestand en
- o aanvullende metadata.

Testbed beveelt het gebruik van XML aan als kapstok om e-mailberichten te bewaren. Dit bewaarobject bevat de belangrijkste contextuele metadata van het digitale document (bijvoorbeeld informatie over het onderwerp, de afzender en ontvangers en gegevens over datum en tijd) en bevat de structuur van het e-mailbericht. Bovendien bevat het koppelingen naar aparte bestanden zoals de berichttekst (body), bijlagen (attachments) en eventuele verwante objecten. Tevens geeft de kapstok de relatie tussen deze bestanden aan. Voor een uitvoerige beschrijving van de 'XML als kapstok'-strategie verwijzen we naar hoofdstuk 4. De structuur van het recordobject wordt geïllustreerd in onderstaand schema dat in hoofdstuk 5 werd besproken.



Afbeelding 14. Schematische weergave van het e-mail bewaarobject

Het mechanisme voor het koppelen van de onderdelen van het e-mail bewaarobject wordt bepaald door de gekozen implementatiestrategie, al is het van wezenlijk belang dat elk onderdeel een unieke ID heeft. Dit kan onder andere een database ID, een URL of de padnaam van een bestandssysteem zijn. Bij deze laatste ID moet er wel op worden gelet dat er geen naamconflicten optreden. Het XML bestand dat het e-mailbericht representeert, kan via de ID's van de onderdelen aan de aparte bestanden met de berichttekst en bijlagen worden gekoppeld.

De structuur van het XML-bestand moet worden beschreven in een DTD- of XML-schema, zodat afzonderlijke XML-bestanden kunnen worden gevalideerd. Het voordeel van het scheiden van de onderdelen van het bericht (zie afbeelding) is dat de structuur en inhoud van het bericht gemakkelijker bewaard worden. Ook kunnen hierdoor de verschillende bestandstypen in het systeem makkelijker worden bijgehouden (bijvoorbeeld onbewerkte tekst, HTML, MS Word, PDF of JPEG), zodat op elk van deze typen de meest geschikte bewaarstrategie kan worden toegepast.

Deze e-mailbewaarmethode kan op verschillende manieren worden geïmplementeerd. Welke manier geschikt is, hangt af van de bestaande computeromgeving en de eisen van andere systemen. Hieronder volgen drie punten die aandacht verdienen.

- o De e-mailberichten kunnen worden opgeslagen via een RMA (Record Management Applicatie) of DMS (Document Management System). Hoewel dergelijke systemen veel nuttige voorzieningen en een goed ontworpen omgeving voor gecontroleerde informatieopslag bieden, moet toch nog worden nagedacht over de langetermijnopslag van digitale documenten, waarin het systeem nog niet voorziet. Het is cruciaal dat het bewaarobject met al zijn facetten in het RMA wordt opgeslagen
- o De standaard e-mailapplicatie (bijvoorbeeld MS Outlook) moet zodanig worden aangepast dat metadata kunnen worden verzameld en communicatie met een centrale service mogelijk is met het oog op bewaring. Het implementeren van de berichtopslaginterface van het centrale systeem als een webservice, waarbij de communicatie tussen client en server plaatsvindt door middel van SOAP-berichten die via HTTP worden verzonden, is een betrekkelijk eenvoudige oplossing die goed wordt ondersteund door moderne softwaretools zoals .NET en J2EE.
- o Het centrale opslagsysteem kan bestaan uit een bestandssysteem, een database of een combinatie van beide. Welke methode geschikt is, hangt af van de grootte van het systeem, van de vraag of het een lange- of kortetermijnoplossing is en van de eisen voor toegang tot de opgeslagen e-mailberichten. Een methode die veel wordt toegepast, is het opslaan van metadata over het e-mailbericht in een relationele database (zodat de gegevens makkelijk kunnen worden opgezocht) en het maken van koppelingen naar bestanden die op een server zijn opgeslagen. Bij deze methode kan een 'native' XML-database worden gebruikt in plaats van een relationele database.

Een opgeslagen e-mailbericht zoeken, vinden en openen

Een opslagsysteem voor e-mail is pas bruikbaar als de informatie die erin is opgeslagen toegankelijk is. Belangrijk hierbij is het opzoeken van specifieke e-mailberichten. Het centrale systeem moet blader- en zoekfuncties bieden waarmee de gebruiker de informatie kan opzoeken die hij nodig heeft. Zo zou er op metadata moeten kunnen worden gezocht (bijv. op afzender, ontvanger, datum, onderwerp, dossier, werkproces, enz.), maar ook full text in de inhoud van een e-mailbericht. Overleg met de records manager voor de juiste classificatiecodes en benamingen die als zoekingen gebruikt worden. Deze zijn van belang voor de samenhang van de documenten.

Als de e-mail is gevonden, moet de gebruiker deze ook kunnen lezen en interpreteren. Daartoe zou het e-mailbericht kunnen worden omgezet in het formaat dat door de standaard e-mailapplicatie van de organisatie wordt gebruikt of zou bijvoorbeeld een op HTML gebaseerde internetviewer kunnen worden toegepast.

Installatie, onderhoud, training, ondersteuning zijn noodzakelijk

Ervan uitgaande dat de meeste overheidsmedewerkers werken met e-mail, zullen velen ook gebruik maken van het e-mail bewaarsysteem. De installatie van de aangepaste e-mailapplicatie, zoals ontwikkeld door Testbed, zou binnen bestaande systemen kunnen plaatsvinden met het oog op een centraal beheer en een centrale uitrol van applicaties. Het gebruik van een 'thin client'-strategie op basis van webbrowsers is één manier om controleproblemen bij installatie en configuratie tot een minimum te beperken. Het nadeel hiervan is echter dat de functionaliteit van het systeem wordt beperkt en dat de gebruikers moeten overstappen op een nieuwe, minder bekende, e-mailapplicatie.

Aangezien deze strategie nieuwe werkmethoden en aanpassing van de gebruikte e-mailapplicatie vereist, zullen de gebruikers training nodig hebben. Door de software zo gebruiksvriendelijk mogelijk te maken en ervoor te zorgen dat deze zoveel mogelijk op bestaande standaard applicaties lijkt, kan training tot een minimum worden beperkt. Niettemin blijft training belangrijk om de gebruikers doeltreffend te leren omgaan met de software. Tevens zal de ICT afdeling samen met de records manager hen ondersteuning moeten bieden bij wat een aanzienlijke verandering van werkmethoden zal zijn.

c. Praktische kwesties

Bij het ontwerp en de configuratie van het bewaarsysteem moeten de volgende praktische kwesties worden overwogen:

- o Veiligheid: de toegang tot het centrale opslagsysteem moet zorgvuldig worden gecontroleerd om onbedoelde of opzettelijke aantasting van de opgeslagen informatie te voorkomen (het opstellen van een toegangsclassificatieschema, zie ook NEN_ISO 15489);
- o back-up: zoals bij elk belangrijk IT-systeem is een geschikte back-up strategie vereist, zodat het systeem opnieuw kan worden opgezet in geval van een systeemstoring, onbedoelde of opzettelijke beschadiging van het systeem of een ramp zoals brand of overstroming;
- o flexibiliteit: verschillende groepen binnen een organisatie kunnen andere metadata nodig hebben en deze behoeften kunnen na verloop van tijd veranderen, zodat het een voordeel is om dit aspect van het systeemontwerp zo flexibel mogelijk te houden. De records manager moet dit in samenspraak met de gebruiker aangeven;
- o gedistribueerde e-mail opslagsystemen: in een grote organisatie met veel afdelingen is het wellicht praktischer over een aantal kleinere e-mail opslagsystemen te beschikken dan over één heel groot systeem. Aandachtspunt is dan wel dat de samenhang bewaakt en geregeld word, want zoeken 'door alles heen' is een vereiste;
- o responsetijden en betrouwbaarheid: omdat de gebruikers bij hun dagelijkse werkzaamheden het bewaarsysteem vaak nodig zullen hebben, zijn korte responstijden en een optimale betrouwbaarheid vereist. Hierbij zijn twee zaken van belang. Ten eerste moet de gebruiker snel en eenvoudig een e-mailbericht in het bewaarsysteem kunnen opslaan. Ten tweede moet de informatie die reeds in het systeem is opgeslagen gemakkelijk bereikbaar en vindbaar zijn. Hierbij bestaan per bedrijfsproces mogelijk heel andere gebruikspatronen.

6.4 Actieplan voor eindgebruikers

Inleiding

In de publicatie *Van digitale vluchtigheid naar digitaal houvast. Het bewaren van e-mail* heeft u kunnen lezen wat de voordelen zijn van digitaal werken, maar ook welke specifieke problemen optreden bij het duurzaam bewaren van digitale documenten in het algemeen en van e-mailberichten in het bijzonder. Testbed Digitale Bewaring heeft bewaarstrategieën getoetst en afgezet tegen het documenttype 'e-mail'. De beste manier om e-mail te bewaren is op dit moment door XML te gebruiken. In de publicatie is eveneens uitvoerig besproken hoe de voorgestelde toepassing van XML kan worden geïmplementeerd.

Maar daarmee zijn we er nog niet. Binnen een organisatie zijn verschillende mensen betrokken bij het duurzaam bewaren van e-mailberichten: van de lijnmanagers in een organisatie tot en met de eindgebruikers die de beschikking hebben over de e-mail faciliteiten. De onderstaande concrete acties zijn specifiek gericht op:

- o algemene (lijn-)managers;
- o records managers;
- o ICT-ers en
- o eindgebruikers.

De genoemde vier actoren hebben een specifieke verantwoordelijkheid in deze. In dit laatste hoofdstuk wordt per doelgroep uiteengezet welke concrete stappen zij moeten zetten om het duurzaam bewaren van e-mailberichten tot een succes te maken. Aan de concrete stappen of acties gaat een beschrijving van de randvoorwaarden vooraf.

Randvoorwaarden

Uiteindelijk bent u het als gebruiker van e-mail die bepaalt of het mogelijk zal zijn om de aanbevelingen in deze uitgave daadwerkelijk te realiseren. U staat immers aan het begin van de keten, aan de bron, waarmee we bedoelen dat u de e-mailberichten aanmaakt, verzendt, ontvangt en beheert. U bepaalt daarmee in hoge mate of uw organisatie in staat is e-mailberichten duurzaam te bewaren.

Het uitgangspunt is dat uw organisatie e-mail beleid heeft geformuleerd en dat er binnen uw organisaties afspraken en procedures zijn hoe met dit fenomeen om te gaan. Verschillende partijen in deze hebben hierbij een rol, zoals het management, de records managementafdeling, de ICT-afdeling en u als eindgebruiker. In de volgende paragraaf staat beschreven waar u aan moet denken met name bij het creëren van e-mailberichten. Want al er iets naar voren is gekomen in ons onderzoek dan is het wel dat digitale duurzaamheid aan de bron begint, dus bij u.

Concrete acties

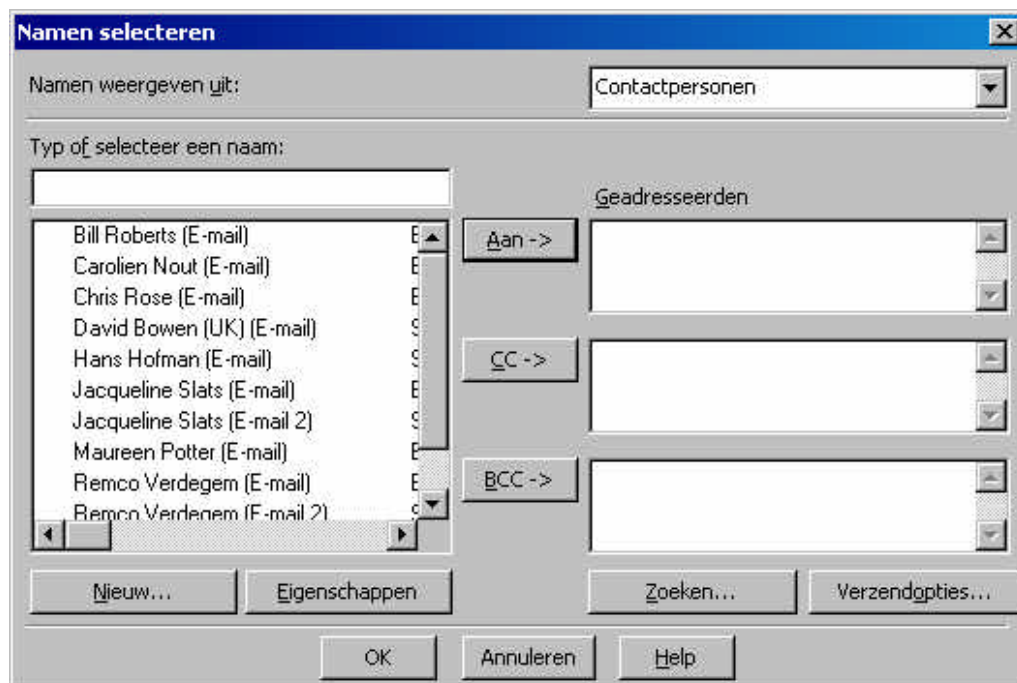
Bij het invullen van de concrete acties onderscheiden we drie aandachtsgebieden:

- a. Het adresseren van e-mailberichten (headerinformatie);
- b. het opstellen van een e-mailbericht (het bericht en eventuele bijlagen of attachments) en
- c. het beheren van e-mailberichten (inkomende en uitgaande).

a. E-mailberichten adresseren

De informatie die in de header van een e-mail wordt getypt (de invulvakken Aan; CC, BCC, Onderwerp), is erg belangrijk, omdat die met name contextuele informatie levert zoals wie zijn de zender en afzender, waar gaat het over, enzovoort. Het is dan ook belangrijk dat u deze velden correct invult. Die informatie moet reeds aan de bron door de maker worden vastgelegd, omdat u bij uitstek degene bent die de context, waarin het e-mailbericht wordt gemaakt, kent.

Maak altijd gebruik van het adresboek van uw e-mailapplicatie. In het adresboek is aanvullende informatie opgeslagen over de mensen naar wie u berichten stuurt. Deze informatie wordt samen met het e-mailbericht opgeslagen, zodat deze contextinformatie altijd aanwezig is. Als u adressen vanuit het adresboek in de header van e-mailberichten invoegt, wordt niet alleen het e-mailadres, maar bijvoorbeeld ook de naam van de ontvanger, zoals opgenomen in het adresboek, automatisch aan het bericht toegevoegd. Dit is van belang met name in die gevallen, waarin mensen gebruik maken van een 'exotische' e-mailadres, bijvoorbeeld ens-p45ht@planet.nl. Dit veronderstelt overigens wel dat het adresboek op een correcte manier wordt gevuld en onderhouden. Het algemene adresboek van een organisatie wordt meestal centraal beheerd. Ook voor uw persoonlijke contactpersonen bevelen wij aan deze zo compleet mogelijk in te vullen en bij te houden.



Afbeelding 14. Voorbeeld van het gebruik van het adresboek

Wees terughoudend in het gebruik van distributielijsten. We hebben reeds aangegeven dat het belangrijk is de namen en adressen (en bij voorkeur de functie) van de geadresseerden te kennen. In distributielijsten wordt deze informatie echter niet altijd (correct) geregistreerd. Afhankelijk van de e-mailapplicatie en het soort e-mail (intern of extern) komen de naam van de lijst en de namen van de personen op de lijst al dan niet in het bericht voor. Het komt voor dat een distributielijst bij een intern e-mailbericht wordt verstuurd, zonder dat de namen van de leden van de distributielijst te zien zijn. Aan de hand van het adresboek kunt u echter zelf vaststellen wie lid zijn van de betreffende distributielijst. Echter een distributielijst is een dynamische lijst. Na verloop van tijd worden nieuwe namen aan de lijst toegevoegd, terwijl oude worden verwijderd. Vaak wordt de status van de distributielijst tijdens het bestaan ervan niet geregistreerd, zodat vrijwel onmogelijk wordt om na te gaan wie op welk moment op de lijst voorkwam.

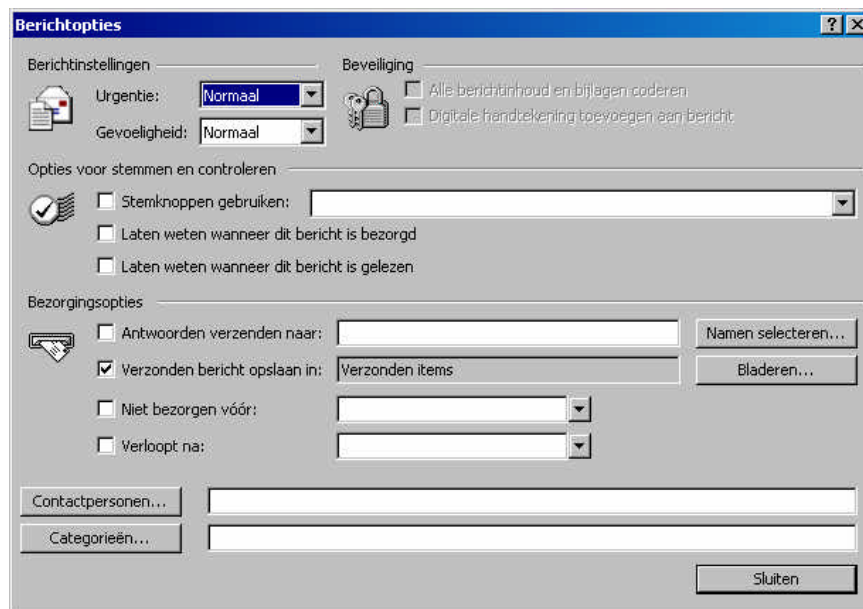
Als u de namen wilt vasthouden van de mensen naar wie u een bericht verzendt en waarvoor u een distributielijst gebruikt, voeg deze informatie dan duidelijk toe aan de inhoud van het e-mailbericht. Of dit nodig is, hangt af van het type bericht dat u maakt. Zo hoeven in een bericht dat via listserv voorzieningen is verzonden niet alle namen van de geadresseerden te worden opgeslagen in de inhoud. Dat zou immers een inbreuk betekenen op de privacy van die personen. Daarentegen moeten in een e-mailbericht dat naar leden van een distributielijst is verzonden om deze te raadplegen over conceptdocumenten wellicht wel de namen van de geadresseerden worden opgeslagen, aangezien het mogelijk moet zijn om te bewijzen dat bepaalde mensen het document hebben kunnen zien en becommentariëren, voordat het werd vrij gegeven. In deze situatie kunt u er beter voor kiezen de e-mailadressen van de geadresseerden op te nemen in het 'Aan' veld en niet gebruik te maken van een distributielijst.

Bij een externe e-mailbericht in combinatie met een distributielijst wordt de distributielijst in de regel door de e-mailapplicatie 'vertaald' naar de afzonderlijke leden van de lijst. De naam van de distributielijst zelf gaat daarbij verloren. Indien u belang hecht aan deze naam (bijvoorbeeld als contextinformatie), dan moet u de naam van de distributielijst opnemen in de body van het e-mailbericht.

Geef uw e-mailberichten altijd een onderwerp. Dat klinkt logisch, maar het is verrassend hoeveel mensen daar niet aan denken. De onderwerpregel moet uniek en informatief zijn. Het gaat hier om belangrijke informatie waarmee u uw berichten kunt sorteren en beoordelen. De onderwerpregel van een e-mail is vaak ook de titel van het e-mailbericht. Zorg daarom dat het onderwerp relevant en bruikbaar is.

Gebruik berichtopties zoals urgentie alleen als dat echt nodig is. Bij veel moderne e-mailapplicaties kan de gebruiker de Urgentie en de Gevoeligheid van het e-mailbericht instellen en markeringen aan het bericht toevoegen die rechtstreeks vanuit de header aanvullende informatie over het bericht verschaffen (bijvoorbeeld Prioriteit).

Dit advies heeft niet zozeer te maken met het bewaren ervan – markeringen worden door de e-mailapplicatie betrouwbaar verzonden en kunnen worden bewaard – als wel met de verschillen tussen e-mailapplicaties. Hoewel markeringen en urgentie-instellingen zinvol kunnen zijn, zijn niet alle e-mailapplicaties uitgerust om ze correct weer te geven. In sommige gevallen krijgt de ontvanger ze helemaal niet te zien en kan eventuele aanvullende informatie of betekenis die via markeringen of instellingen aan het bericht zijn meegegeven, verloren gaan. Als u zeker weet dat de geadresseerden dezelfde e-mailapplicatie gebruiken als u, kunt u markeringen en instellingen gebruiken. Weet u dat echter niet zeker, probeer dan deze extra informatie op een andere manier aan uw e-mailbericht toe te voegen, bijvoorbeeld door de gegevens rechtstreeks in de onderwerpregel of in de body van het e-mailbericht op te nemen.



Abbeelding 15. De berichtopties voor urgentie in Outlook2000.

b. E-mailberichten opstellen

Uit ons onderzoek is gebleken dat e-mailberichten in onbewerkte tekstindeling (plain text) gewoonlijk zonder problemen kunnen worden verzonden en ontvangen. Hoe eenvoudiger het e-mailbericht, des te groter de kans dat de geadresseerde (ontvanger) precies datgene ziet wat u als afzender heeft bedoeld. Naarmate u de e-mail complexer maakt, neemt die kans echter af. De moraal van dit verhaal? Houd uw e-mail zo eenvoudig mogelijk.

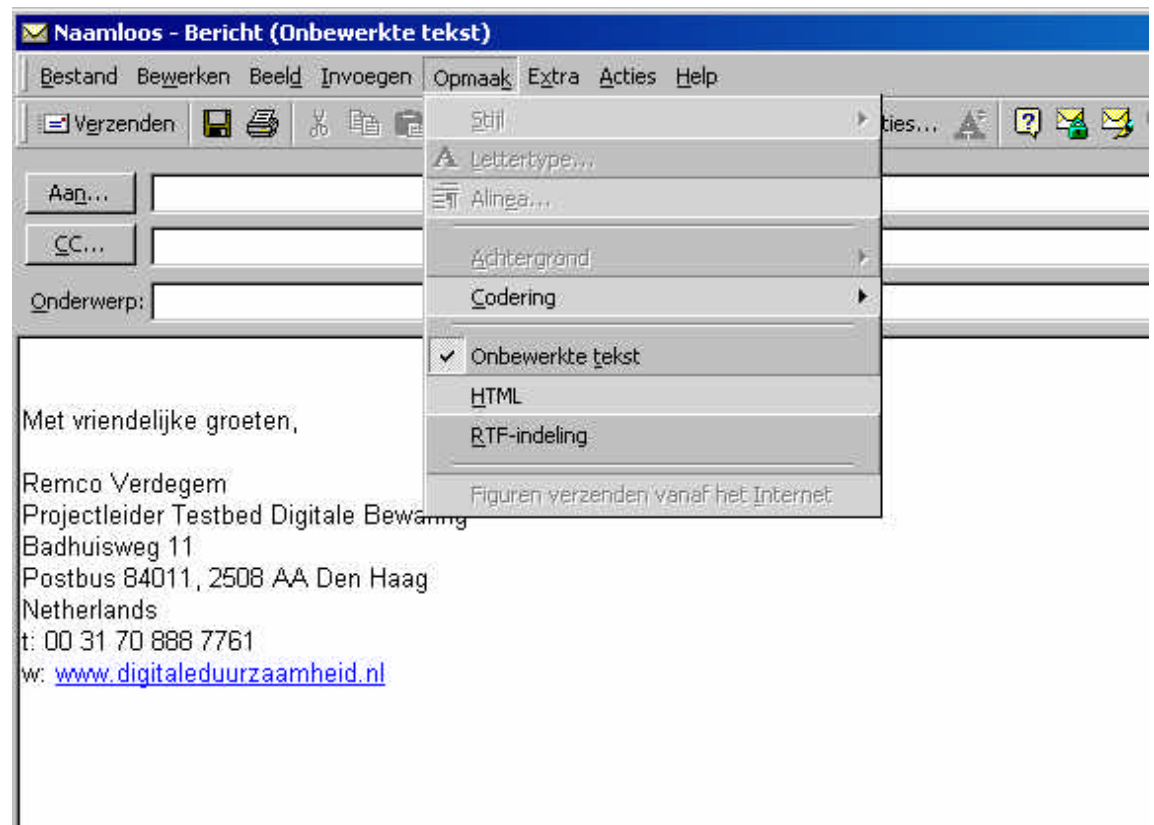
Maak en verzend uw berichten waar mogelijk in onbewerkte tekstindeling of in HTML-formaat²⁴. Berichten in onbewerkte tekstindeling zijn meestal prima geschikt voor eenvoudige communicatie. Meer formele en officiële e-mailberichten waarin ook afbeeldingen en logo's kunnen zijn opgenomen, kunnen in HTML worden gemaakt. Houd er echter rekening mee dat naarmate e-mailberichten complexer zijn, het lastiger is de e-mailberichten duurzaam te bewaren. Bovendien ondersteunen oude e-mailapplicaties geen HTML.

Maak bij voorkeur dus geen gebruik van het Rich Text Format (RTF) van MS Outlook, omdat dit formaat specifiek is voor Outlook en bij verzending wordt gecodeerd in een bestand dat mogelijk ook bijlagen en opmaakinformatie bevat. Deze gegevens moeten worden vertaald, zodat ze door andere e-mailapplicaties kunnen worden gelezen. MS

²⁴

Via de instellingen van uw e-mailapplicatie kunt u het formaat bepalen waarin uw berichten worden verzonden. Outlook 2002 biedt de opties Onbewerkte tekst, Rich Text, en HTML. Veel overheidsorganisaties werken met Outlook en hebben dit programma daarom als voorbeeld genomen. Andere systemen, zoals Novell Groupwise en Eudora, bieden andere opmaakformaten. In zulke gevallen moeten gebruikers hun keuzen onderzoeken en handelen in overeenstemming met het bovenstaande algemene advies.

Exchange, dat vaak samen met MS Outlook wordt gebruikt, zorgt voor deze vertaling. Wat daarna wordt verzonden, is afhankelijk van de instellingen van MS Exchange.



Abbeelding 16. De verschillende opmaak mogelijkheden in Outlook.

Gebruik in uw e-mailberichten geen velden die automatisch worden bijgewerkt, zoals automatische datum- en tijdvakken. Voer deze gegevens altijd in als 'harde' tekst. Automatische velden zijn niet stabiel en worden telkens bijgewerkt als u het e-mailbericht opent, waardoor de inhoud en context van het e-mailbericht direct verloren gaat.

Gebruik uw bijlagen of attachments verstandig. Zorg ervoor dat u uw bestanden in het juiste bestandsformaat verzendt. Verzend een afbeelding bijvoorbeeld als bitmap of JPEG-bestand. Maar al te vaak worden afbeeldingen in een applicatie, zoals PowerPoint of Word geplakt. Deze applicaties maken bestanden aan in het leverancierseigen formaat, terwijl vrijwel elke viewer bitmap- en JPEG-bestanden kan weergeven.

Bij beantwoorden van e-mail niet tussenvoegen. Gewoonlijk wordt op een ontvangen e-mailbericht gereageerd door op <Beantwoorden> te drukken. Daar is niets mis mee, zolang het maar goed gebeurt. Als u op de inhoud van een bericht wilt reageren, type dan uw eigen opmerkingen boven het oorspronkelijke bericht en laat ruimte tussen uw handtekening en de headers van het oorspronkelijke bericht. Zo kunnen uw opmerkingen direct worden onderscheiden van die van de ander. Vergeet niet dat uw bericht mogelijk over twintig, dertig of veertig jaar wordt gelezen en u er dan zelf niet bij bent voor tekst en uitleg.

Deze werkwijze is nog belangrijker als u in één en hetzelfde e-mailbericht met meerdere personen correspondeert. Als vijf personen achtereenvolgens reageren op dit e-mailbericht en elke persoon zijn opmerkingen in uw e-mailbericht voegt, is het al snel niet meer duidelijk wie wat heeft geschreven.

Gebruik een handtekeningenblok. Door een 'handtekeningenblok' aan het eind van uw e-mailbericht toe te voegen, verschaft u de ontvanger belangrijke contextuele informatie. Let op: een handtekeningenblok is geen digitale handtekening. De term 'digitale handtekening' wordt veelal verkeerd gebruikt, maar heeft tegenwoordig betrekking op de techniek, waarmee een bepaald type codering op bestanden kan worden toegepast voor verzending via het internet.

Handtekeningenblokken zijn blokken informatie die aan het eind van een e-mail worden toegevoegd en waarmee de identiteit van de afzender en enkele contactgegevens worden meegedeeld. Het is een toevoeging waarmee toekomstige gebruikers van het digitale document aanvullende contextuele informatie verkrijgen.

Het is zinvol met twee handtekeningenblokken te werken: één voor intern gebruik onder de leden van uw team, afdeling, project of organisatie en één voor extern gebruik die aan alle uitgaande officiële correspondentie wordt toegevoegd. Een interne handtekeningenblok moet minimaal die gegevens bevatten die nodig zijn voor het identificeren van een medewerker. Dit zijn:

- o de naam van de afzender (bijvoorbeeld Jacqueline Slats);
- o de officiële functie van de afzender (bijvoorbeeld Programma manager);
- o het project of de afdeling (bijvoorbeeld Testbed Digitale Bewaring);
- o de organisatie (bijvoorbeeld Stichting ICTU).

Externe handtekeningenblokken moeten zoveel informatie bevatten dat ontvangers de afzender kunnen achterhalen zonder op de knop Beantwoorden te klikken:

- o de naam van de afzender (bijvoorbeeld Jacqueline Slats);
- o de officiële functie van de afzender (bijvoorbeeld Programma manager);
- o het project of de afdeling (bijvoorbeeld Testbed Digitale Bewaring);
- o de organisatie (bijvoorbeeld Stichting ICTU);
- o het adres en de vestigingplaats (bijvoorbeeld Nieuwe Duinweg 24 – 26, Den Haag);
- o het telefoonnummer (bijvoorbeeld +31(0)70 888 77 69);
- o het e-mailadres van de afzender (bijvoorbeeld Jacqueline.Slats@ictu.nl);
- o de website van het project- of de afdeling (bijvoorbeeld <http://www.digiteduurzaamheid.nl>).

c. Inkomende en uitgaande e-mailberichten beheren

U bent in eerste instantie zelf verantwoordelijk voor het beheer van inkomende en uitgaande e-mailberichten. Dit betekent dat als u de ontvanger bent van een formeel e-mailbericht van een externe contactpersoon, dat voor bewaring in aanmerking komt, u dat bericht moet bewaren in de daarvoor bestemde map. Bent u de afzender van een te bewaren e-mailbericht, dan bent u daar ook verantwoordelijk voor.

Interne (administratieve) berichten binnen de organisatie behoeven alleen door de afzender te worden beheerd en bewaard. Ze hoeven dus niet te worden bewaard door iedereen die een kopie heeft ontvangen. Niet-administratieve e-mailberichten tussen verschillende afdelingen binnen dezelfde organisatie kunnen betrekking hebben op verschillende werkprocessen en moeten daarom door zowel verzender als ontvanger worden bewaard.

Zorg voor een goed beheer van het Postvak IN. Verwijder inkomende tijdelijke berichten onmiddellijk, nadat u ze hebt gelezen. Zo houdt u het Postvak IN werkbaar.

Verwijder ook de tijdelijke berichten uit het venster Verzonden items en maak regelmatig de map Verwijderde items leeg.

Tijdelijk goed bewaren. Als er nog geen aangepast systeem is geïmplementeerd voor het bewaren van uw e-mailberichten, moet u de e-mail op een andere manier opslaan, totdat een dergelijk systeem is geïnstalleerd. Een eenvoudige manier is het aanmaken van mappen voor die e-mailberichten die bewaard moeten worden, zodat niet alle e-mailberichten in één grote map worden onderbracht, hetgeen het terugzoeken ernstig bemoeilijkt. Maak de mappen aan in overleg met de records manager in uw organisatie. Uw e-mailberichten worden in dat geval tijdelijk in hun 'oorspronkelijke' formaat opgeslagen totdat het mogelijke is deze om te zetten naar XML. Hetzelfde geldt voor e-mailberichten die u deelt met anderen, zoals een postbus op afdelingsniveau. Zorg er wel voor dat inkomende en uitgaande e-mailberichten die bij elkaar horen in dezelfde map terechtkomen om de samenhang te bewaren.

Zoals gezegd is dit niet meer dan een tijdelijke oplossing tot een aangepast e-mailsysteem is geïnstalleerd, die het mogelijk maakt uw e-mailberichten direct in XML te genereren en op een centrale locatie op te slaan. Bijlagen moeten in het oorspronkelijke formaat samen met het bericht worden opgeslagen. Aangezien bijlagen vaak zijn opgemaakt in een leverancierseigen formaat zoals Microsoft Word, Excel, PowerPoint of Adobe's PDF, vergen ze mogelijk wel een andere benadering voor langdurige bewaring dan het e-mailbericht zelf. Door ze apart op te slaan, wordt de verzendcodering ervan onmiddellijk gedecodeerd, kan het betreffende documenttype worden gecontroleerd op naderende veroudering en kan de geëigende bewaaractie worden ondernomen, bijvoorbeeld de migratie naar een hogere versie.

Plak in geen geval de inhoud van een e-mailbericht in een andere applicatie zoals Microsoft Word om vervolgens het oorspronkelijke bericht te verwijderen. Zowel de authenticiteit als de integriteit van het bericht worden dan namelijk ernstig aangetast. Zodra u deze handeling uitvoert, gaat een groot aantal belangrijke metadata verloren die niet meer kunnen worden teruggehaald uit het nieuwe formaat. Uiteindelijk kunnen de in XML opgemaakte e-mailberichten worden overgeheveld naar een Document Management Systeem (DMS) of Record Management Applicatie.

Verklarende woordenlijst

Achterwaartse compatibiliteit

Hiermee wordt bedoeld dat software in staat is bestanden te decoderen of in te lezen die gemaakt zijn met eerdere versies van dezelfde software. De meeste software is overigens slechts beperkt achterwaarts compatibel.

ASCII

American Standard Code for Information Interchange.

Algemeen geaccepteerde standaard waarin de betekenis van tekens (bytes) is vastgelegd. Computers begrijpen immers alleen getallen. Zo staat ASCII-code 65 voor de hoofdletter A. ASCII is tevens een protocol voor het overbrengen van bestanden van de ene computer naar de andere. Het is eigenlijk alleen geschikt voor tekstbestanden.

Attachments

Binaire of platte tekst bijlagen die worden meegestuurd met e-mailberichten.

Authenticiteit

De mate waarin de weergave van een archiefbescheid volledig en volstrekt in overeenstemming is met de oorspronkelijke vastlegging van het bescheid en bovendien de functie die bij ontstaan bedoeld was, behouden blijft.

BCC

Blind Carbon Copy.

De optie in e-mailprogramma's om een e-mailbericht, zonder melding aan de ontvanger daarvan in kennis te stellen, in afschrift te sturen naar een ander/anderen. Zie ook CC

Bewaring

Processen en activiteiten die betrekking hebben op de zorg voor de technische en intellectuele instandhouding van authentieke archiefbescheiden door de tijd heen.

CC

Carbon Copy

In dit geval kan de ontvanger in de header zien wie hetzelfde e-mailbericht in afschrift hebben ontvangen.

Computerbestand

Een geheel van gegevens in een zelfde opslagformaat

Context

Het geheel van administratief-organisatorische, bestuurlijk-juridische en technische gegevens, waarbinnen de functie van het archiefstuk in relatie tot de activiteiten en taken van de archiefvormer moet worden geïnterpreteerd.

Conversie

Het omzetten in of overzetten van gegevens in een ander opslagformaat

Digitale duurzaamheid

Het resultaat van de waarborging van de raadpleegbaarheid, authenticiteit en leesbaarheid van digitale documenten gedurende de geldende bewaartermijn.

DIV

Documentaire Informatie Voorziening. Het proces van communicatie door middel van documenten; dit begrip impliceert dus papieren en digitale documenten, zowel teksten als financiële, meet- en regelgegevens en afbeeldingen.

DMS

Document Management Systeem

Een systeem dat functionaliteit biedt voor het acquireren, opslaan, archiveren en opvragen van documenten, inclusief het managen tijdens het uitvoeren, administreren, doorgeven en autoriseren van gebruikers. Document Management Systemen bewaken de toegang tot bestanden en houden een historisch bestand bij van de inhoud van de bestanden.

E-mail

'Electronic mail', oftewel elektronische post

Emulatie

Het softwarematig nabootsen van de oude hardware- en/of softwareomgeving. De emulatiesoftware draait op huidige en toekomstige hardwareplatforms, zodat het probleem van veroudering van hard- en software wordt vermeden.

Gedrag

Gedrag is één van de vijf attributen van digitale documenten, zoals beschreven door Jeff Rothenberg en Tora Bikson in "Carrying Authentic, Understandable and Usable Digitale Records Through Time". Gedrag maakt het de gebruiker mogelijk te interacteren met het digitale document, bijvoorbeeld door het openen van een attachment of door het activeren van een hyperlink. De andere vier attributen zijn inhoud, context, structuur en opmaak.

Integriteit

Eigenschap van een digitaal document dat zijn vorm, inhoud en structuur bij raadpleging gelijk zijn aan de vorm, inhoud en structuur op het tijdstip dat het werd opgemaakt.

HTML

Hyper Text Mark-up Language

J2EE

Staat voor Java 2 Enterprise Edition. Is een ontwikkel platform dat de laatste jaren is uitgegroeid tot een industrie standaard voor het ontwikkelen van grootschalige Java-applicaties.

JPEG

Staat voor Joint Pictures Expert Group en is met name een bestandsformaat voor foto's op websites. In JPEG wordt het beeld opgedeeld in blokken. Van elk blok wordt alleen de meest relevante informatie opgeslagen

Mark-Up language

Een ander woord voor meta talen, speciaal bedoeld voor het aanbrengen van structuur in complexe documenten. De bekendste varianten zijn HTML en XML

Metadata

Gegevens die context, inhoud, vorm en structuur van digitale documenten en hun beheer door de tijd heen beschrijven.

Migratie

Het overzetten van bestanden van de ene hardware- en/of softwareomgeving naar de andere

MIME

Een Internet-protocol dat het coderen van binaire inhoud in e-mailberichten mogelijk maakt. Zo kan MIME worden gebruikt om een grafisch bestand of Worddocument te coderen en als bijlage toe te voegen aan een op tekst (ASCII) gebaseerd bericht. De ontvanger moet ook met MIME werken om de bijlage te kunnen decoderen.

Opslag

Het structureel bewaren van digitale informatie, zoals bestanden en documenten op magnetische of optische media.

PDF

Portable Document Format. Een door de firma Adobe ontwikkeld bestandsformaat voor het uitwisselen van documenten met behoud van kwaliteit en vormgeving.

PKI

Public Key Infrastructure. Een systeem van digitale certificaten, certificate autoriteiten en andere registratie autoriteiten die voor een elektronische transactie de validiteit van iedere partij kunnen verifiëren.

Platform

Geheel van apparatuur en besturingsprogrammatuur waarop de toepassingsprogrammatuur werkt.

RMA

Records Management Applicatie. Toepassingsprogrammatuur voor het opnemen, beheren en beschikbaar stellen van archiefbescheiden.

RTF

Rich Text Format. Formaat van een tekstdocument met opmaak. Een Microsoft-protocol voor bestandsindeling dat vet, markeringen, onderstrepingen en vele andere opmaakkenmerken omvat.

Structuur

Het logische verband tussen de elementen van een digitaal document of van een archief.

Toegankelijkheid

Mate waarin de authentieke weergave van een (digitaal) document zonder belemmering kan worden geraadpleegd.

Transmissiebestand

Het e-mailbestand zoals dat door een mailserver wordt ontvangen of verstuurd.

URL

Uniform Resource Locator (adres). Een internet naamgevingsconventie voor resources die beschikbaar zijn via diverse TCP/IP-toepassingsprotocollen. Bijvoorbeeld: [HTTP://www.digitaleduurzaamheid.nl](http://www.digitaleduurzaamheid.nl) is de URL van de website van het programma Digitale Duurzaamheid.

Viewer

Softwareprogramma om bepaalde bestanden te kunnen bekijken.

Vorm

De uiterlijke verschijning waarin de structuur en opmaak zichtbaar zijn.

W3C

World Wide Web Consortium ontwikkelt standaarden voor het World Wide Web (WWW), momenteel de belangrijkste applicatie op het internet. Een belangrijk werkgebied van W3C betreft mark-up languages voor het definiëren en structureren van webdocumenten. Zie ook www.w3c.org

Wrapper

Een begrip dat staat voor een benadering waarbij XML wordt gebruikt als een soort envelop, een omhulsel.

XML

Staat voor eXtended Markup Language en is een tekstgebaseerde taal om gegevens te verrijken met informatie over structuur en betekenis. Het is een open standaard, gedefinieerd door het World Wide Web Consortium en is hard- en softwareonafhankelijk.

XSLT

Extensible Stylesheet Language Transformations: een tool om XML documenten om te zetten, bijvoorbeeld naar HTML. Zie ook: www.w3c.org/Style/XSL/

Bibliografie

Boudrez, Philip	<XML/> en Digital Archiveren (2002) http://www.antwerpen.be/david/teksten/xml_digitaalarchiveren.pdf
Crocker, David H	Standard for the Format of ARMA Internet Text messages – RFC # 822 http://www.ietf.org/rfc/rfc0822.txt
Giesbers, Saskia (RMC)	Records Management Terminologie (6 maart 2002)
Horsman, Peter	Archivering van Elektronische Post; methoden, meningen en alternatieven (Programma Digitale Duurzaamheid, Den Haag 1999)
Feeney, Mary (Ed)	Digital Culture: Maximising the Nation's Investment (National Preservation Office UK, 1999)
Hovy, Lodewijk	Sporen nalaten of uitwissen? Het bewaren van persoonsgegevens / (Archievenblad, december 2001)
InterPARES Project	Authenticity Task Force Final Report (2002) http://www.interpares.org/book/interpares_book_d_part1.pdf
InterPARES Project	Preservation Task Force Final Report (2002) http://www.interpares.org/book/interpares_book_f_part3.pdf
Lorie, Raymond	<i>A Project on the Preservation of Digital Data</i> http://www.rlg.org/preserv/diginews/diginews5-3.html
Lourens, Wim, et al	<i>Emulation and Conversion: Organisational and Architectural Overview of an electronic Archive</i> http://www.library.tudelft.nl/e-archive/Documenten/Resultaten/reportone13.pdf
Mellor, Paul et al	<i>Migration On Request, a Practical Technique for Preservation</i> (2002) http://www.si.umich.edu/CAMILEON/reports/migreq.pdf
Natural Resources Canada	Guidelines on Managing Electronic Mail Messages (January 2000)
Ploeg, Dr. F. van der	Regeling geordende en toegankelijke staat archiefbescheiden
Prins, prof. mr. J.E.J. Matthijssen, dr. L.J.	De digitale overheid en de wet; juridische kaders voor gebruik van digitale documenten bij overheden (Programma Digitale Duurzaamheid, Den Haag, 2000)
Programma Digitale Duurzaamheid	<i>De Digitale Overheid en de wet</i> http://www.digitaleduurzaamheid.nl/bibliotheek/docs/dig-overh-wet.pdf
Programma Digitale Duurzaamheid	<i>Archivering van Elektronische Post</i> http://www.digitaleduurzaamheid.nl/bibliotheek/docs/archelp.pdf

Programma Digitale Duurzaamheid	<i>Naar een verantwoorde archivering van e-mail (Programma Digitale Duurzaamheid, The Hague, 1998)</i>
Redactie	E-mail dertig jaar oud (Archievenblad, December 2001)
Rothenberg, Jeff & Bikson, Tora	Carrying Authentic, Understandable and Usable Records Through Time(1999) http://www.digitaleduurzaamheid.nl/bibliotheek/docs/final-report_4.pdf
Rijksarchiefinspectie	Wet- en regelgeving www.rijksarchiefinspectie.nl/wetgeving/
Testbed Digitale Bewaring	XML en digitale bewaring (2002) http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white-paper_xml-nl.pdf
Thomas, Wimpe	XML: de mogelijkheden en valkuilen voor de overheid; 19 september 2002.
VERS	Victorian Electronic Records Strategy Final Report http://www.prov.vic.gov.au/vers/published/final.htm
Werkgroep E-mailgebruik	Richtsnoer E-mailgebruik t.b.v. de Rijksoverheid (Ministerie van Binnenlandse Zaken en Koninkrijkrelaties, Den Haag, 2001)
Zuurmond, A, Mies, K.	Winst met ICT in uitvoering; Zenc, Den Haag, juni 2002.

Appendix A

E-mail/XML demo: technische beschrijving

Introductie

Testbed Digitale Bewaring heeft een demo ontwikkeld die laat zien op welke wijze het mogelijk is om e-mailberichten duurzaam op te slaan in XML, zonder de gebruiksvriendelijkheid van het medium e-mail aan te tasten.

Doelen

Er zijn drie aspecten van het e-mailgebruik die hierbij in ogenschouw zijn genomen:

- E-mail wordt steeds meer ingezet voor officiële communicatie, terwijl men geneigd is dit communicatiemiddel op een informele manier te gebruiken. Ons doel was de inhoud en de vorm van officiële e-mailberichten meer te laten lijken op die van een brief, geschreven op officieel briefpapier (met briefhoofd of logo).
- Als een e-mail bewaard moet blijven, doen de meeste organisaties dit door het bericht uit te printen en te bewaren in een papieren archief. Sommige medewerkers bewaren ad hoc e-mailberichten in hun persoonlijke e-mail brievenbus. Testbed zocht een manier om e-mailberichten op een centrale plek te bewaren, zonder deze in een ingewikkeld document management systeem onder te hoeven brengen.
- De e-mailberichten moesten zodanig worden opgeslagen dat ze eenvoudig en correct voor de lange termijn bewaard kunnen worden, zodat de toegankelijkheid en leesbaarheid gegarandeerd is voor zolang als het nodig is, dus dat kan 5, 10 of zelfs 100 jaar zijn.

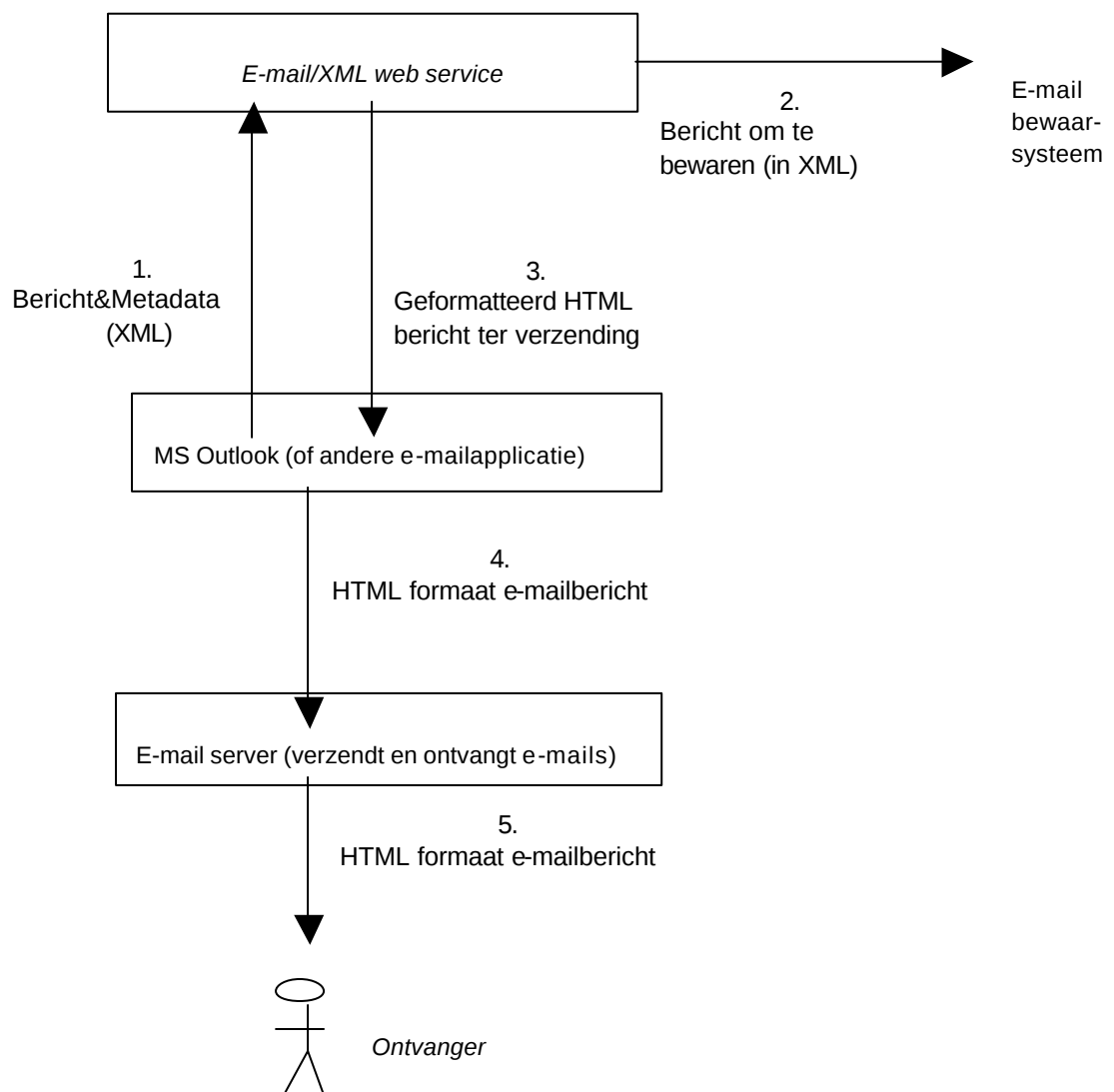
Er is gezocht naar een oplossing die gemakkelijk in het gebruik is, weinig kosten met zich meebrengt en eenvoudig is te implementeren.

Onze oplossing

Uitgangspunt bij het ontwikkelen van de demo is Microsoft Outlook geweest. De reden om hiervoor te kiezen ligt in het feit dat dit het meest gebruikte e-mailprogramma is binnen de Nederlandse overheid. De te ontwikkelen oplossing zou goed inpasbaar moeten zijn in de huidige kantoorautomatiseringomgeving van de gebruikers, waarbij we tevens gebruik zouden kunnen maken van de mogelijkheden die Outlook zelf al biedt, zodat wij die zelf niet zouden hoeven ontwerpen. Hoewel gekozen is voor Outlook, betekent dit geenszins dat onze oplossing niet op andere e-mailprogramma's kan werken, in tegendeel. Onze keuze voor Outlook betekent evenmin dat wij hiermee suggereren dat dit programma beter of slechter is dan andere e-mailprogramma's.

Onze benadering houdt een aanpassing van Outlook in, waarbij een ActiveX DLL is geschreven in Visual Basic. Wij hebben een aantal mogelijkheden aan de standaard

versie van Outlook toegevoegd en tegelijk een aantal reguliere Outlook functies ontoegankelijk gemaakt. Het (overall) systeem volgt op een client-server architectuur. De client draait op de computer van de gebruiker en communiceert met de centrale server, waar de berichten worden opgeslagen en die zorgt voor het omzetten in de huisstijl van de uitgaande e-mailberichten. Dit is de 'webservice server', te onderscheiden van de reguliere e-mailserver, die in alle gevallen steeds nodig blijft.



Afbeelding A1. Omzetting naar XML met behulp van webservice.

Metadata

Het standaard formaat voor e-mailberichten (zoals gedefinieerd door de Internet Engineering Taskforce) bevat veel nuttige informatie, maar voor de lange termijn bewaring is het noodzakelijk informatie toe te voegen. In onze demo hebben wij de volgende metadata elementen toegevoegd:

- Context informatie zoals dossier en werkproces;
- Informatie over de afzender van het e-mailbericht, zoals naam, functie, adres, telefoon, etc.;
- Informatie over de ontvanger, zoals volledige naam en organisatie.

Voordat een formele e-mail verstuurd kan worden, moet een deel van deze informatie eerst worden ingevuld in een invulformulier. Dit vraagt slechts een kleine inspanning van de e-mail gebruiker: de persoonlijke informatie hoeft slechts eenmalig ingevuld te worden; deze gegevens worden dan bij volgende berichten automatisch door de computer ingevuld. De informatie over het dossier en werkproces kan bijvoorbeeld geselecteerd worden uit een keuzelijst.

Het e-mailadres alleen is vaak niet voldoende om een persoon te kunnen identificeren als het de bedoeling is de context van een bericht voor langere tijd te bewaren. Om die reden maakt onze demo gebruik van een adreslijst in Outlook Contacts, waar aanvullende informatie over de ontvanger kan worden ingevuld. Voordat men iemand een formele e-mail stuurt, moeten de volledige naam en organisatiegegevens van de ontvanger in de Contact map worden ingevuld. Deze informatie wordt er door de software uitgehaald en opgeslagen in de metadata van het bericht.

Opslag in XML formaat

In onze demo worden zowel de inhoud van het bericht als de metadata opgeslagen in een XML document. Hiervoor zijn twee redenen: de eerste is dat XML een goed gedefinieerde open standaard is en dat het een goed formaat is voor langdurige bewaring. De tweede reden is dat het gebruik van XML voor de inhoud van het bericht het mogelijk maakt om eXtensible Style sheet Language (XSL) te gebruiken om een omzetting van XML naar HTML te bewerkstelligen. Hiermee worden de inhoud van een bericht en de metadata verzameld, in een geformatteerde vorm weergegeven, waarbij de overall lay out van het bericht, de keuze van het lettertype, kleuren en een logo of afbeelding worden toegevoegd. Aangezien dit centraal gebeurt, kan de huisstijl worden toegepast in alle formele e-mailberichten. Een verandering van de (huis-)stijl hoeft dan slechts op één centrale plaats te worden aangebracht.

De Outlook user interface is aangepast om de gebruiker te begeleiden bij het maken van zijn e-mailbericht. Het e-mailbericht wordt achter de schermen geconverteerd naar XML. De Outlook add-in combineert de inhoud van het bericht met de extra metadata in een enkel XML document. Dit document wordt overgezet naar de webservice server, ingekapseld in een SOAP-bericht. Deze server slaat een kopie van dit XML document op in het e-mail bewaarsysteem. Vervolgens wordt XSL stylesheet gebruikt om het geformatteerde HTML bericht te maken, dat teruggestuurd wordt naar Outlook, als een SOAP bericht.

Een andere functie van webservice is te controleren of de XML die Outlook heeft gemaakt, tegemoet komt aan het XML-schema dat is gedefinieerd voor het bericht. Aangezien we controleren of het XML bericht voldoet aan het schema, kunnen we er zeker van zijn dat de XSL transformatie goed is gelukt.

(Binaire) bijlagen kunnen gewoon aan een e-mailbericht worden toegevoegd., Deze worden overgezet van Outlook naar de webservice server. Iedere bijlage wordt apart in het bewaarsysteem opgeslagen. Een link naar dit document wordt samen met de belangrijkste metadata opgeslagen in het opgeslagen XML document. Nu is het zo dat in principe elk soort document als bijlage in een e-mail kan worden meegestuurd. Of deze ook op de lange termijn goed bewaard kunnen blijven is een moeilijk probleem, dat niet direct met onze demo voor het bewaren van e-mail opgelost kan worden. Maar onze benadering inzake het bewaren van e-mail in de vorm van een XML document met metadata over de bijlagen, inclusief informatie over het type bijlage, maakt het gemakkelijk om de bijlagen te beheren en terug te vinden in het bewaarsysteem en het is wellicht een eerste stap naar een verdere oplossing.

Ontvangen e-mailberichten

De software voorziet tevens in de mogelijkheid om e-mailberichten vanuit de Outlook Inbox (of andere persoonlijke folders) duurzaam op te slaan als XML document. De gebruiker moet de noodzakelijke metadata invullen en vervolgens wordt het e-mailbericht omgezet naar XML. Daarbij wordt hetzelfde XML schema (zie appendix C) gehanteerd als voor uitgaande e-mailberichten. Het enige verschil is de manier waarop wordt omgegaan met de feitelijke inhoud van het e-mailbericht. Aangezien de ontvanger geen controle heeft over het formaat en de opmaak van inkomende e-mail, moet het XML schema in staat zijn om te gaan met ieder soort bericht dat valt binnen de algemene standaarden voor e-mail. Om die reden wordt de inhoud van het e-mailbericht opgeslagen in een apart bestand (meestal in de vorm van een HTML of platte tekst bestand). Een link naar dit bestand is opgeslagen in de XML representatie van het e-mailbericht.

Appendix B

XML Schema

XML schema

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Dit is een schema dat als afdwingbaar sjabloon de structuur van een emailbericht vastlegt.-->
<!--Gemaakt door Hette Bakker voor Testbed, ICTU, nov 2002. -->
<xs:schema targetNamespace="xmail.ictu.nl" xmlns="xmail.ictu.nl" xmlns:xs="http://www.3.org/2001/XMLSchema" elementFormDefault="qualified">
  <!--Het element e-mail is de 'wortel' van het schema, dit element bestaat weer uit de elementen meta, transmissiebestand, onderwerp, etc. -->
  <xs:element name="e-mail">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="meta" type="metaType"/>
        <xs:element name="transmissiebestand" type="bestandsgegevensType"/>
        <xs:element name="onderwerp"/>
        <xs:element name="geadresseerdenlijst">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="geadresseerde" type="geadresseerdeType" maxOccurs="unbounded"/>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
        <xs:element name="cclijst" minOccurs="0">
          <xs:complexType>
            <xs:sequence>
```

```

        <xs:element name="cc" type="geadresseerdeType" maxOccurs="unbounded"/>
    </xs:sequence>
</xs:complexType>
</xs:element>
<xs:element name="bcclijst" minOccurs="0">
    <xs:complexType>
        <xs:sequence>
            <xs:element name="bcc" type="geadresseerdeType" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="afzender" type="afzenderType"/>
<xs:element name="inhoud">
    <xs:complexType>
        <xs:sequence>
            <xs:choice>
                <xs:element name="nietXMLinhoud" type="nietXMLinhoudType"/>
                <xs:element name="XMLinhoud" type="XMLinhoudType"/>
            </xs:choice>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="bijlagelijst" type="bijlagelijstType" minOccurs="0"/>
</xs:sequence>
</xs:complexType>

```

```

</xs:element>
<!-- Einde van het element e-mail. -->
<!-- Hieronder volgen de verschillende 'types' waarnaar hierboven wordt verwezen. -->
<xs:complexType name="metaType">
  <xs:sequence>
    <xs:element name="datum" type="xs:dateTime"/>
    <xs:element name="dossier">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="Experiment 49"/>
          <xs:enumeration value="Experiment 53"/>
          <xs:enumeration value="Jansen 12"/>
          <xs:enumeration value="Pietersen 19"/>
          <xs:enumeration value="Overzichten adviesorganen van de centrale overheid"/>
          <xs:enumeration value="E-mailXML Demo"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="werkproces">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="Het uitvoeren van migratie-experimenten met spreadsheets"/>
          <xs:enumeration value="Het verstrekken van bijstandsuitkeringen"/>
          <xs:enumeration value="Het houden van enquête t.b.v. het verzamelen van informatie over adviesorganen"/>
          <xs:enumeration value="Het verstrekken van advies over het langdurig bewaren van e-mails"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
  </xs:sequence>
</xs:complexType>

```

```

        </xs:restriction>
    </xs:simpleType>
</xs:element>
</xs:sequence>
</xs:complexType>
<xs:complexType name="geadresseerdeType">
    <xs:sequence>
        <xs:element name="naam"/>
        <xs:element name="e-mail"/>
        <xs:element name="organisatie" minOccurs="0"/>
    </xs:sequence>
</xs:complexType>
<!-- Hieronder volgen de twee types inhoud -->
<xs:complexType name="nietXMLinhoudType">
    <xs:sequence>
        <xs:element name="body" type="bestandsgegevensType"/>
        <xs:element name="gerelateerdebestanden" type="bestandsgegevensType" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
</xs:complexType>
<xs:complexType name="XMLinhoudType">
    <xs:sequence>
        <xs:element name="XMLzelf">
            <xs:complexType>
                <xs:sequence>
                    <xs:element name="aanhef"/>

```

```

<xs:element name="blok" maxOccurs="unbounded">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="kop" minOccurs="0"/>
      <xs:choice maxOccurs="unbounded">
        <xs:element name="regel"/>
        <xs:element name="lijst">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="lijstelement"
maxOccurs="unbounded"/>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:choice>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="afsluitingfrase" maxOccurs="unbounded"/>
</xs:sequence>
</xs:complexType>
</xs:element>
<xs:element name="stylesheet" type="bestandsgegevensType"/>
</xs:sequence>
</xs:complexType>

```

```

<xs:complexType name="bijlagelijstType">
  <xs:sequence>
    <xs:element name="bijlage" type="bestandsgegevensType"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="bestandsgegevensType">
  <xs:sequence>
    <xs:element name="bestandsnaam"/>
    <xs:element name="bestandstype"/>
    <xs:element name="bestandslokatie"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="afzenderType">
  <xs:sequence>
    <xs:element name="naam"/>
    <xs:element name="e-mail"/>
    <xs:element name="functie" minOccurs="0"/>
    <xs:element name="organisatie" minOccurs="0"/>
    <xs:element name="telefoonnummer" minOccurs="0"/>
  </xs:sequence>
</xs:complexType>
</xs:schema>

```

Optionele and herhaalbare elementen zijn duidelijk herkenbaar aan een 'Unbounded' waarde. (Overheids-) organisaties moeten hun eigen opsomming formuleren voor Dossier en Werkproces. De waarden in dit schema dienen slechts als voorbeeld.

Appendix C

Bewaar Logbestand

De exacte inhoud van de Bewaar Logbestand is afhankelijk van de gekozen bewaar procedure. In ieder geval dient het log bestand minimaal de volgende informatie te bevatten:

Technische Metadata

- Details van de originele computer omgeving: client software = applicatie (bijv. Outlook) + server (bijv. Exchange) + Besturingssysteem;
- Details van de tussenformaten (bijv.ACSII, RTF);
- Details van de nieuwe computeromgeving (voldoende details moeten worden vastgelegd teneinde toegang te kunnen garanderen tot de digitale documenten in het huidige formaat).

Preservation Action Metadata

- Datum en tijdstip van iedere bewaaractie;
- Verantwoordelijke persoon van iedere bewaaractie;
- Details van de transformatie (omzettings-) software en
- Resultaten van de omzetting

Metadata die betrekking heeft op de toegang tot de records

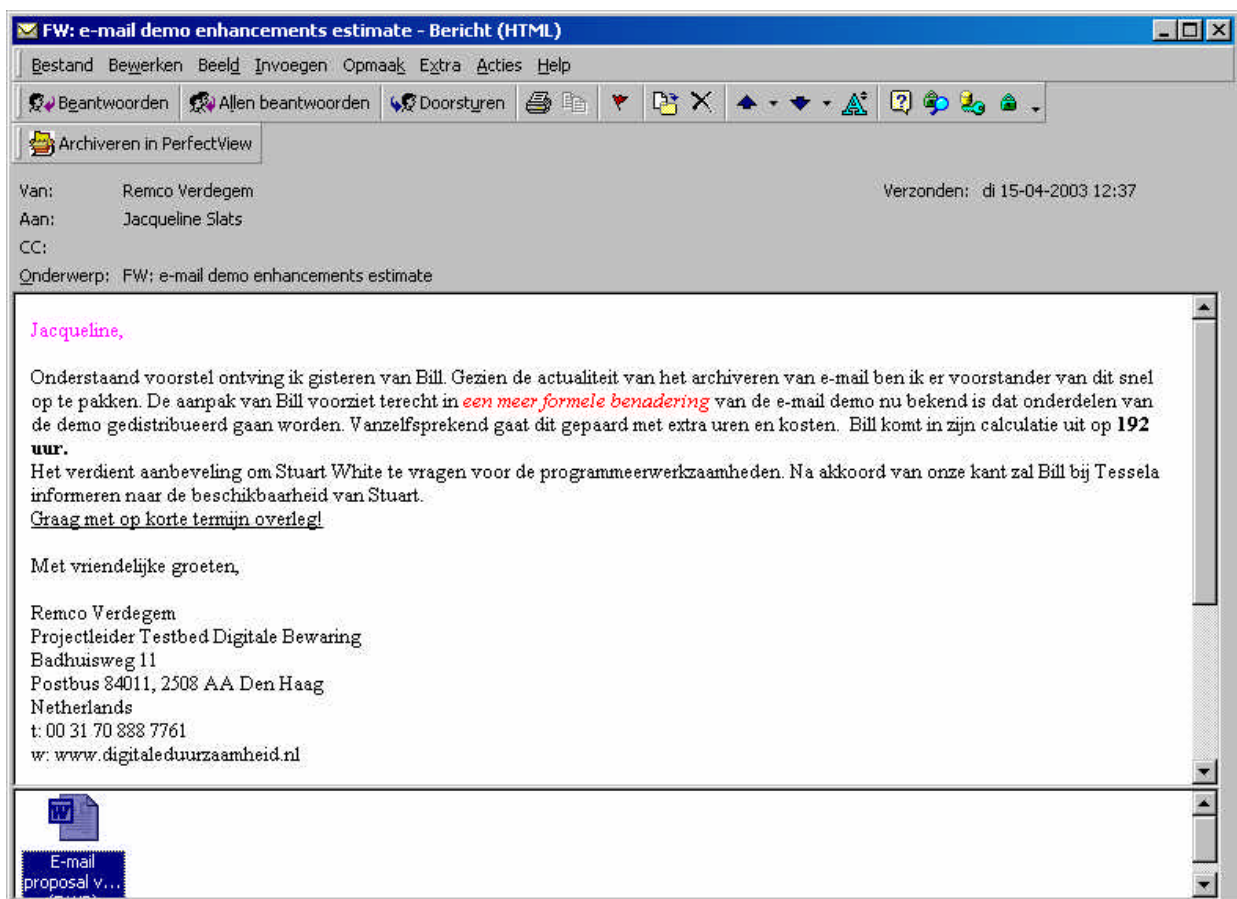
- Privileges/rechten en
- Geschiedenis

Appendix D

Voorbeeld van een e-mailbericht

Appendix D laat de verschillende representaties zien van een e-mailbericht.. Afbeelding D1 is een voorbeeld van een e-mailbericht, zoals getoond door de e-mail applicatie, in dit geval Outlook 2000. D2 toont het volledige transmissie bestand van hetzelfde e-mail bericht, inclusief alle transmissiegegevens en de verschillende representaties van de body van het bericht. D3 is een weergave van de body van het bericht in de vorm van platte tekst. D4 tenslotte laat het e-mail bericht zien als een XML bestand zoals gegenereerd door de door het Testbed ontwikkelde e-mail/XML demo.

Afbeelding D1: HTML representatie van het e-mailbericht



Afbeelding D2: Complete Transmissie Bestand

X-MimeOLE: Produced By Microsoft Exchange V6.0.6249.0
content-class: urn:content-classes:message
MIME-Version: 1.0
Content-Type: multipart/mixed;
 boundary="-----_NextPart_003_01C3033A.2C220780"
Subject: FW: e-mail demo enhancements estimate
Date: Tue, 15 Apr 2003 12:31:49 +0200
Message-ID: <F5714832E4D128479178CA6877D2280805F542@gw02.dh01.ictu.nl>
X-MS-Has-Attach: yes
X-MS-TNEF-Correlator:
Thread-Topic: FW: e-mail demo enhancements estimate
To: "Jacqueline Slats" <Jacqueline.Slats@ictu.nl>

This is a multi-part message in MIME format.

-----_NextPart_003_01C3033A.2C220780
Content-Type: multipart/alternative;
 boundary="-----_NextPart_004_01C3033A.2C220780"

-----_NextPart_004_01C3033A.2C220780
Content-Type: text/plain;
 charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable

Jacqueline,

Onderstaand voorstel ontving ik gisteren van Bill. Gezien de actualiteit =
van het archiveren van e-mail ben ik er voorstander van dit snel op te =
pakken. De aanpak van Bill voorziet terecht in een meer formele =
benadering van de e-mail demo nu bekend is dat onderdelen van de demo =
gedistribueerd gaan worden. Vanzelfsprekend gaat dit gepaard met extra =
uren en kosten. Bill komt in zijn calculatie uit op 192 uur.
Het verdient aanbeveling om Stuart White te vragen voor de =
programmeerwerkzaamheden. Na akkoord van onze kant zal Bill bij Tessela =
informerend naar de beschikbaarheid van Stuart.
Graag met op korte termijn overleg!

Met vriendelijke groeten,

Remco Verdegem
Projectleider Testbed Digitale Bewaring
Badhuisweg 11
Postbus 84011, 2508 AA Den Haag
Netherlands
t: 00 31 70 888 7761
w: www.digitaleduurzaamheid.nl

> -----Oorspronkelijk bericht-----
> Van: Bill Roberts
> Verzonden: maandag 16 september 2002 14:48
> Aan: Remco Verdegem

> Onderwerp: e-mail demo enhancements estimate
>
>

-----_=_NextPart_004_01C3033A.2C220780
Content-Type: text/html;
 charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<HTML><HEAD>
<META HTTP-EQUIV=3D"Content-Type" CONTENT=3D"text/html; =
charset=3Diso-8859-1">
<TITLE></TITLE>

<META content=3D"MSHTML 5.00.3103.1000" name=3DGENERATOR></HEAD>
<BODY>
<P><FONT =
color=3D#ff00ff>Jacqueline,

Onderstaand=20
voorstel ontving ik gisteren van Bill. Gezien de actualiteit van het =
archiveren=20
van e-mail ben ik er voorstander van dit snel op te pakken. De aanpak =
van Bill=20
voorziet terecht in een meer formele=20
benadering van de e-mail demo nu bekend is dat onderdelen =
van de=20
demo gedistribueerd gaan worden. Vanzelfsprekend gaat dit gepaard met =
extra uren=20
en kosten. Bill komt in zijn calculatie uit op 192=20
uur.
Het verdient aanbeveling om Stuart White te vragen voor =
de=20
programmeerwerkzaamheden. Na akkoord van onze kant zal Bill bij Tessela=20
informereren naar de beschikbaarheid van Stuart.
<U><FONT =
size=3D2>Graag=20
met op korte termijn overleg!

</U>Met =
vriendelijke=20
groeten,

Remco Verdegem
Projectleider Testbed Digitale=20
Bewaring
Badhuisweg 11
Postbus 84011, 2508 AA Den=20
Haag
Netherlands
t: 00 31 70 888 7761
w:=20
www.digitaleduurzaamheid.nl

> -----Oorspronkelijk=20
bericht-----
> Van: Bill Roberts
> Verzonden: maandag 16 =
september=20
2002 14:48
> Aan: Remco Verdegem
> Onderwerp: e-mail demo=20
enhancements =
estimate
>
>

</P></BODY></HTML>

-----_=_NextPart_004_01C3033A.2C220780--

-----_=_NextPart_003_01C3033A.2C220780
Content-Type: application/msword;
 name="E-mail proposal version 3.doc"
Content-Transfer-Encoding: base64

d

/ /

106

```
----- = NextPart 003_01C2E6F3.34C731CA--
```

Afbeelding D3: Inhoud van het e-mailbericht als platte tekst

Beste Jacqueline,
=20

Onderstaand voorstel ontving ik gisteren van Bill. Gezien de actualiteit =
van het archiveren van e-mail ben ik er voorstander van dit snel op te =
pakken. De aanpak van Bill voorziet terecht in een meer formele =
benadering van de e-mail demo nu bekend is dat onderdelen van de demo =
gedistribueerd gaan worden. Vanzelfsprekend gaat dit gepaard met extra =
uren en kosten. Bill komt in zijn calculatie uit op 192 uur.
Het verdient aanbeveling om Stuart White te vragen voor de =
programmeerwerkzaamheden; hij heeft immers de juiste ervaring. Na =
akkoord van onze kant zal Bill bij Tessela informeren naar de =
beschikbaarheid van Stuart.
Graag op korte termijn overleg!!!

Met vriendelijke groeten,

Remco Verdegem
Projectleider Testbed Digitale Bewaring
Badhuisweg 11
Postbus 84011, 2508 AA Den Haag
Netherlands
t: 00 31 70 888 7761
w: www.digitaleduurzaamheid.nl

> -----Oorspronkelijk bericht-----
> Van: Bill Roberts=20
> Verzonden: maandag 16 september 2002 14:48
> Aan: Remco Verdegem
> Onderwerp: e-mail demo enhancements estimate
>
>=20

Afbeelding D4: XML representatie van het e-mailbericht

```
<?xml version="1.0" encoding="UTF-8" ?>
- <ictu:e-mail xmlns:ictu="xmail.ictu.nl"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="xmail.ictu.nl E-mail.xsd">
-   <ictu:meta>
      <ictu:datum>2002-09-16T12:16:08+01:00</ictu:datum>
      <ictu:dossier>E-mailXML Demo</ictu:dossier>
      <ictu:werkproces>Het verstrekken van advies over het langdurig
        bewaren van e-mails</ictu:werkproces>
    </ictu:meta>
-   <ictu:transmissiebestand>
      <ictu:bestandsnaam>TransmissieBestand.txt</ictu:bestandsnaam>
      <ictu:bestandstype>text/plain</ictu:bestandstype>
      <ictu:bestandslokatie>c:\xmail_files\archive\E-mailXML
        Demo\mail9\TransmissieBestand.txt</ictu:bestandslokatie>
    </ictu:transmissiebestand>
      <ictu:onderwerp>FW: e-mail demo enhancements estimate
        </ictu:onderwerp>
-   <ictu:geadresseerdenlijst>
      - <ictu:geadresseerde>
          <ictu:naam>Jacqueline Slat</ictu:naam>
          <ictu:email>jacqueline.slat@ictu.nl</ictu:email>
        </ictu:geadresseerde>
      </ictu:geadresseerdenlijst>
-   <ictu:afzender>
      <ictu:email>remco.verdegem@ictu.nl</ictu:email>
      <ictu:naam>Remcol Verdegem</ictu:naam>
      <ictu:functie>Projectleider</ictu:functie>
      <ictu:organisatieonderdeel>Testbed Digitale
        Bewaring</ictu:organisatieonderdeel>
    </ictu:afzender>
-   <ictu:inhoud>
      - <ictu:nietXMLinhoud>
          - <ictu:body>
              <ictu:bestandsnaam>Body.html</ictu:bestandsnaam>
              <ictu:bestandstype>text/html</ictu:bestandstype>
              <ictu:bestandslokatie>c:\xmail_files\archive\E-mailXML
                Demo\mail9\Body.html</ictu:bestandslokatie>
            </ictu:body>
          </ictu:nietXMLinhoud>
        </ictu:inhoud>
    </ictu:e-mail>
```