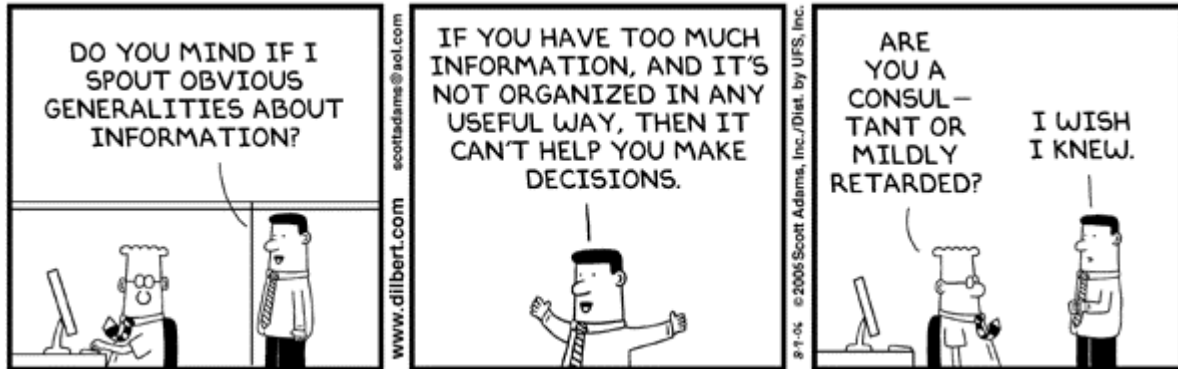


Orde in de digitale chaos?

Strategieën voor digitaal informatiebeheer. Enterprise Content Management en het Handboek Architectuur van de Gemeente Amsterdam, een vergelijkend onderzoek



Universiteit	Universiteit van Amsterdam (UvA)
Faculteit	Faculteit der Geesteswetenschappen
Leerstoelgroep	Leerstoelgroep Archief- en Informatiewetenschap
Opleiding	Documentaire Informatiewetenschap (DIW)
Onderdeel	Master-scriptie
Naam student	Drs. Niels Postma
Student nr.	9446206
E-mail	Postma75@gmail.com
Scriptiebegeleider	Drs. G.J. van Bussel
Referent	Prof. dr. J.S. Mackenzie Owen
Datum	25 mei 2008
Versie	1.0

Hoofdstukindeling

1. Probleemstelling & onderzoeksopzet	7
1.1. Inleiding.....	7
1.2. Probleemstelling.....	8
1.3. Onderzoeksopzet.....	10
2. Een definitie van ECM.....	11
2.1. Wat is belang van informatie.....	11
2.1.1. Waar moet het archief aan voldoen.....	12
2.2. Informatie en technologie.....	13
2.3. Wat is ECM.....	18
2.4. Definitie van ECM.....	19
3. De vijf componenten van ECM.....	24
3.1. Inleiding.....	24
3.2. De eerste component: Capture.....	27
3.2.1. Document imaging & Optical Character Recognition.....	27
3.2.2. Automated data capture.....	29
3.3. De tweede component van ECM: Manage.....	31
3.3.1. Document Management & Record Management	31
3.3.2. Workflow Management & Business Process Modelling.....	35
3.3.3. Content Management.....	38
3.3.4. Collaborative Software.....	43
3.4. De derde component: Store.....	45
3.4.1. Information retrieval.....	45
3.4.2. Ordeningsstructuur.....	47
3.4.3. Fulltext retrieval.....	48
3.4.4. Metadata search.....	51
3.5. De vierde component: Deliver.....	54
3.5.1. Transformatie & conversie.....	54
3.5.2. Distributie & outputmanagement.....	55
3.5.3. Samenstellen van informatie.....	55
3.5.4. Personalisatie.....	56
3.5.5. eXtended Markup Language	56
3.6. De vijfde component: Preserve.....	59
3.6.1. Opslagtechnologieën.....	59
3.6.2. Digitale duurzaamheid.....	61
4. ECM: voorwaarden en meerwaarde	63
4.1. Inleiding.....	63
4.2. Voorwaarde voor ECM: applicatie-integratie.....	63
4.2.1. Eilandautomatisering.....	63
4.2.2. Enterprise Application Integration	64
4.2.3. ECM suite.....	66
4.3. Voorwaarde voor ECM: standaardisering van metadata.....	68
4.3.1. Waarom standaardisering.....	68
4.3.2. Metadata-norm: NEN-ISO 23081-1:2006	70
4.3.3. De inhoud van metadata.....	72
4.4. Meerwaarde van ECM: afbreken van informatiesilo's.....	75
4.4.1. Wat zijn informatiesilo's.....	75

4.4.2. De risico's van informatiesilo's.....	76
4.5. Meerwaarde van ECM: Information Lifecycle Management.....	80
4.5.1. Wat is Information Lifecycle Management.....	80
4.5.2. De inhoudelijk benadering van ILM.....	82
4.5.3. De technische benadering van ILM.....	85
4.5.4. ECM & ILM	87
4.6. Conclusie theoretisch kader ECM.....	88
5. Casestudy.....	90
5.1. Inleiding.....	90
5.2. Onderzoekopzet.....	90
5.3. Karakteristieken van ECM.....	92
5.4. Samenvatting handboek architectuur Gemeente Amsterdam.....	95
5.5. Vergelijking karakteristieken ECM & handboek architectuur	101
5.6. Reflectie	111
5.6.1. Uitkomst vergelijking	111
5.6.2. Kantekening bij vergelijking.....	113
5.7. Conclusie.....	115
5.8. Nawoord: positionering ECM.....	115
Lijst van geraadpleegde literatuur.....	117
Bijlage 1: Procesbeschrijving.....	123
Bijlage 2: Britney Spears.....	124
Bijlage 3: Records Continuüm model.....	125

Lijst met afbeeldingen

Figuur 1: opbouw scriptie.....	10
Figuur 2: the expanding digital universe.....	15
Figuur 3: verhouding gestructureerde- en ongestructureerde informatie.....	16
Figuur 4: groei ontwikkeling gestructureerde- en ongestructureerde informatie.....	17
Figuur 5: ICT ontwikkeling 1970 – heden	19
Figuur 6: poster ECM 101.....	22
Figuur 7: digitalisering papieren documenten.....	28
Figuur 8: resultaten OCR	28
Figuur 9: relatie data, documenten & records.....	32
Figuur 10: schematische weergave RMA.....	34
Figuur 11: DMS & RMA.....	35
Figuur 12: hiërarchische opbouw procesarchitectuur.....	36
Figuur 13: schematische weergave workflow-systeem.....	37
Figuur 14: modelmatige representatie van een werkproces.....	38
Figuur 15: CMS granulariteit	39
Figuur 16: opbouw van content management systeem.....	40
Figuur 17: CMS, scheiding content en opmaak.....	42
Figuur 18: werking van content management systeem.....	42
Figuur 19: ongeordende samenwerking.....	43
Figuur 20: repository gebaseerde samenwerking.....	43
Figuur 21: concept information retrieval systeem.....	46
Figuur 22: ordeningsstructuur.....	48
Figuur 23: zoekterm Britney Spears.....	49
Figuur 24: inverted file.....	50
Figuur 25: inverted file.....	50
Figuur 26: zoekscherm DMS.....	52
Figuur 27: conversie & transformatie.....	55
Figuur 28: personalisatie.....	56
Figuur 29: XML-document.....	57
Figuur 30: XML - DTD.....	57
Figuur 31: XML-document met DTD.....	58
Figuur 32: Network Attached Storage.....	60
Figuur 33: duurzaamheid informatiedragers.....	62
Figuur 34: spaghettiarchitectuur.....	65
Figuur 35: Enterprise Application Integration.....	65
Figuur 36: schematische weergave verhouding metadata NEN-ISO 23081-1: 2006.....	71
Figuur 37: beschrijving van registratieveld.....	73
Figuur 38: informatiesilo's versus ECM.....	75
Figuur 39: GFO zaken model.....	78
Figuur 40: zaakdossier DMS.....	79
Figuur 41: zaakdossier ECM.....	79
Figuur 42: informatiebeheer in papieren omgeving.....	83
Figuur 43: levenscyclus van document.....	84
Figuur 44: relatie werkproces bewaarbeleid.....	85
Figuur 45: raadpleegfrequentie documenten.....	85
Figuur 46: kostenvergelijking opslagmedia.....	86
Figuur 47: relatie werkproces ILM.....	87

Figuur 48: typologie applicatielandschap (globaal).....	105
Figuur 49: model 7.9 het toekomstige applicatielandschap: wat kan gemeenschappelijk?	107
Figuur 50: verhouding ECM - handboek architectuur: informatie.....	111
Figuur 51: verhouding ECM - handboek architectuur: applicatie.....	112
Figuur 52: verhouding ECM - handboek architectuur: scope.....	112
Figuur 53: positionering ECM in informatie huishouding.....	116
Figuur 54: zoektermen 'Britney Spears' bij Google.....	124
Figuur 55: records continuüm model.....	125

1. Probleemstelling & onderzoeksopzet

1.1. Inleiding

Ik ben een digitaal document, althans dat was ik bij mijn geboorte. Ik ben de laatste versie in een lange lijn van voorgangers. Geen van hen bleek te voldoen aan de eisen van mijn auteur, maar voor de zekerheid zijn ze toch maar bewaard. Hoewel ik me uniek voel, heb ik de eigenaardige eigenschap dat ik me kan vermenigvuldigen zonder dat er onderscheid is te maken tussen mij en mijn nieuwe tweelingbroer of -zus. Kort na mijn geboorte gingen ik en enkele tweelingbroers en -zussen op reis naar verschillende bestemmingen om gelezen en bekeken te worden. Mijn auteur maakt vele documenten, zowel privé als voor zijn werk. Mijn broers en zussen die hij voor zijn werk tot leven wekt, leiden een luxueus leven vergeleken met mij. Zo worden zij gecreëerd en beheerd in systemen die daar speciaal voor zijn ontwikkeld. Ook hun geboorte kan een zwaar proces zijn, maar vindt wel plaats in een veilige omgeving. Ze noemen deze digitale kraamkamer ook wel een document management systeem. Bij de creatie zijn vaak meerdere mensen betrokken, die zich allemaal met onze inhoud bemoeien. Schaamteloos wordt er aan ons gesleuteld en van alles veranderd. Vindt men dat we er klaar voor zijn, dan gaan we op reis. Ieder van ons heeft een gepredestineerde reis die hij/of zij zal maken. Volgens een vooraf gedefinieerd pad gaan we langs afdelingen en personen die ons gebruiken voor de uitvoering van hun werk. De reisorganisatie die dit allemaal verzorgt en in goede banen leidt noemen we workflow. Zijn we volgroeid dan krijgen we het stempel *definitief* en mogen we eindelijk uitrusten in de veilige omgeving van de record management applicatie. Hier kan niemand meer iets aan ons veranderen en wachten we geduldig tot iemand ons weer komt opzoeken. Voor sommigen van ons is een internationale carrière weggelegd. Met de snelheid van het licht cirkelen zij rond de wereld in de vorm van e-mail of hebben ze een plekje verworven op het internet.

Ze noemen ons wel documenten, maar we zijn in veel opzichten niet te vergelijken met onze papieren naamgenoten. Achter ons uiterlijk gaat een wereld aan informatie schuil. Door middel van geavanceerde systemen houden wij bijvoorbeeld bij welke versie we zijn, wie ons heeft bekeken en wie ons heeft aangepast. Grappend zeg ik wel eens dat George Orwell er bang van zou worden als hij wist wat wij allemaal weten en bijhouden. Wij zijn de laatste jaren een veel besproken onderwerp. Ons aantal groeit hard en het ziet er dan ook naar uit dat ik er nog een groot aantal broers en zussen bij ga krijgen. Bij de laatste telling op het werk van mijn auteur waren er al meer dan een miljoen van ons in omloop. Een strategie om ons te beheren is dan ook geen overbodige luxe, maar bittere noodzaak geworden. Ik ga over twee van deze strategieën die het beheer van digitale informatie tot doel hebben.

1.2. Probleemstelling

Zowel bij de overheid als in het bedrijfsleven is digitalisering een veelbesproken onderwerp. Veel gehoorde kreten hierbij zijn onder andere ‘digitaliseren van het archief’ en ‘digitaliseren van de werkprocessen’. Vaak wordt gesproken over digitalisering als iets wat men van plan is in de (al dan niet nabije) toekomst te gaan doen.

Naar mijn mening is dit een foute voorstelling van zaken, om de simpele reden dat het proces van digitalisering al lang gaande is. Zo is er bijna geen document meer te verzinnen dat niet digitaal gecreëerd wordt. Wat in mijn optiek dan ook het onderwerp van gesprek zou moeten zijn, is niet zozeer digitalisering, maar *digitaal informatiebeheer*. Het gaat daarbij om de vraag hoe we de aanwezige digitale informatie zodanig kunnen organiseren dat zij optimaal wordt gebruikt om de doelstellingen van de organisatie te verwezenlijken. In dit verband wordt vaak Enterprise Content Management genoemd als dé oplossing voor het beheer van digitale informatie.

‘Enterprise Content Management is de totaaloplossing voor het beheren van alle ongestructureerde informatie (content) in uw gehele organisatie. Deze informatie is aanwezig in vele verschillende formaten: tekstdocumenten, technische tekeningen, XML bestanden, beeldmateriaal, PDF bestanden etc. Met Enterprise Content Management is uw organisatie in staat al deze bestanden te beheren’.¹

Vanuit mijn werk voor de Gemeente Amsterdam ben ik een ander initiatief tegengekomen voor het beheer van digitale informatie, namelijk het *Handboek Architectuur*. Dit handboek van de Gemeente Amsterdam is een door het college van B&W vastgestelde blauwdruk voor de gemeentelijke informatievoorziening. In de architectuur wordt op hoofdlijnen weergegeven hoe de verschillende componenten en initiatieven in de organisatie en de informatievoorziening in elkaar grijpen, zowel functioneel als technisch. Het handboek architectuur vormt hiermee het bestemmingsplan voor de toekomstige organisatie en informatievoorziening van Amsterdam. In dit plan wordt circa vier jaar vooruit gekeken.²

Tijdens het lezen van het handboek drong zich bij mij direct de vraag op of Enterprise Content Management (ECM) en het *Handboek Architectuur* eigenlijk hetzelfde zijn, of dat het twee verschillende technieken zijn om digitale informatie te beheren.

Dit lijkt in eerst instantie een simpele vraag. Toch kon ik er niet zomaar een antwoord op geven. De reden hiervoor is dat het handboek een (tot op zekere hoogte) concrete uitwerking is van de vraag hoe de informatiehuishouding van de Gemeente Amsterdam eruit moet komen te zien en hoe ze beheerd moet worden. De vraag wat ECM is en wat hier precies onder moet worden verstaan, is minder eenvoudig te beantwoorden. Omschrijvingen van ECM als ‘een totaaloplossing voor het beheren van alle ongestructureerde informatie (content) in uw gehele organisatie’ verschaffen weinig duidelijkheid over wat ECM nu eigenlijk precies is. Is ECM een technologie? Is het een strategie? Een hype? Of misschien wel een combinatie daarvan?

¹ <http://www.nl.capgemini.com/diensten/technology/ecm/> (24-05-2008)

² *Handboek architectuur, de samenhang in organisatie en informatievoorziening van de Gemeente Amsterdam* - adviesgroep architectuur versie 1.0 (april 2007) pag. 1-2

Een korte zoektocht op het internet geeft een veelvoud aan uiteenlopende interpretaties over wat er onder ECM dient te worden verstaan. Een quickscan van de wetenschappelijke literatuur bood evenmin direct een antwoord. Het eerste wat opvalt, is dat er in de wetenschappelijke literatuur nog maar weinig is geschreven over ECM, een bevinding die ook tot uiting wordt gebracht in het artikel ‘Contemporary issues of enterprise content management, the case of Statoil’.

‘Our search in major academic outlets and databases showed that there has not yet been much focus in information systems (IS) research on the concept of ECM from the viewpoint of an organization utilizing content management technology. (We searched for the phrase “content management”, in title or abstract, in the electronic databases of ACM, AIS, IEEE, Springer, and EBSCOhost. This was assumed to give a general-level picture of the current status of content management as a research topic within the IS discipline.) The articles that explicitly address content management mainly report on particular technical functionality of content management software (Kerer et al. 2002; Surjanto et al. 2000; Tyrväinen et al. 2003), or provide purely conceptual suggestions (Goodwin and Vigden 2002; Han and Pape 2002). Few articles refer to an organizational context for content management.’³

Om de vraag te kunnen beantwoorden of ECM en het handboek architectuur twee verschillende termen zijn voor hetzelfde principe moet dus eerst duidelijk worden wat ECM precies is.

Met bovenstaande constatering was mijn **onderzoeksvraag** geboren:

‘Zijn Enterprise Content Management (ECM) en het Handboek Architectuur twee verschillende termen voor hetzelfde principe of zijn het twee verschillende benaderingen voor het beheren van dezelfde digitale informatie?’

Zoals ik hierboven al heb aangegeven, moet eerst duidelijk worden wat er precies onder ECM wordt verstaan. Om mijn onderzoeksvraag adequaat te kunnen beantwoorden zal ik eerst op een aantal deelvragen ingaan. Die deelvragen zijn:

Wat is de definitie van ECM?

Wat zijn de belangrijkste elementen van ECM?

Wat zijn de voorwaarden voor ECM?

Wat is de meerwaarde van ECM ?

³ Munkvold, Bjørn; Erik, Päivärinta; Tero, Hodne; Anne, Kristine; Stangeland, Elin - *Contemporary Issues of Enterprise Content Management: The Case of Statoil* – In: Scandinavian Journal of Information Systems, 2006, 18(2):69-100, alhier pag. 71

1.3. Onderzoeksopzet

Om de onderzoeksvraag en bijbehorende deelvragen te beantwoorden gebruik ik de volgende methode.

Theoretisch kader

Eerst schep ik op basis van een literatuuronderzoek een theoretisch kader waarin de vraag wat ECM is, kan worden beantwoord. Dit gebeurt in drie hoofdstukken.

Hoofdstuk 2 is een inleidend hoofdstuk waarin ik eerst inga op het belang van informatie, de informatie-explosie die al een aantal jaren gaande is en de verschillende typen informatie die kunnen worden onderscheiden.

Vervolgens geef ik een heldere definitie van ECM. Zowel in dit hoofdstuk als in de daarop volgende hoofdstukken is een groot aantal plaatjes en modellen opgenomen. Dit is bewust gedaan en heeft als doel de soms erg abstracte materie door middel van visuele ondersteuning inzichtelijker te maken. De definitie van ECM in hoofdstuk 2 gaat ervan uit dat de in ECM gebruikte technieken en applicaties zijn onder te verdelen in vijf componenten.

In hoofdstuk 3 beschrijf ik eerst op hoofdlijnen de vijf componenten. Vervolgens bespreek ik per component de belangrijkste technieken en applicaties.

Hoofdstuk 4 bestaat uit twee delen.

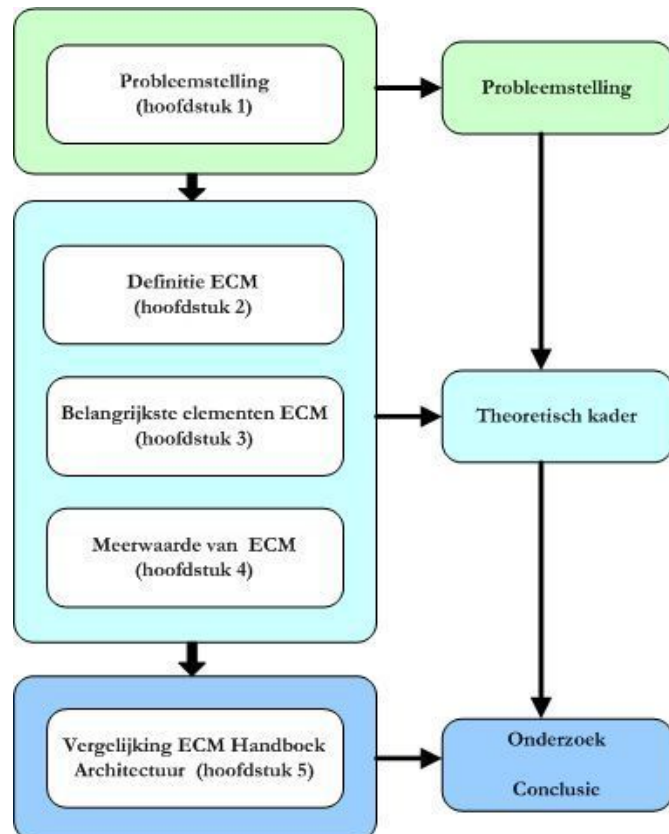
Eerst ga ik in op de vraag wat ervoor nodig is om ECM te implementeren, oftewel, ik ga in op 'de voorwaarden voor ECM'. In het tweede deel behandel ik de vraag wat nu precies de meerwaarde van ECM is.

De drie hoofdstukken tezamen vormen het theoretische kader waarin ik antwoord geef op de vraag wat ECM is.

Onderzoek

Op basis van deze drie hoofdstukken abstraheer ik een negental karakteristieken van ECM. Het gaat daarbij om zowel kenmerken als doelstellingen van ECM. Deze karakteristieken vergelijk ik vervolgens met het handboek architectuur en ik ga na op welke punten ECM en het handboek architectuur overeenkomen en op welke punten dit mogelijk niet het geval is.

De hypothese is dat als ECM en het handboek architectuur verschillende termen voor hetzelfde principe zijn, de voor ECM gedefinieerde karakteristieken ook in het handboek tot uiting moeten komen. Op basis van deze vergelijking geef ik vervolgens een reflectie en ten slotte een antwoord op de onderzoeksvraag.



Figuur 1: opbouw scriptie

2. Een definitie van ECM

2.1. Wat is belang van informatie

Organisaties creëren, ontvangen en versturen grote hoeveelheden informatie die zowel digitaal als analoog kan zijn. Om die informatie te kunnen beheren, worden ze opgenomen in het archief, dat ook zowel digitaal als analoog kan zijn. Bij archief denken de meeste mensen in eerste instantie aan een donkere kelder waar dozen vol documenten worden opgeslagen die niet meer gebruikt worden, maar waarvan om een of andere reden toch nog geen afscheid kan worden genomen. Dit is zeker niet wat in deze scriptie met archief wordt bedoeld; in deze scriptie staat de opvatting centraal dat archief procesgebonden informatie is. Deze definitie gebiedt mogelijk enige uitleg. Binnen een organisatie zijn het creëren, ontvangen en versturen van informatie geen op zichzelf staande activiteiten, maar maken ze altijd deel uit van een groter kader te weten het werkproces waarbinnen de informatie een rol speelt. Archief is dus informatie die voortkomt uit werkprocessen, daarin gebruikt en bewaard wordt, om vervolgens op een later moment mogelijk opnieuw een rol in die werkprocessen te spelen. Documenten worden in het kader van werkprocessen gecreëerd, gebruikt, ontvangen, verstuurd en vormen daarmee de administratieve neerslag van de wijze waarop die processen hebben plaatsgevonden. Archief is dus gedocumenteerde informatie die voortkomt uit of ontvangen wordt tijdens de uitvoering van werkprocessen binnen een organisatie.⁴ Nu is vastgesteld wat archief is, kan een nog belangrijkere vraag worden gesteld, namelijk: waarom hebben we het eigenlijk?

Er is binnen organisaties een overvloed aan informatie die op de een of andere manier beheerd moet worden. Of dit nu in papieren of digitale vorm gebeurt het is en blijft een belangrijke kostenpost. Waarom slaan we al die informatie op? Met andere woorden, wat is het belang van informatie voor een organisatie?

De belangrijkste reden is dat informatie onontbeerlijk is voor de bedrijfsvoering. Bij de uitvoering van hun werkzaamheden hebben mensen de ondersteuning van informatie nodig. Maar de waarde van informatie is moeilijk in te schatten. Iedere organisatie dient zichzelf dan ook de volgende vragen te stellen: wat is het waard om bepaalde kennis en informatie in huis te hebben? Wat is de potentiële waarde van de informatie nu en voor het toekomstig gebruik? En hoe is deze waarde te berekenen? Naast ondersteuning van de bedrijfsvoering zijn er voor een organisatie nog een aantal andere belangrijke redenen om een archief te hebben en te onderhouden.⁵

a. *Bedrijfseconomische redenen.* Iedere organisatie dient vanuit bedrijfseconomisch oogpunt een analyse te maken van risico's en bedreigingen die gemoeid zijn met het verlies van gegevens over werkprocessen. Er kan hierbij bijvoorbeeld worden gedacht aan het verlies van klantenbestanden, opgestelde rapporten en openstaande facturen. Risico's en bedreigingen doen zich voor wanneer niet de juiste documenten worden gecreëerd, de documenten niet voldoen aan de vereiste kwaliteit en wanneer de noodzakelijke documenten niet op de juiste wijze worden opgeslagen.

b. *Kennisbron.* Documenten kunnen een waardevolle bron van kennis zijn en vormen daarmee een belangrijk bedrijfskapitaal. Door middel van kennismanagement en *data mining* kan de vastgelegde informatie benut worden voor het ontdekken van trends en het beter van dienst zijn van de klant.

c. *Verantwoording.* Documenten kunnen aantonen dat een activiteit is uitgevoerd of juist niet is uitgevoerd. Zo kan er op basis van documenten op een later moment verantwoording worden afgelegd over het gevoerde beleid. Er kan gereconstrueerd worden welke beslissingen zijn genomen en op basis van welke documenten die beslissingen zijn genomen. Vooral bij de

⁴ Horsman, Peter J. - *Archiveren. Een inleiding* - Stichting Archiefpublicaties, 's-Gravenhage (2004) pag. 10, 21

⁵ Mensink, Dennis - *Je zit op een goudmijn maar je weet het niet* - Facto magazine nr. 4, (2005); pag. 10-15 alhier pag. 12

overheid is verantwoording een belangrijke reden voor het houden van een archief. Dit omdat een voorwaarde voor het democratisch proces is, dat de overheid zich ten allen tijde moet kunnen verantwoorden voor het gevoerde beleid. Dat informatie en het beheer ervan vitaal belang is wordt nog eens duidelijk onderstreept in het advies: *Informatie: grondstof met toekomstwaarde*, opgesteld door de Raad voor het Openbaar Bestuur en de Raad voor Cultuur.

‘Een goede en betrouwbare informatiehuishouding is van vitaal belang voor de overheid. Alle producten en prestaties van de overheid hangen hiermee samen. De juiste beleidsinformatie, correcte kengetallen, wetgevingskwaliteit, informatie aan burgers en vooral: het vermogen om verantwoording af te leggen zijn hier van afhankelijk. De kwaliteit van de informatiehuishouding staat in rechtstreeks verband met de algemene beginselen van goed bestuur.’⁶

d. *Juridische reden*. Hierbij gaat het voor een organisatie om de mogelijkheid om recht te halen of zich te verdedigen. Documenten kunnen een belangrijke rol spelen als bewijs, de beschikbaarheid van een document waarborgt de bewijsfunctie. Het gevolg is echter dat een organisatie geconfronteerd wordt met de noodzaak tot het bewaren en opslaan van een enorme hoeveelheden informatie. De kosten van het opslaan dienen wel in verhouding te staan met de verwachte schade bij het ontbreken van de documenten. Er dient dan ook een analyse te worden gemaakt welke risico's aanvaardbaar zijn wanneer bij een claim vanuit de omgeving de organisatie in de rechtzaal zijn gelijk wil halen.

e. *Cultureel erfgoed*: een deel van de geproduceerde informatie kan beschouwd worden als cultureel erfgoed. Met behulp van die documenten kunnen onderzoekers op een later moment een beeld vormen van een gebeurtenis uit het verleden. Vooral voor de archiefbescheiden van de overheid speelt dit een belangrijke rol. Om die reden is in de archiefwet vastgelegd dat bepaalde documenten niet voor vernietiging in aanmerking komen en na verloop van tijd door de organisatie aan een archiefbewaarplaats moeten worden overgedragen.⁷

2.1.1. Waar moet het archief aan voldoen

Een organisatie kan dus op verschillende manieren profiteren van het onderhouden van een archief. Als een archief onderdeel uitmaakt van de bedrijfsvoering moet men wel kunnen vertrouwen op de informatie die zich hierin bevindt. Daarom wordt aan archiefdocumenten een viertal eisen gesteld:

1. *Authenticiteit*. Met authenticiteit wordt bedoeld of het document ook echt is wat het beweert te zijn. Is het bijvoorbeeld geen vervalsing? Is het document ook echt afkomstig van degene die beweert het te hebben opgemaakt of verzonden? Is het document opgemaakt of verzonden op het aangegeven tijdstip? Hierbij dient er op gelet te worden dat authenticiteit en waarheid niet door elkaar worden gebruikt. De inhoud van een authentiek document kan onwaar zijn en de inhoud van een niet-authentiek document kan toch een correcte weergave van de werkelijkheid zijn.

⁶*Informatie: grondstof met toekomstwaarde* - advies opgesteld door Raad van Cultuur en de Raad voor het Openbaar Bestuur (2008); pag. 5

http://www.rob-rfv.nl/default.aspx?skin=Rob&inc=detail&nieuws_id=851&type=actueel

⁷ Horsman, Peter J. - *Archiveren. Een inleiding*; pag. 16-17

NEN-ISO 15489-1(nl), *Informatie en documentatie – Archiefbeheer. Deel 1: Functionele eisen voor informatie- en archiefmanagement* - Nederlands Normalisatie-instituut (2001); pag. 7

Arnold, A.G.; Glashouwer, B.. - *De moderne informatiehuishouding van de digitale overheid, het archief op het bureau* - Stichting Het Expertise Centrum, Amsterdam (2005); pag. 31

http://www.hec.nl/index.php?page_id=58

2. *Betrouwbaarheid*. De inhoud van het document moet in overeenstemming zijn met de feiten. De gebruiker van het document moet op de inhoud kunnen vertrouwen en ervan uit kunnen gaan dat het document geen onwaarheden bevat.
3. *Integriteit*. Het document moet volledig en ongewijzigd zijn. Het is noodzakelijk dat informatie wordt afgeschermd tegen ongeautoriseerde wijzigingen. Elke geautoriseerde wijziging behoort expliciet te worden aangegeven en herkenbaar te zijn.
4. *Bruikbaarheid*. Een document dat is gearhiveerd, moet kunnen worden teruggevonden, gepresenteerd en begrepen. Daarnaast behoort de directe relatie tot het werkproces of de transactie waaruit het document is voortgekomen te kunnen worden weergegeven. Met het begrijpen van een document wordt bedoeld dat een document ook door iemand anders dan de opsteller moet kunnen worden begrepen. Een document met losse cijfers zegt de toekomstige gebruiker waarschijnlijk weinig, maar als uit het document kan worden opgemaakt dat het een jaarrekening is, dan krijgen de losse cijfers contextwaarde en kan het document worden begrepen. Als aan deze vier voorwaarden niet wordt voldaan, dan is het archiveren van documenten een zinloze bezigheid.⁸

2.2. Informatie en technologie

Nu het belang van informatiebeheer is vastgesteld, kan er gekeken worden naar welke digitale hulpmiddelen hiervoor gebruikt worden. Zelfs voor een leek is waarneembaar dat de informatietechnologie zich in een snel tempo ontwikkelt. Er is dan ook een groot verschil waarneembaar tussen de manier waarop informatie in de beginnendagen van de informatietechnologie werd gebruikt en verwerkt en hoe dit tegenwoordig gebeurt. In de beginnendagen werd informatietechnologie gebruikt voor *data processing* en *number crunching*. De computer werd gebruikt voor de opslag en verwerking van grote hoeveelheden numerieke data. Ondanks dat numerieke dataverwerking nog steeds een voornaam onderdeel is van het gebruik van computers, is er een duidelijk verschil tussen de beginperiode van de informatietechnologie en de huidige situatie. Het belangrijkste verschil is dat naast numerieke data tekstuele informatie en grafische afbeeldingen thans alom vertegenwoordigd zijn in de informatietechnologie. Tekstuele en grafische informatie zijn bij uitstek geschikt voor betekenisoverdracht en spelen daarmee dus een belangrijke rol in de overdracht van kennis, inzichten, meningen en ideeën. Deze ontwikkeling heeft tot gevolg dat het gebruik van informatietechnologie zich tegenwoordig veel meer richt op de creatie, distributie, opslag en de ontsluiting van producten van de geest in plaats van alleen maar op het verwerken van numerieke data. Deze ‘switch’ van dataverwerking naar uitwisseling van betekenisvolle informatie tussen mensen is ook waarneembaar in de gebruikte terminologie. Werd in de beginperiode gesproken van ‘automatisering’, tegenwoordig heeft men het over ‘informatie- en communicatietechnologie’ (ICT).⁹

Informatie-explosie

Onder invloed van de ontwikkelingen op het gebied van informatie- en communicatietechnologie zijn er nieuwe vormen van communicatie ontstaan. Daarmee heeft het digitale tijdperk ons ook een exponentiële groei aan informatie gebracht. Belangrijke reden voor de enorme groei aan informatie is dat voor de verspreiding van digitale informatie gebruik wordt gemaakt van uiteenlopende verschijningsvormen. Informatie komt tegenwoordig in talloze vormen zoals foto's, tekstdocumenten, webpagina's, spreadsheets, presentaties, grafische afbeeldingen, tekeningen, e-mail en video. De informatie-explosie, ook wel aangeduid als het uitdijende digitale

⁸ NEN-ISO 15489-1:2001 - pag. 11

Horsman - *Archiveren. Een inleiding* - pag. 23-24, 85

⁹ Mackenzie Owen, John - *The scientific article in the age of digitization* - (2005); pag. 129-130
<http://dare.uva.nl/document/17843>

universum, heeft ook een keerzijde. De enorme hoeveelheid digitale informatie kan complex en moeilijk te beheren zijn. En door de exponentiële groei ervan wordt het probleem snel groter. Een bijkomend probleem is dat al deze elektronische documenten ook nog eens moeten worden geïntegreerd met een reeds complexe verzameling papieren informatie.¹⁰ Onderstaande statistieken geven inzicht in de omvang van de exponentiële groei van informatie wereldwijd en het daaruit voortvloeiende belang deze informatie te beheren. Daarnaast is, om enige grip op de omvang van de geproduceerde informatie te krijgen, ook een overzicht gemaakt van de opslagformaten.

- In 1999, zo heeft de Universiteit van Berkeley berekend, is er wereldwijd 2 exabytes, oftewel 2 miljard gigabyte aan informatie geproduceerd. Het gaat daarbij om zowel analoge als digitale informatie
- In 2002 is er 2,5 keer zoveel informatie geproduceerd als in 1999, 5 exabytes in totaal. Om grip te krijgen op de omvang is het handig het met boeken te vergelijken. Als alle 19 miljoen boeken en ander gedrukte documenten van grootste bibliotheek ter wereld, de Library of Congress, gedigitaliseerd zouden worden zou dit ongeveer tien terabytes aan informatie zijn. Als deze vergelijking wordt doorgetrokken dan staat vijf exabytes aan informatie gelijk aan een half miljoen keer de omvang van de Library of Congress.
- 92 % van de in 2002 geproduceerde informatie is digitaal opgeslagen.¹¹
- In een onderzoeksrapport van onderzoeksbedrijf IDC naar de groei van wereldwijd digitaal geproduceerde informatie wordt berekend dat alleen al in 2006 de astronomische omvang van meer dan 160 exabytes aan digitale informatie is geproduceerd. Om de omvang van 160 exabytes in perspectief te plaatsen: dit is te vergelijken met 3 miljoen keer de informatie uit alle boeken die ooit zijn geschreven of 12 stapels met boeken van de aarde naar de zon.
- Driekwart van de in 2006 gecreëerde 160 exabyte aan informatie wordt door particulieren gecreëerd, 25 procent, 40 exabyte, wordt door organisaties gecreëerd.
- In 2007 zal de hoeveelheid digitaal geproduceerde informatie voor het eerst de

Hoeveel is een zettabyte?

Kilobyte (KB) is 1000 bytes (10^3 bytes)

1 KB: een getypte pagina

100 KB: een lage resolutie foto

Megabyte (MB) is 1000 kilobyte

1.000.000 byte (10^6 bytes)

1 MB: een novelle

5 MB: de complete Shakespeare

100 MB: 1 meter boeken

500 MB: een CD-ROM

Gigabyte (GB) is 1.000 megabyte

1,000,000,000 bytes (10^9 bytes)

1 GB: een vrachtwagen vol boeken

5 GB: een DVD-ROM

Terabyte (TB) is 1000 gigabytes

1.000.000.000.000 bytes (10^{12} bytes)

1 TB: bedrukt papier van 50,000 bomen

2 TB: een universiteitsbibliotheek

10 TB: gedrukte collectie van de Library of Congress

(totaal 19 miljoen gedrukte werken)

20 TB: het .gov domein in februari 2004

Petabyte (PB) is 1000 terabytes

1.000.000.000.000.000 bytes (10^{15} bytes)

2 Petabytes: Alle informatie uit alle academische onderzoeksbibliotheken in de V.S.

200 Petabytes: Al het gedrukte materiaal

Exabyte (EB) is 1000 petabytes

1.000.000.000.000.000.000 bytes (10^{18} bytes)

2 Exabytes: totale omvang van informatie geproduceerd in 1999

Zettabyte (ZB) is 1000 exabytes

1.000.000.000.000.000.000.000 bytes (10^{21} bytes)

¹⁰ *Enterprise Content Management, de route naar brede acceptatie* - Microsoft (2006); pag. 6

<http://www.microsoft.com/netherlands/adviseurs/productiviteit/ecmoplossingen.aspx>

¹¹ *How Much Information? 2003* - Universiteit van Berkeley (2003); pag.1-2

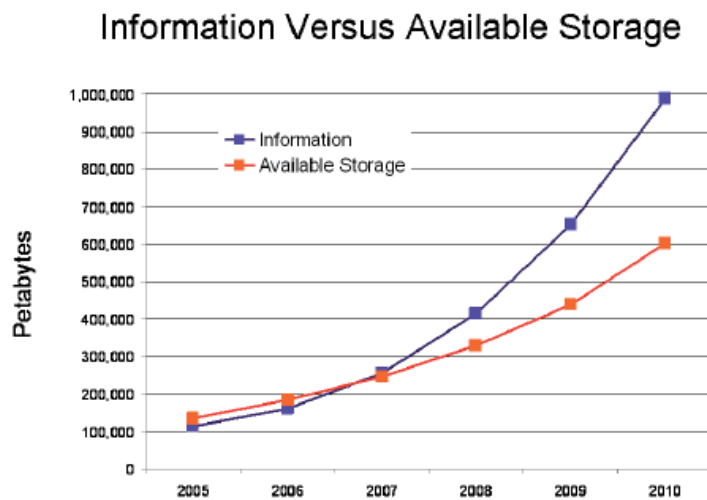
http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf

totale beschikbare opslagruimte overstijgen, een trend die zich in de daaropvolgende jaren zal voortzetten (zie hiervoor ook figuur 2).

- IDC voorspelt dat de omvang van de in 2010 geproduceerde digitale informatie is toegenomen tot ongeveer één zettabyte. Als we bij de vergelijkingen met boeken blijven, dan zal in 2010 een stapel boeken kunnen worden gemaakt van de zon naar Pluto en weer terug.

- Het percentage van de informatie door organisaties gecreëerd zal in 2010 zijn toegenomen tot 30 procent, ongeveer 330 exabyte.

- Ondanks dat er in 2010 een zettabyte aan digitale informatie zal worden geproduceerd, zal de totale beschikbare opslagruimte 'slechts' zijn gegroeid tot 600 exabyte. Ter vergelijking, een zettabyte staat gelijk aan een 160 gigabyte harddisk voor elke bewoner van de aarde.¹²



Figuur 2: the expanding digital universe

Gestructureerd versus ongestructureerd

Hierboven is consequent gebruik gemaakt van de term 'informatie'. Het is echter mogelijk om verschillende typen informatie te onderscheiden. Zo is er bij de beschrijving van de ontwikkeling van het computergebruik een onderscheid gemaakt tussen enerzijds numerieke data en anderzijds tekstuele en grafische informatie. Deze twee typen informatie kunnen worden aangeduid als gestructureerde en ongestructureerde informatie. Onder gestructureerde informatie wordt informatie verstaan die op een vooraf bepaalde gestructureerde manier in een database is opgeslagen. Over het algemeen betreffen dit numerieke data. Ongestructureerde informatie is alle overige informatie. Ongestructureerde informatie kan dan ook het best omschreven worden als 'alle informatie die niet gestructureerd is'. Het gaat daarbij om e-mails, rapporten, audio, video, etc. Onder ongestructureerde informatie worden vaak woorden/tekst verstaan. Dit onderscheid tussen gestructureerde (cijfers) en ongestructureerde (woorden, audio, video enz.) informatie behoeft wel enige nuancering. Het belangrijkste criterium voor gestructureerde informatie is of de betreffende informatie past binnen een vooraf gedefinieerd informatiemodel, bijvoorbeeld een tabel in een database. Uitgaande van deze definitie kunnen bijvoorbeeld persoonsnamen (woorden) die in een database worden opgenomen als gestructureerde informatie worden beschouwd en zijn cijfers die in een tekst voorkomen ongestructureerde informatie.

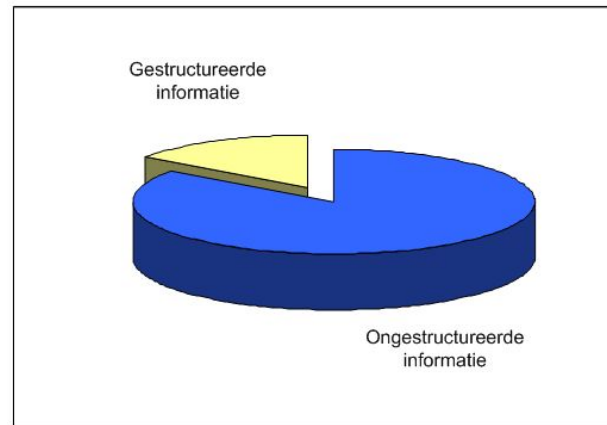
¹² Gantz, John F. - *The expanding digital universe, a forecast of worldwide information Growth through 2010* - IDC (2007); pag. 1-4

http://www.emc.com/about/destination/digital_universe/pdf/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf

De definitie van ongestructureerde informatie die in deze scriptie gebruikt wordt, luidt als volgt:

‘Unstructured information (UI) is any document, file, image, report, form, etc. that has no defined, standard structure that would enable convenient storage in unit record or similar automated processing devices; it cannot be defined in terms of rows and columns or records; and the data cannot be examined with standard unit record access.’¹³

Zoals hierboven aangegeven, is onder invloed van de mogelijkheden van de informatietechnologie de hoeveelheid digitale informatie sterk toegenomen. Binnen organisaties is digitale ongestructureerde informatie meestal opgeslagen in een verscheidenheid aan applicaties maar ook op de gedeelde schijven of gewoon op de persoonlijke harde schijf. De vraag welk percentage van de informatie gestructureerd of ongestructureerd is, valt niet gemakkelijk te beantwoorden, omdat dit mede wordt bepaald door het type organisatie. Als uitgangspunt kan worden genomen dat rond de tachtig procent van de informatie binnen organisaties ongestructureerd is.¹⁴



Figuur 3: verhouding gestructureerde- en ongestructureerde informatie

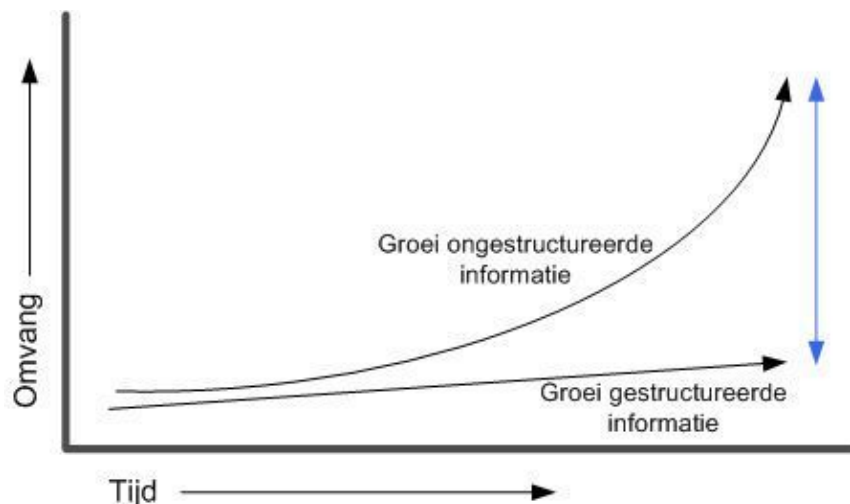
Nog belangrijker dan de vraag welk percentage van de informatiehuishouding op dit moment uit ongestructureerde informatie bestaat, is de vraag hoe de verhouding zich in de toekomst gaat ontwikkelen. Zoals hierboven reeds aangegeven, lag de focus voorheen voornamelijk op het beheer van gestructureerde informatie. De ontwikkeling die op dit moment plaatsvindt en zich naar verwachting in de toekomst zal voortzetten, is dat de focus steeds meer verschuift naar het beheer van ongestructureerde informatie. Hiervoor zijn twee belangrijke argumenten te geven. Ten eerste wordt men zich steeds meer bewust van de waarde van ongestructureerde informatie. Men realiseert zich steeds meer dat de betekenis van informatie door nieuwe ontwikkelingen of wisseling van perspectief altijd aan verandering onderhevig is. Ongestructureerde informatie kan verborgen kennis bevatten die pas van waarde blijkt wanneer de informatie op een later moment nodig is.

Een tweede argument is de exponentiële groei van de hoeveelheid informatie en van ongestructureerde informatie in het bijzonder. Analisten geven aan dat de hoeveelheid ongestructureerde informatie binnen grote organisaties zich iedere twee maanden verdubbelt. Mogelijk is deze inschatting enigszins aan de hoge kant, maar ook als de omvang zich iedere jaar verdubbelt, is dat een enorme groei. Zeker als men bedenkt dat gestructureerde informatie een lineaire groei van rond de tien procent vertoont. De exponentiële groei van ongestructureerde informatie maakt het noodzakelijk dat er meer aandacht wordt besteed aan hoe die gecreëerd en beheerd moet worden.¹⁵

¹³ Ladley, John - *Beyond Rows and Columns, Unstructured Information, Part 3* - DM Review Online (2003); http://www.dmreview.com/article_sub.cfm?articleId=6284

¹⁴ Jenkins, Tom; Kölher, Walter; Shackleton, John - *Enterprise Content Management Methods, what you need to know* - Ontario (2006); pag. 91
Ladley, John - *Beyond Rows and Columns, Unstructured Information*.

¹⁵ Jenkins, Tom; Forquer, Bill; Jelinski Peter - *Enterprise Content Management Solutions, what you need to know* - Ontario (2006); pag. 18-20



Figuur 4: groei ontwikkeling gestructureerde- en ongestructureerde informatie

Wet en regelgeving

De exponentiële groei van ongestructureerde informatie heeft niet alleen tot gevolg dat het steeds tijdrovender wordt om uit de informatieberg de juiste en waardevolle informatie te vinden, het brengt ook nieuwe bedrijfsrisico's met zich mee die betrekking hebben op wet- en regelgeving. Wat betreft het informatiemanagement bestaat er een keur aan wetten waarmee rekening dient te worden gehouden, onder meer de Archiefwet, Archiefbesluit 1995, Ministeriële regelingen ex. art. 11, 12, 13, Archiefverordening, WOB, WBP, AWB, Wet bestuurlijk berichtenverkeer, Wet elektronische handtekening en Wet op de politieregisters.¹⁶ Voor het bedrijfsleven is er naast de reeds langer bestaande wetgeving recentelijk Amerikaanse wetgeving bijgekomen die ook van invloed is op Nederlandse bedrijven. Als gevolg van grote fraudezaken bij beursgenoteerde ondernemingen (zoals Enron) is in de Verenigde Staten in 2002 de Sarbanes-Oxley wet (SOx) in werking getreden. De wet, vernoemd naar de twee indieners, legt tal van regels op aan zowel binnenlandse als buitenlandse bedrijven die aan een Amerikaanse beurs genoteerd zijn. In dit verband wordt ook vaak de term *compliance* gebruikt.

Compliance betekent letterlijk 'naleving'. De laatste decennia heeft het woord een meer specifieke betekenis gekregen. De term compliance wordt gehanteerd bij de naleving van wet- en regelgeving bij met name financiële instellingen. Vanaf de jaren tachtig van de vorige eeuw worden er door de wetgever (en later de toezichthouders) steeds meer eisen gesteld aan financiële instellingen zoals de hierboven genoemde SOx-wetgeving. Financiële instellingen (en ook andere instellingen die raakvlakken hebben met de financiële sector) beschikken over een compliance-regeling. Hierin staan de voorschriften waaraan de in die instelling werkzame personen zich moeten houden, alsmede voorschriften omtrent de wijze van controle op de naleving ervan. Een dergelijke regeling bevat bijvoorbeeld voorschriften over de wijze waarop personen voor eigen rekening in effecten mogen handelen, restricties in de omgang met zakenrelaties en (in relatie tot het onderwerp van de scriptie van groot belang) de omgang met (vertrouwelijke) informatie. De reikwijdte van compliance is niet zondermeer aan te geven. Dit, in combinatie met de complexiteit van het onderwerp, resulteert in een bewuste keuze om compliance buiten de scriptie te houden.¹⁷

¹⁶ Voor een uitgebreid overzicht van de wettelijke kaders voor de overheid met betrekking tot digitale documenten zie: Prins J.E.J.; Matthijssen L.J. - *De Digitale Overheid en de wet, de juridische kaders voor gebruik van digitale documenten bij overheden* - Den Haag (2000)

¹⁷ Claassen, Laurent - *Compliance is een' modewoord'* - Ernst & Young
http://www.ey.nl/download/publicatie/Compliance_is_een_modewoord.pdf

2.3. Wat is ECM

Voor de creatie en het beheer van zowel gestructureerde als ongestructureerde informatie zijn in de afgelopen decennia verschillende applicaties ontwikkeld. Voor het beheer van gestructureerde informatie heeft dit uiteindelijk in de jaren negentig geleid tot de ontwikkeling van ERP, wat staat voor Enterprise Resource Planning. ERP is een reactie op de beperkingen die kleefden aan de eerste automatiseringsgolf, waarin voor verschillende bedrijfsfuncties aparte applicaties werden gebruikt. Zo zijn er aparte programma's ontwikkeld voor de boekhouding, verkoop, urenregistratie, planning en dergelijke. Het grootste nadeel hiervan is de inefficiëntie, die wordt veroorzaakt doordat dezelfde gegevens keer op keer opnieuw ingevoerd moeten worden en de verschillende applicaties vaak niet met elkaar kunnen communiceren. ERP software gaat uit van een integrale aanpak van gestructureerde informatie. Informatie zit niet langer in verschillende, los van elkaar opererende applicaties, maar wordt beheerd in een geïntegreerd systeem dat is opgebouwd uit verschillende modules. Uitgangspunt van ERP software is dan ook dat informatie eenmalig wordt ingevoerd en eenmalig wordt opgeslagen. De ingevoerde gegevens kunnen vervolgens in alle modules worden benut, zodat alle opgeslagen informatie door het hele bedrijf, door iedereen kan worden gebruikt. ERP software ondersteunt op deze wijze werkprocessen als relatiebeheer, verkoop, inkoop, werkvoorbereiding, planning, urenregistratie, voorraad en financiën.¹⁸

Ook voor het creëren, beheren en distribueren van ongestructureerde informatie zijn in de afgelopen jaren verschillende applicaties ontwikkeld. De belangrijkste (in hoofdstuk 3 wordt hier uitgebreid op ingegaan) zijn document management systemen voor creatie, versiebeheer, check-in/checkout en korte termijnopslag van documenten. Record management applicatie die zich specialiseren in de fixatie en langdurige opslag van documenten. Omdat organisaties nog steeds een belangrijk deel van hun documenten in papieren vorm ontvangen en versturen, zijn document capture en imaging applicaties ontwikkeld voor de digitalisering van zowel inkomende als uitgaande stromen papier. Workflow-applicaties voor het ondersteunen van de werkprocessen door het op vooraf opgestelde wijze routeren van documenten door de organisatie voor controle, revisie en goedkeuringen en het gelijktijdig creëren van audittrails. Al deze typen applicaties zijn voorhanden en worden momenteel ingezet voor het beheer van de informatiehuishouding. Om dezelfde reden als ERP software is ontwikkeld, namelijk om te voorkomen dat informatie wordt opgesloten in een applicatie en daardoor niet beschikbaar is voor de rest van de organisatie, is er een nieuwe ontwikkeling gaande op het gebied van informatietechnologie: Enterprise Content Management (ECM).

ECM staat voor een integrale aanpak van ongestructureerde informatie die zich binnen organisaties bevindt. ECM heeft als doel het integreren van de bestaande informatietechnologieën voor het beheer van ongestructureerde informatie zoals document management, record management, webcontent management en workflow. ECM wordt in het algemeen beschouwd als een verzameling van afzonderlijke, maar onderling verbonden toepassingen om content uit verschillende bronnen vast te leggen, te verwerken en te beheren. Doel van ECM is dat alle bronnen van content worden geïntegreerd, onafhankelijk van de locatie van de content.

¹⁸ http://nl.wikipedia.org/wiki/Enterprise_Resource_Planning

De belangrijkste uitdagingen waar informatiemanagement op dit moment voor staat en waar ECM een oplossing voor probeert te bieden zijn:

- Het terugdringen van redundante informatie
- Het uitschakelen van informatiesilo's
- Voorkomen dat intellectueel kapitaal verloren gaat
- Voldoen aan wet- en regelgeving (compliance)
- Het vinden van de juiste, specifieke, waardevolle informatie tussen alle opgeslagen informatie (information retrieval)
- Al deze dingen doen op een efficiënte en kosteneffectieve manier

De hierboven geschetste ontwikkeling van de informatietechnologie kan als volgt schematisch worden weergegeven:

	1970s	1980s	1990s	2000s
Who	IBM	Microsoft	SAP	Open Text
Why	Clerical Productivity	Personal Productivity	Departmental Productivity	Org. Compliance & Productivity
How	Data Processing	Email; Desktop Publishing	ERP	ECM
Computing Environment	Mainframe	Personal Computer	LAN	Ubiquitous Computing

Figuur 5: ICT ontwikkeling 1970 – heden¹⁹

2.4. Definitie van ECM

In het voorgaande is een globale omschrijving gegeven van ECM. Voor een scriptie volstaat dit niet en het is dan ook nodig een heldere definitie van ECM te formuleren. Echter er bestaan meerdere definities van ECM. Forrester hanteert in haar rapport *Topic overview: Enterprise Content management* de volgende:

‘In its totality, ECM must be a strategy for: 1) how to manage all unstructured information — images, Web content, rich media assets, and corporate records; 2) how to mitigate content-related risks; and 3) how to put content to use for business people and business processes. By doing this, ECM technologies and strategies help information and knowledge management professionals manage the company's unstructured information to reduce risk, boost productivity, and improve customer experiences.’²⁰

Kenmerkend voor de definitie van Forrester is dat ervan uit wordt gegaan dat ECM een strategie is. Zij stelt dat ECM een strategie is waarin wordt bepaald hoe er wordt omgegaan met alle ongestructureerde informatie die binnen een organisatie aanwezig is, hoe de in verschillende *repositories* opgeslagen informatie met elkaar wordt geïntegreerd en hoe de informatie vervolgens bruikbaar wordt gemaakt door er context aan te geven. De definitie stelt dat ECM een strategie is

¹⁹ Bron: Jenkins, Tom; e.a. - *Enterprise Content Management Solutions* - pag. 17

²⁰ <http://www.forrester.com/Research/Document/Excerpt/0,7211,40327,00.html> (29-04-2008)

voor het beheer van de ongestructureerde informatie, maar zegt niets over hoe dit moet gebeuren of welke technologieën hiervoor gebruikt kunnen worden. Om die reden is de definitie misschien wel erg globaal. Een veel concretere definitie van ECM wordt gegeven door Gartner in het rapport *Magic Quadrant for Enterprise Content Management (2007)*. Het geheel van Enterprise Content Management bestaat volgens Gartner uit de volgende zes componenten:

1. *Document Management* for check-in/check-out, version control, security and library services for business documents.
2. *Document Imaging for capturing*, transforming and managing paper documents.
3. *Records Management* for long-term archiving, automation of retention and compliance policies, and ensuring legal, regulatory and industry compliance.
4. *Workflow* for supporting business processes, routing content, assigning work tasks and states, and creating audit trails.
5. *Web Content Management* for automating the Webmaster function and managing dynamic content and user interaction.
6. *Document-Centric Collaboration* for document sharing and supporting project teams.²¹

Is de definitie van Forrester mogelijk wat te globaal, een punt van kritiek op de definitie van Gartner is dat deze zich eigenlijk geheel richt op de specifieke technologie waarvan ECM gebruik maakt voor het managen van de ongestructureerde informatie. Dat de definitie zich beperkt tot de technologie heeft als belangrijk nadeel dat technologische veranderingen in de toekomst automatisch met zich mee brengen dat de definitie verouderd raakt en zal moeten worden aangepast.

Zoals hierboven aangegeven, bestaat ECM technologie uit verschillende applicaties voor informatiebeheer. Wat ECM van de losse applicaties voor informatiebeheer onderscheidt, is dat het de technologieën met elkaar integreert. ECM is dan ook meer dan alleen technologie. Het is tevens een strategie voor het beheer van informatie gedurende de gehele levenscyclus. Een definitie van ECM die hierbij goed aansluit is de definitie die de AIIM²² (Association of Information and Image Management) geeft van ECM.

‘Enterprise Content Management (ECM) is the technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists.’

‘While there are ECM technologies, more importantly, ECM is an ongoing and evolving strategy for maximizing how your content is to be used. (...) Technology can enable streamlined management of content, but the underlying strategy must come first.’²³

Bij de strategie voor het managen van de volledige levenscyclus van informatie, van ontstaan tot vernietiging, staan de tools en technologieën van ECM centraal. Hierbij wordt er door de AIIM van uitgegaan dat het niet genoeg is om content simpelweg te beheren, maar dat content zodanig ‘gemanaged’ moet worden dat deze bijdraagt aan de realisatie van bedrijfsdoelstellingen.

²¹ Shegda Karen M.; Bell, Toby; Chin, Kenneth; Gilbert, Mark R.- *Magic Quadrant for Enterprise Content Management (Gartner RAS Core Research Note G00150426)* - Gartner (2007); pag. 3
<http://mediaproducts.gartner.com/gc/reprints/ibm/external/volume2/article16/pdf/article16.pdf>

²² <http://www.aiim.org/>

²³ <http://www.aiim.org/about-ecm.asp> (18-03-2008)

'It's not enough to "manage" content. Of course, the ability to access the correct version of a document or record is important, but companies must go further. Content must be managed so that it is used to achieve business goals. Central to this strategy are the tools and technologies of ECM, which manage the complete lifecycle of content, birth to death.'²⁴

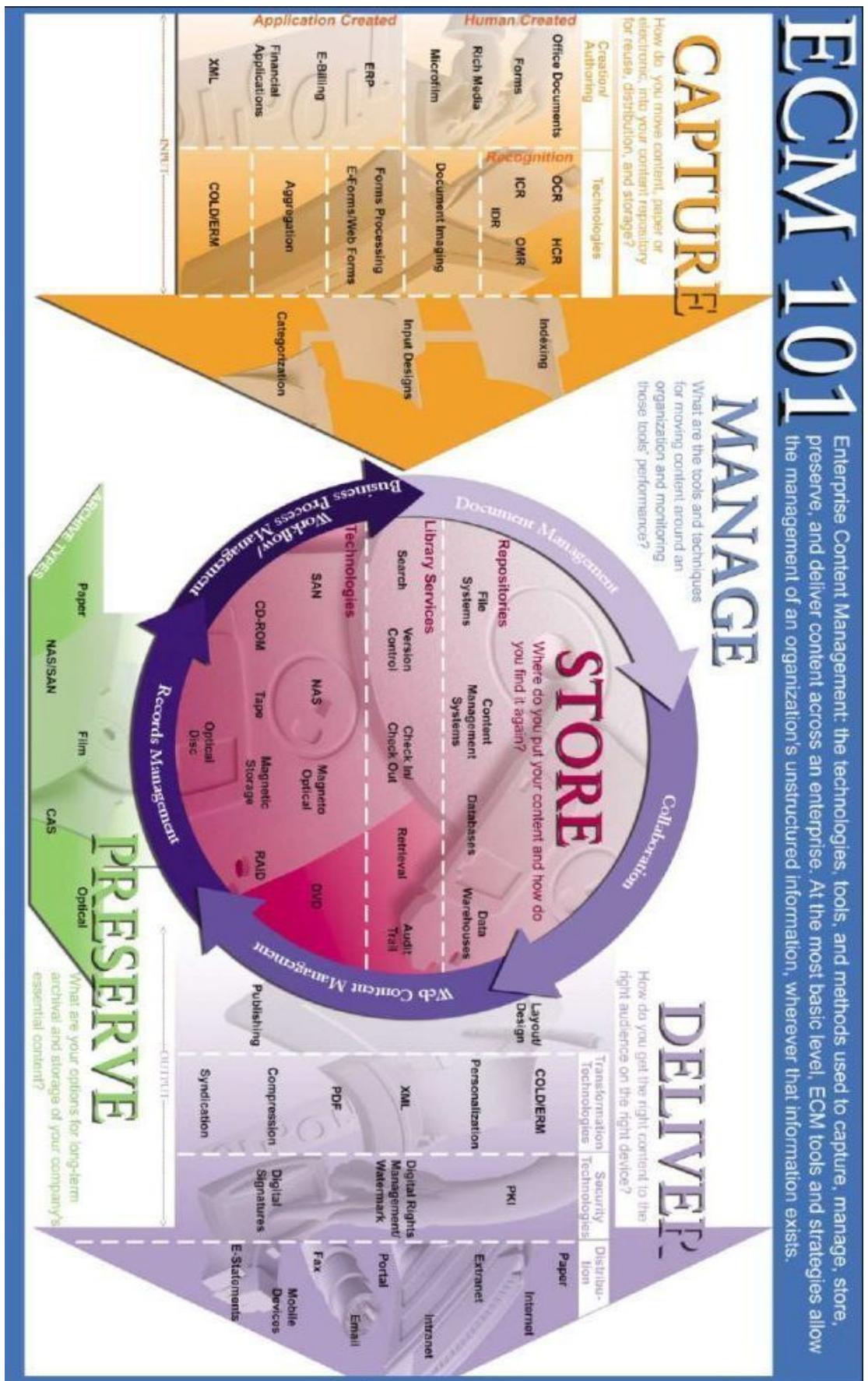
ECM is in de definitie van de AIIM dan ook een combinatie van strategie en technologie voor het managen van content van de organisatie. Veelzeggend in dit opzicht is de visie van Dr. Ulrich Kampffmeyer, voormalig lid van de raad van bestuur van AIIM Europa.

'The important thing is whether the term ECM or Enterprise Content Management means more functionality and benefit for the user. This applies to subsets of ECM as well as to its overall claim of managing enterprise content. With all the comprehensive claims and all the countless components of Enterprise Content Management, at the end of the day ECM is a vision, a strategy, or even a new industry, but it is not a closed system solution or a distinct product.'²⁵

In de definitie van AIIM komt duidelijk naar voren dat het principe van ECM uitgaat van nauw samenwerkende applicaties. Waarbij het bij ECM gaat om het samenspel van de applicaties en niet om de losstaande applicaties. Een complete ECM oplossing dient dan ook alle genoemde functionaliteiten te bieden, niet slechts een of twee. In de AIIM definitie worden er vijf componenten van ECM onderscheiden: capture (opnemen/ontsluiten), manage (beheren), store (opslag), preserve (lange termijnopslag) en deliver (distribueren). Elk van deze componenten is een onlosmakelijk onderdeel van ECM en bestaat uit een aantal tools en technologieën. De samenhang tussen de verschillende componenten is mooi gevisualiseerd in de AIIM poster : ECM 101. Per component worden de bijbehorende tools en technologieën genoemd.

²⁴ <http://www.aiim.org/about-ecm.asp> (18-03-2008)

²⁵ Kampffmeyer, Ulrich - *ECM enterprise content management* - (2006); pag. 2
http://www.project-consult.net/Files/ECM_White%20Paper_kff_2006.pdf



Figuur 6: poster ECM 101

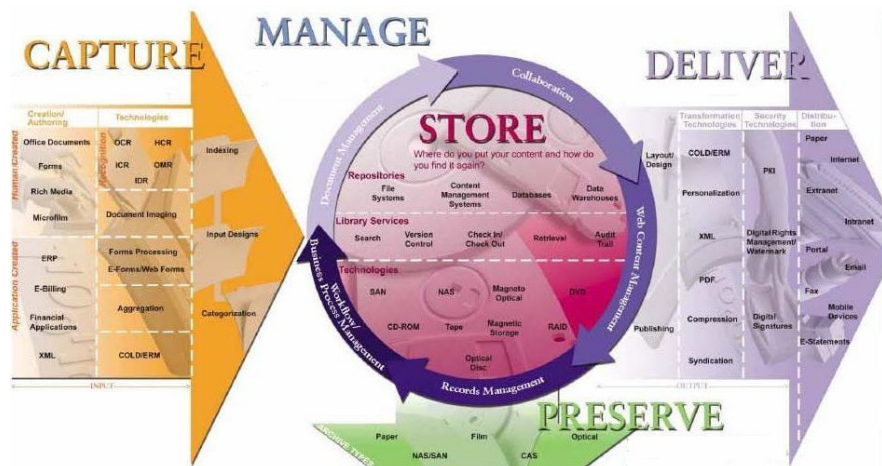
Er zijn twee redenen waarom de definitie van de AIIM in deze scriptie als uitgangspunt wordt genomen. De belangrijkste is dat die definitie uitgaat van een combinatie van strategie en technologie en zich niet slechts op een van die elementen richt. De definitie sluit aan bij mijn opvatting dat het één niet zonder het ander kan functioneren, geen succesvolle technologie zonder strategie en vice versa. De tweede reden is dat de technologische aspecten van ECM zeer concreet door de AIIM zijn neergezet in de poster EMC 101 die bij de definitie van de AIIM hoort. Dat de bijbehorende tools en technologie concreet genoemd worden, brengt het risico met zich mee dat de definitie snel verouderd. Omgekeerd wordt de definitie van ECM helder en begrijpelijker als duidelijk wordt welke tools en technologieën deel uitmaken van ECM. Door de tools en technologieën niet in de definitie zelf te zetten, zoals bij Gartner, maar los van de definitie in een poster te verwerken, wordt voorkomen dat door technologische veranderingen de definitie verouderd en is het toch mogelijk om concreet de ECM tools en technologieën te benoemen. Nu de definitie van ECM duidelijk is, wordt het tijd om dieper in te gaan op de vijf componenten van ECM.

3. De vijf componenten van ECM

3.1. Inleiding

Zoals ik in hoofdstuk 2 heb aangegeven, wordt Enterprise Content Management door de AIIM niet beschouwd als een gesloten systeem of specifiek product, maar als een strategie die bestaat uit een veelheid aan technologieën. Als we naar de poster (en de bijbehorende uitleg op de website van de AIIM) kijken, dan wordt daar een duizelingwekkend aantal termen genoemd waaruit ECM is opgebouwd.²⁶ Storage, repositories, content integration, backup/recovery, Bpm, wfm, dms, rma, cms, scanning, form processing, search/retrieval, security, collaboration, XML en portals om er maar een paar te noemen. Het moge duidelijk zijn dat het onmogelijk is om op alle technieken en applicaties in te gaan. Toch denk ik dat het voor een goed begrip van wat ECM is van belang is om de belangrijkste technieken en applicaties waaruit het bestaat op hoofdlijnen te beschrijven. De technologieën waaruit de ECM-strategie is opgebouwd worden door de AIIM onderverdeeld in vijf componenten:

- Capture
- Manage
- Store
- Deliver
- Preserve



Men dient zich hierbij te realiseren dat de indeling van het AIIM-model in componenten functiegericht is opgesteld en niet als een architectuur beschouwd dient te worden. Hieronder geef ik per component kort aan wat daaronder wordt verstaan en hoe de componenten zich tot elkaar verhouden. In paragraaf 3.2. tot 3.6. bespreek ik per component de belangrijkste technologieën.

Capture

De capture-component bestaat uit applicaties en technieken voor het genereren, opnemen, prepareren en verwerken van zowel analoge als digitale informatie. Het doel is het aanleveren van informatie aan de manage-component voor verdere verwerking en opslag. De capture-component wordt vaak als de input-component beschouwd. Het opnemen van informatie, die vervolgens in andere applicaties verder wordt verwerkt, kan op verschillende niveaus plaatsvinden en daar kunnen verschillende technologieën bij worden gebruikt. De mogelijkheden variëren van het op relatief simpel niveau inscannen van papieren documenten tot het op gecompliceerde manieren verwerken van informatie waarbij gebruik wordt gemaakt van automatische classificatie. Door dit brede spectrum aan mogelijkheden kunnen technologieën binnen de capture-component betrekking hebben op vele soorten informatie zoals papieren

²⁶ <http://www.aiim.org/about-ecm.asp>

documenten, in bureauapplicaties opgemaakte elektronische documenten, xml-documenten, formulieren, multimediatekstbestanden en microfilm.²⁷

Manage:

De manage-component heeft betrekking op het managen, verwerken en gebruiken van informatie. De ‘traditionele’ toepassingsgebieden zijn:

1. *Document management* voor check-in/checkout, versiebeheer, security en bibliotheekservices voor documenten.
2. *Document-centric collaboration* voor het delen van documenten door projectteams.
3. *Web content management* voor de automatisering van de taken van de webmaster.
4. *Record management* voor duurzame bewaring van documenten conform wet- en regelgeving.
5. *Workflow / Business process management* voor de routing van documentstromen.

De eerste drie applicaties, document management, document-centric collaboration en web content management richten zich met name op het dynamische deel van informatiebeheer. Record management daarentegen richt zich op de statische fase, waar in de opgeslagen informatie geen veranderingen meer (mogen) plaatsvinden. Workflow en business process management zijn de link tussen alle hiervoor genoemde applicaties.

Store

De store-component wordt gebruikt voor de tijdelijke opslag van informatie. De store-component, en de technologieën die hierin worden gebruikt, kunnen in drie categorieën worden onderverdeeld. Als eerste zijn er de repositories, die als opslaglocatie van informatie fungeren. Het gaat hier o.a. om filesystemen en databases. De tweede categorie betreft de library services. Deze fungeren als administratieve componenten die de toegang regelen tot de informatie in de repositories. Library services hebben o.a. betrekking op functionaliteiten als versiebeheer, zoekmogelijkheden, audittrails en autorisatie. De derde categorie betreft de opslagtechnologieën, waarbij het gaat om technologieën als SAN (storage area network), NAS (network attached storage), DVD en optische disks.²⁸

Deliver

De deliver-component wordt gebruikt voor de presentatie van de informatie die zich bevindt in de manage-, store- en preserve- componenten. De deliver-component wordt ook wel aangeduid als output management. De twee belangrijkste functies met betrekking tot output zijn:

- Layout/ontwerp: tools waarmee de vormgeving van de output wordt bepaald.
- Publiceren: applicaties voor het distribueren en publiceren van de informatie.

De deliver-component is onder te verdelen in drie subgroepen: transformation technologies, security en distribution.

Onder transformation technologies verstaat men technologieën als COLD, XML, PDF en personalisatie. Het gaat hierbij om technologieën die betrekking hebben op de output van de informatie. In het bijzonder gaat het om de mogelijkheid om (inhoudelijk) dezelfde informatie in verschillende vormen (web pagina, pdf of gedrukte folder) te presenteren. Dit is mogelijk door de inhoud en opmaak van documenten van elkaar te scheiden. Deze technologieën zijn over het algemeen actief op de achtergrond en zijn dan ook voor de gebruiker vaak onzichtbaar.

²⁷ Kampffmeyer, Ulrich - *ECM enterprise content management* - pag. 30-36

²⁸ Kampffmeyer - *ECM enterprise content management* - pag. 54-57

Security-technologieën hebben betrekking op informatiebeveiliging. Beveiliging beperkt zich niet tot alleen de deliver-component. Zij heeft betrekking op alle componenten en dient dan ook beschikbaar te zijn voor alle componenten. Zo wordt een digitale handtekening bijvoorbeeld niet alleen gebruikt bij de verzending van informatie maar ook bij de ontvangst (capture).

Distribution heeft betrekking op de verzending en presentatie van informatie. Alle drie de technologieën maken mogelijk dat verschillende soorten informatie via verschillende routes aan de gebruiker geleverd kunnen worden. Voor de distributie van informatie kan men gebruik maken van zowel actieve technieken als e-mail, als van passieve technieken, waarbij de gebruiker zelf informatie zoekt op bijvoorbeeld de website of het intranet. Doel van de distribution is de gevraagde informatie zo goed mogelijk bij de gebruiker aan te leveren.²⁹

Preserve

De preserve-component wordt gebruikt voor lange termijnopslag, veilige opslag en backup van statische informatie. De technologieën die hiervoor gebruikt worden, komen voor een deel overeen met die van de store-component. Enkele technologieën die hiervoor worden gebruikt, zijn NAS, SAN, WORM, maar ook papier en microfilm.³⁰

In de definitie van de AIIM bestaat ECM uit een groot aantal technieken en applicaties, waarvan in de meeste organisaties op dit moment al één of meerdere in gebruik zijn. Het zijn dan ook niet de losse componenten zelf maar de strategie om de verschillende componenten met elkaar te integreren die ECM zijn meerwaarde geeft. Voordat we gaan kijken naar de meerwaarde van ECM, is het van belang om niet alleen een opsomming te geven, maar ook om de belangrijkste componenten van ECM te beschrijven. Hieronder ga ik per categorie dieper in op de technologieën en op de vraag waar ze voor gebruikt kunnen worden. Zoals hierboven ook al aangegeven bestaat ECM uit een enorme hoeveelheid technologieën. Vanwege de beperkte omvang van deze scriptie heb ik er voor gekozen om per component alleen op de belangrijkste technologieën in te gaan. De nadruk zal hierbij liggen op de manage-component.

²⁹ Ibidem, pag. 66-71

³⁰ Ibidem, pag. 62-64

3.2. De eerste component: Capture

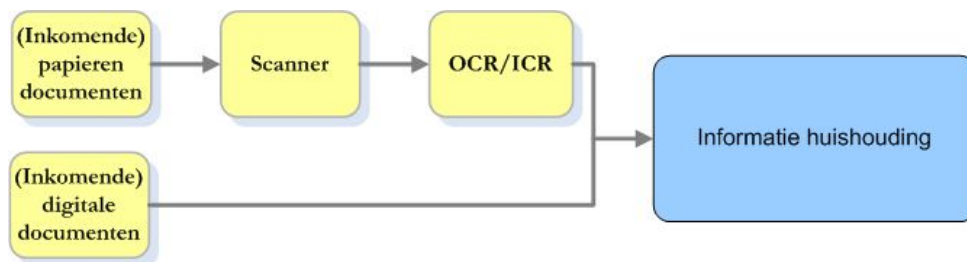


De capture-component draait om de vraag ‘How do you move content, paper or electronic, into your content repository for reuse, distribution, and storage?’. De capture-component bestaat hiervoor uit technieken en applicaties voor het genereren, opnemen, prepareren en verwerken van zowel analoge als digitale informatie. Het doel van de component is het aanleveren van informatie aan de manage-component voor verdere verwerking en opslag. Hieronder zal ik twee belangrijke technieken uit de capture-component bespreken, namelijk document *imaging*/OCR en *automated data capture*.

3.2.1. Document imaging & Optical Character Recognition

Ondanks het vergevorderde stadium van digitalisering wordt veel informatie nog steeds in papieren vorm gedistribueerd. Om die analoge informatie op te kunnen nemen in de manage-component zal zij gedigitaliseerd moeten worden. Document imaging is het proces van digitalisering van papieren documenten dat tot doel heeft deze als elektronische bestanden in de informatiehuishouding op te nemen. Gedigitaliseerde documenten kunnen op twee manieren worden opgeslagen:

1. Op een grafische manier als image. Er wordt dan als het ware een elektronische fotokopie gemaakt van het document. Dit betekent dat de computer het image ook als fotokopie behandelt, dus als één geheel. Het doet er niet toe of de inhoud van het document bestaat uit een tekst, een foto, een tekening of een combinatie daarvan. In een image kunnen afzonderlijke elementen van het document dus niet worden herkend.
2. Als tekst. De papieren documenten worden eerst met behulp van scanners geconverteerd naar een elektronisch formaat. Vervolgens wordt het elektronische bestand met behulp van Optical Character Recognition (OCR) of Intelligent Character Recognition (ICR) ingelezen. Het grote voordeel hiervan is dat het bestand hiermee fulltext doorzoekbaar is geworden. Er kan nu binnen het bestand op elk woord of wordelement waaruit de tekst bestaat worden gezocht.



Figuur 7: digitalisering papieren documenten

Het gebruik van OCR en ICR maakt het mogelijk om op een efficiënte manier informatie te onttrekken aan papieren documenten die zijn geconverteerd naar elektronische bestanden. Het voornaamste doel van de herkenningstechnologieën als OCR, ICR, Optical Mark Reader (OMR), barcodes en patchcodes is de reductie van de tijd die nodig is voor indexering en de toekenning van metadata aan geconverteerde documenten. De tijd die nodig is voor het handmatig indexereren van documenten vormt vaak de bottleneck in de bedrijfsbrede ontsluiting van (fysieke) documenten. De kwaliteit en de bruikbaarheid van documenten, die zijn gescand en vervolgens ingelezen met OCR wordt bepaald door twee factoren. Ten eerste de kwaliteit van de analoge documenten die gescand worden. Factoren hierbij zijn het gebruikte lettertype en of er bijvoorbeeld koffievlekken op het document zitten. De tweede bepalende factor is de nauwkeurigheid van de gebruikte OCR-technologie. De resultaten van de bewerking met OCR kunnen sterk van elkaar verschillen zoals het volgende voorbeeld laat zien.³¹

New Humphrey mystery

Origineel document

HUMPHREY the Downing Street cat, cleared of killing four robins a fortnight ago, is back under a cloud. He has been spotted by police and security men returning to No 10 with a baby duck in his mouth, suspiciously like one of several birds reported missing from Duck Island in nearby St. James's Park. Malcolm Kerr, head keeper, said: "Humphrey is far from innocent." The cat, aged six, is believed to operate under cover of darkness and to be an expert at dodging traffic on his park trips.

Resultaat na bewerking met OCR systeem 1

New Humphrey mystery
 HUMPHREY the DownLng Street cat, cleared of killi'ig fotir robins a fortnight ago, la back under a cloud. He has been spotted by police and security men retilrning to No 10 with a baby duck in Ms mouth, suspiciously like one of several birds reported missing from Duck Island In nearby St. James's PariL Malcolm Kerr, head keeper, said: "Humphrey la far from Innocent." The cat, aged six, la believed to operate wider coe'er of darkness and to be an expert at dodge tram£ on `lie park trips.

Resultaat na bewerking met OCR systeem 2

New Humphrey mystery
 HREY the Downing Street cat, c four robim a fortnight ago, is back under a cloud. He has been spotted by PoUce and security men returning to No 10 with a baby d-uck in his mouth, suspiciously Uke one of several b@ reported ffrom Duck Is@d in nearby St. James's Park. Maloolm Kerr, head keeper, said, "Humphrey is far fmm bmocent.' The cat, aged s@ is believed to operate under of darkness and to be an expert at dodging trame on his park trips.

Figuur 8: resultaten OCR

³¹ *Analysis, Selection, and Implementation Guidelines Associated with Electronic Document management Systems (EDMS)* - An AIIM Recommended Practice Report prepared by the Association for Information and Image Management International (2006); pag. 7, 14-15
<http://www.aiim.org/documents/standards/arp1-2006.pdf>.

Er zijn op dit moment OCR-programma's die een nauwkeurigheid van 99.3% behalen. Dit betekent dat, mits het te scannen document van goede kwaliteit is, minder dan 1 op de 100 karakters verkeerd door het systeem wordt geïndexeerd. Ervan uitgaand dat de gemiddelde lengte van een woord zes letters is, betekent dit dat bij gemiddeld 1 op de 17 woorden een letter fout wordt geïndexeerd. Ook als het nauwkeurigheidsperscentage minder hoog ligt, levert dat niet noodzakelijkerwijs problemen op. Tom Landaur laat in zijn onderzoek zien dat zelfs bij een foutenpercentage van 50% per woord (1 op de 2 woorden bevat één of meer foute karakters) het vinden van de juiste informatie nog steeds 20% beter verloopt dan bij het gebruik van handmatig toegekende trefwoorden. Zijn onderzoek richt zich specifiek op medische wetenschappelijke artikelen en de resultaten kunnen dan ook niet zomaar gegeneraliseerd worden. Het onderzoek toont wel aan dat een hoog foutenpercentage niet een onoverkomelijk probleem hoeft te zijn. De reden hiervoor is volgens Landaur dat woorden die van groot belang zijn voor de inhoud van de tekst naar alle waarschijnlijkheid vele malen in de tekst worden gebruikt, waardoor het OCR-programma de term minimaal één keer correct inleest.³² Om het nauwkeurigheidsperscentage te verbeteren is het, afhankelijk van de gebruikte software, ook mogelijk dat er nabewerkingen op de scan plaatsvinden. Woorden die een fout geïndexeerd karakter bevatten, worden handmatig verbeterd. Echter: 'Use of image 'clean up' and other post-scanning processing should only be used to improve legibility. Caution should be exercised when using these tools, as any material modification to the image may affect the ability to authenticate the document in a legal proceeding.'³³

3.2.2. Automated data capture

Herkenningstechnologieën als OCR zijn met name effectief als de informatievraag zich richt op specifieke delen van gestandaardiseerde documenten. Bij deze gelijkvormige documenten is het mogelijk om specifieke 'zones' te identificeren. De daarin aanwezige informatie kan men door middel van OCR naar tekst converteren en vervolgens in de applicatie opnemen. Dit proces wordt *automated data capture* genoemd. In de praktijk blijkt dat hiermee niet alleen tijd en geld wordt bespaard, maar dat ook de kwaliteit van het indexeren verbetert. Desondanks verdwijnt de noodzaak tot handmatig indexeren hiermee niet helemaal. De hoeveelheid informatie die automatisch uit de documenten wordt gewonnen, wordt bepaald door de kwaliteit van de input en de nauwkeurigheid van het systeem. Kan het systeem niet de gewenste kwaliteit garanderen, dan zal handmatige controle en indexering (zij het in verminderde mate) op enigerlei wijze nodig blijven.³⁴ Op het moment dat men het gebruik van 'automated data capture' overweegt, dient men over de volgende punten na te denken:

- Is het mogelijk typen documenten te identificeren waarvan de hoeveelheid zo groot is dat 'automated data capture' wordt gerechtvaardigd? Het gaat in het algemeen om formulieren die zowel gestructureerde als ongestructureerde informatie bevatten en waarbij standaardisatie het mogelijk maakt om de gevraagde informatie te identificeren.
- Is het mogelijk de gebruikte formulieren te herontwerpen om zo het systeem de documenten beter te laten herkennen? Het gebruik van checkboxes³⁵, barcodes en specifieke lettertypen leiden namelijk tot meer nauwkeurigheid.

³² Lesk, Michael - *Understanding digital libraries* - second edition, San Francisco (2005); pag. 55-57

³³ *Analysis, Selection, and Implementation Guidelines Associated with Electronic Document management Systems (EDMS)* - pag. 15

³⁴ *Analysis, Selection, and Implementation Guidelines* - AIIM, pag. 7,14-15

³⁵ Scanners kunnen beschikken over een patchcode detector die de patchcodes op de documenten herkent waarna er een automatische scheiding van bestanden kan worden gemaakt.

- Hoe worden de documenten aangeleverd voor het scannen? Zitten er verschillende formaten door elkaar? Zitten er verschillende documenttypes door elkaar? Als er verschillende documenttypes door elkaar zitten, is het dan mogelijk die te scheiden?
- Is het mogelijk gebruik te maken van reeds in de organisatie aanwezige informatie? Is het bijvoorbeeld mogelijk om een unieke persoonlijke 'identificer' te ontdekken zoals bijvoorbeeld een polisnummer of een soft-nummer. Zo ja, dan kan die worden gebruikt om automatisch persoonlijke gegevens als naam en adres te vergelijken met de informatie die reeds aanwezig is in de database van de organisatie. Hierdoor neemt de nauwkeurigheid sterk toe.
- In het kader van digitale duurzaamheid is het van belang voor het opslaan van de scans alleen gebruik te maken van breed gedragen bestandsformaten als TIFF, JPEG, JPEG 2000 of PDF-A.³⁶

³⁶ *Analysis, Selection, and Implementation Guidelines* - pag 13-16

3.3. De tweede component van ECM: Manage



De manage-component draait om de vraag ‘What are the tools and techniques for moving content around an organization and monitoring those tools performance?’ De manage-component omvat hiervoor de informatiebeheerssystemen en vormt de ruggengraat van ECM. Vanwege het grote belang van deze component zal ik hier uitgebreider op ingaan dan op de andere componenten. Ik zal achtereenvolgens document management-/record management, workflow, content management en collaborative software bespreken.

3.3.1. Document Management & Record Management

In de literatuur over digitaal werken en digitaal archiveren worden de termen Document Management System (DMS) en Record Management Applicatie (RMA) vaak naast en door elkaar gebruikt. De systemen groeien steeds dichterbij elkaar toe en daardoor worden de onderlinge verschillen steeds kleiner. Desondanks is het van belang om die verschillen duidelijk te maken. Grofweg komt het bij DMS en RMA neer op het onderscheid dat in de archivalische wereld wordt gemaakt tussen een *document* en een *archiefdocument*. Die termen worden overigens niet door iedereen op dezelfde wijze gedefinieerd. Ter illustratie, er worden in het rapport Record Management Terminologie maar liefst zestien verschillende definities gegeven van het woord ‘document’ en niet minder dan 24 van het woord ‘archiefbescheiden’.³⁷ In deze scriptie zal ik (tenzij anders aangegeven) de definities aanhouden zoals die in de ReMANO³⁸ worden gehanteerd. Document wordt hierin gedefinieerd als ‘betekenisvolle, samenhangende groep gegevens die als eenheid een functie vervullen ter overdracht van informatie, vastlegging van een toestand of recht, of ter verandering van een rechtstoestand’.

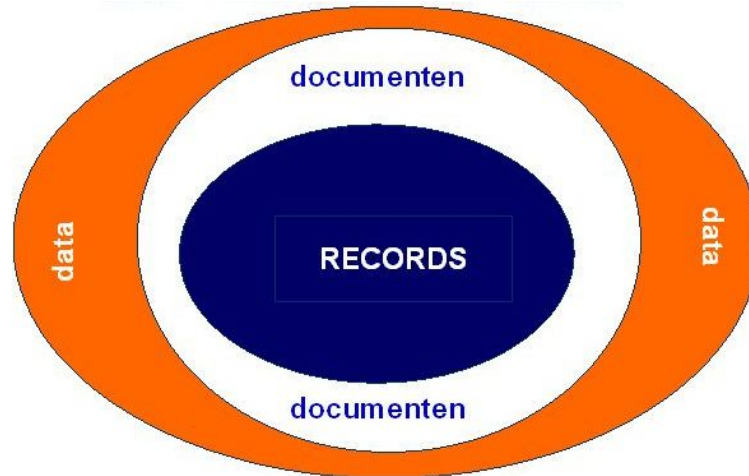
³⁷ Giesbers, Saskia - *Record Management Terminologie, Resultaten van een onderzoek in opdracht van de Record management Conventie* - RMC bureau (2004); pag. 19-23, 42-44

http://www.rmconventie.nl/uploads/RecordsManagementTerminologiev4juli2004_1.pdf

³⁸ Bussel, G.J.; Horsman P.J.; Waalwijk H. - *Softwarespecificaties voor Record Management Applicaties voor de Nederlandse Overheid 2004 (ReMANO 2004)* - Archiefschool, Amsterdam (2004); pag. 17-22

http://www.archiefschool.nl/docs/ReMANO_2004.pdf

Een archiefdocument (engelse term hiervoor is record) wordt als volgt gedefinieerd: ‘een document dat is ontvangen of gemaakt en gebruikt bij de uitvoering van processen, taken en activiteiten. De archiefwettelijke term is ‘archiefbescheiden’. In de ReMANo wordt de term archiefdocument gehanteerd om te onderstrepen dat archiefbescheiden documenten zijn. Simpel gesteld: een archiefdocument is een document, maar niet ieder document is een archiefdocument. Een document wordt pas een archiefdocument als het wordt ontvangen of gemaakt en gebruikt bij de uitvoering van processen, taken en activiteiten.³⁹ De relatie tussen data, documenten en archiefbescheiden (records) kan als volgt worden weergegeven:



Figuur 9: relatie data, documenten & records

Definitie DMS & RMA

Het onderscheid tussen typen documenten komt ook terug in de softwaresystemen die zijn ontwikkeld voor het beheer van deze documenten. Zo wordt een document management systeem in de ReMANo gedefinieerd als: ‘Software, specifiek ter ondersteuning van het genereren, opslaan en opvragen van documenten’.⁴⁰

Het verschil dat in de ReMANo wordt gemaakt tussen een DMS en een RMA heeft betrekking op het type documenten dat ermee wordt beheerd. Een record management applicatie wordt gedefinieerd als software, specifiek ter ondersteuning van het beheer van archiefdocumenten.⁴¹ De hier gegeven definities van een DMS en RMA zijn erg abstract en behoeven mogelijk nog enige verdere uitleg.

Document Management Systeem

Een DMS is bedoeld voor de ondersteuning van werkprocessen. Het is een systeem voor het digitaal managen van de creatie, vastlegging, raadpleging en opslag van elektronische documenten. Het kan hierbij gaan om zowel Word-, Excel- en PDF-bestanden als om gescande formulieren, facturen en foto’s. Het DMS kan ingezet worden om het creëren van documenten in een organisatie te kanaliseren. Documenten worden dan in één omgeving opgeslagen, in plaats van op verschillende locaties, zoals de lokale C:-schijf en de gedeelde netwerkschijven. Door het DMS te integreren met de kantoorautomatisering (bijvoorbeeld MS Office) biedt het de gebruiker de mogelijkheid om zorgvuldig met zijn (in bewerking zijnde) documenten om te gaan. Enkele functionaliteiten die in vrijwel alle document management systemen zijn opgenomen, zijn gecontroleerde toegang tot documenten, check-in/check-out en versiebeheer.

³⁹ Bussel, G.J.; e.a - *Softwarespecificaties voor Record Management Applicaties voor de Nederlandse Overheid* - pag. 17-22

⁴⁰ Ibidem, pag. 19

⁴¹ Ibidem, pag. 21

Gecontroleerde toegang heeft tot doel de informatie in brede zin te beveiligen. Hiermee wordt bepaald wie het recht heeft een document te lezen, te creëren, aan te passen of te vernietigen. De toegang tot documenten wordt bepaald op basis van 'rollen' die vooraf zijn gedefinieerd. Zo kan men bijvoorbeeld bepalen dat auteurs het recht hebben om hun eigen stukken te creëren, wijzigen en vernietigen. Redacteuren mogen alleen documenten wijzigen. Anderen mogen de documenten slechts lezen en hebben niet de mogelijkheid om wijzigingen aan te brengen of het document te vernietigen.

Check-in/check-out is de functionaliteit in het systeem waarmee wordt geregeld dat, alvorens een auteur gaat werken aan een document, dit eerst wordt 'outgecheckt' en als hij er mee klaar is weer 'ingecheckt'. Deze procedure garandeert dat slecht één persoon tegelijk kan werken aan een document. Zouden er twee personen bijvoorbeeld gelijktijdig wijzigingen kunnen aanbrengen in een document, dan zou er een conflict ontstaan over welke versie de juiste is.

Nauw samenhangend met check-in/check-out is versiebeheer. Deze functionaliteit zorgt ervoor dat iedere keer dat een document wordt 'ingecheckt', er een nieuwe versie (kopie) van wordt gemaakt die een nieuw versienummer krijgt. Het achterliggende idee van versiebeheer is dat in documenten de vorm, structuur en inhoud van de opgenomen data vastliggen. En om die reden is het niet de bedoeling dat daarin wijzigingen worden aangebracht. Gebeurt dit wel, dan verandert het document en ontstaat er in principe een nieuw document. Bij wijziging wordt er dan ook een nieuwe versie van het document gecreëerd en opgeslagen. Door de verschillende versies op te slaan is iedere verandering traceerbaar.⁴²

Documenten worden door het DMS opgeslagen, versies worden bijgehouden en een deel van de document management systemen biedt tevens de mogelijkheid documenten rond te sturen. Aan documenten kunnen metadata worden gekoppeld die het mogelijk maken relaties tussen documenten te leggen (dossiervorming) en opgeslagen documenten op een later moment weer op te vragen. Een DMS richt zich in eerste instantie op de creatiefase van een document.

Voor het beheer van documenten op de lange termijn is een DMS minder geschikt. De auteur heeft bijvoorbeeld de mogelijkheid inhoud en metadata te wijzigen en zelfs documenten te verwijderen uit het systeem. Deze mogelijkheden zijn in de fase waarin de documenten gecreëerd worden erg nuttig, maar wanneer die in de definitieve versie in het elektronische archief zijn opgenomen, moet alleen een daartoe geautoriseerde persoon iets kunnen wijzigen of verwijderen. Dus ondanks dat een DMS (archief)documenten creëert en ook gedurende een bepaalde periode opslaat, is een DMS niet direct geschikt voor langdurige archivering.⁴³

Record Management Applicatie

Veel documenten die in het DMS worden opgeslagen, spelen een rol in werkprocessen. Omdat een DMS zich met name richt op de ondersteuning van die processen, blijft de archiveerfunctie (bewijsvoering) veelal onderbelicht. Dit hiaat wordt opgevuld door de Record Management Applicatie (RMA). Om documenten geordend en toegankelijk te houden en de authenticiteit ervan te garanderen moet het beheersregime worden aangepast. Bij de opslag moeten hiervoor aan de documenten metadata worden toegekend, die bijvoorbeeld betrekking kunnen hebben op afzonderlijke documenten, groepen documenten, werkprocessen waaruit documenten voortkomen, maar ook op de hard- en software die nodig is om het document te kunnen lezen.

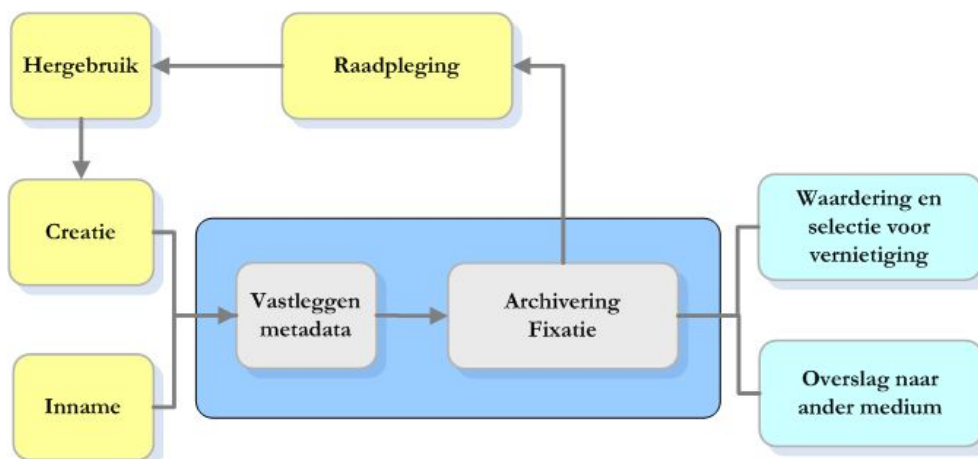
⁴² Rockley, Ann; Kostur, Pamela; Manning, Steve - *Managing enterprise content, a unified content strategy* - Berkeley (2003); 316-317

Bussel, G.J. - *'Fixed content', context en SOX* - VBDS (2005) http://www.vbds.nl/index.php?option=com_content&task=view&id=88&Itemid=63

⁴³ Doekes, M. - *RMA of DMS: cruciaal nuanceverschil* - Digital Display http://digidiv.amsterdam.nl/handleidingen_en_instrumenten/documentaire_informatiesystemen/kenmerken_en_cases/rmad.doc

Essentieel voor een RMA is dat de koppeling tussen documenten en de daarbij behorende metadata volledig en betrouwbaar is.⁴⁴

Een RMA legt de nadruk op archivalische metadata en heeft als functie die documenten die erin thuishoren, in het archiefsysteem op te nemen. Tevens zorgt de RMA ervoor dat de documenten adequaat worden bewaard, ontsloten en beschikbaar gesteld aan degenen die daartoe geautoriseerd zijn. Verder draagt de RMA er zorg voor dat documenten uit het archiefsysteem worden verwijderd op het moment dat ze geen belang meer dienen voor de organisatie. Door de functionaliteiten van een RMA wordt zij vaak gelijkgesteld met een archiefsysteem, dit is echter niet correct. Het verschil wordt duidelijk wanneer men een conceptueel onderscheid maakt tussen RMA en archiefsysteem. Een archiefsysteem is 'het geheel van processen, documenten, gegevens, methoden, mensen en middelen (waaronder de Record Management Applicatie) waarmee een organisatie haar archieffunctie vorm geeft.'⁴⁵ Een RMA is dus slechts een onderdeel van een archiefsysteem, het is de software die voor opname, verwijdering, beschikbaarstelling en bewaring van documenten zorgt. Hoewel RMA's over het algemeen beschikken over de functionaliteit om documenten op te slaan, kan dit in principe ook buiten de applicatie gebeuren. De opslag van de betreffende archiefdocumenten vindt dan plaats in een specifieke database. De RMA bestaat in dat geval enkel uit de verwerkingslogica. Dit zijn de regels volgens welke de RMA opereert en de metadata die worden gebruikt.⁴⁶ Schematisch ziet een RMA er als volgt uit:



Figuur 10: schematische weergave RMA⁴⁷

Het schema laat zich als volgt lezen: bij een organisatie komen documenten binnen (inname) of worden documenten aangemaakt (creatie). Van alle documenten die de neerslag vormen van bepaalde processen, transacties of handelingen, wordt de context vastgelegd in een metadatasysteem. Deze metadata bestaan uit gegevens die automatisch of handmatig uit het document worden gehaald (bijvoorbeeld naam, adres en woonplaats van de auteur, datum van ontvangst/creatie) en die eraan worden toegekend (bijvoorbeeld het bedrijfsproces waaraan het is gekoppeld, de bewaartermijn en de personen die geautoriseerd zijn om de gegevens te bekijken). Vervolgens wordt het document opgenomen in het archiefdeel en daar zodanig gefixeerd dat vorm en inhoud onveranderlijk vastliggen. Hiermee is het document een record (archiefdocument) geworden. We noemen dit document voor het gemak doc.1. Is het record gearchiveerd, dan is het raadpleegbaar voor een ieder die daartoe is geautoriseerd. Een document dat eenmaal is opgenomen in het RMA kan dus niet meer worden gewijzigd.

⁴⁴ Testbed Digitale Bewaring - *Functionele specificaties bewaarsysteem (versie1.0)* - Den Haag (2005); pag. 19 http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Functionele_specs_bewaarsysteem_v1.pdf

⁴⁵ Bussel, G.J.; e.a - *Softwarespecificaties voor Record Management Applicaties* - pag. 18

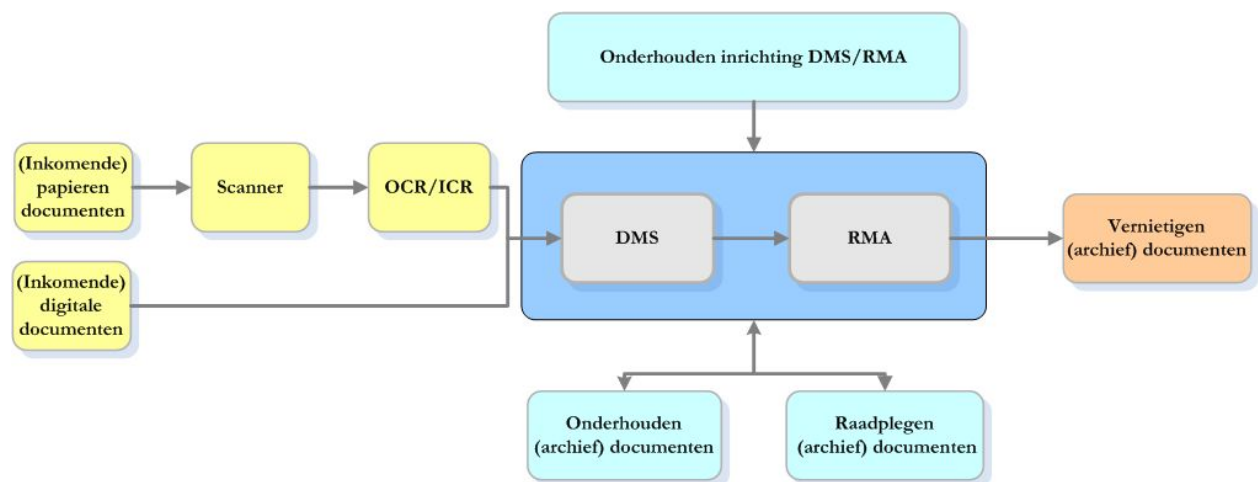
⁴⁶ Ibidem, 16-18

⁴⁷ Syllabus VHIC faculty - *Leergang aankomend records manager* - VHIC pag. 62

Raadpleging van records (bijvoorbeeld doc.1) kan ertoe leiden dat een document wordt hergebruikt en dat er een nieuwe versie van wordt gecreëerd (doc.2). Iedere keer dat een nieuwe versie of een nieuw document in het systeem wordt opgenomen, wordt daarvan de context vastgelegd in een metadatasysteem. Vervolgens kan het archiefdocument worden gefixeerd en gearchiveerd. Doc.2 zal dus nooit doc.1 vervangen. De records die in het systeem zijn opgenomen, hoeven daar natuurlijk niet voor eeuwig te blijven. Archiefdocumenten die geen waarde meer hebben, moeten uit het systeem worden verwijderd.⁴⁸

Combinatie DMS/ RMA

Zoals ik hierboven al heb aangegeven, hebben een DMS en een RMA ieder een specifieke rol binnen de informatiehuishouding. Een informatiehuishouding waarin gebruik wordt gemaakt van zowel een DMS als een RMA, zou er schematisch als volgt uitzien:



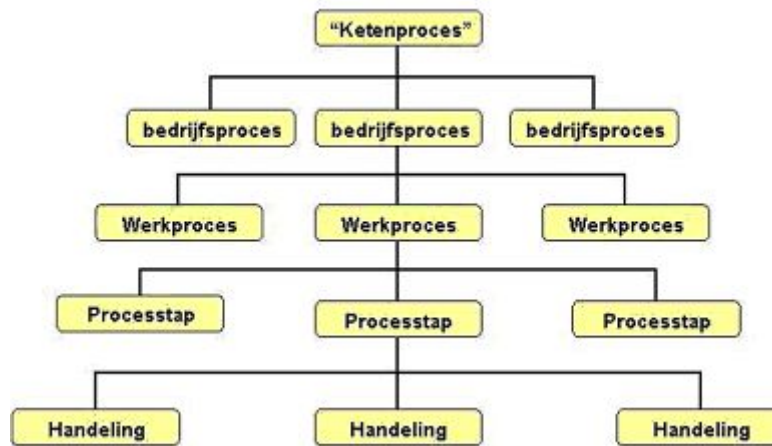
Figuur 11: DMS & RMA

3.3.2. Workflow Management & Business Process Modelling

Workflow, of workflow management (WfM) is een overkoepelend begrip voor het geheel van afspraken en technische middelen waarmee processen geautomatiseerd kunnen worden ondersteund en gestuurd. Bij workflow management worden documenten, informatie of taken van de ene betrokkene naar de andere doorgegeven volgens vooraf gedefinieerde procedures en regels. Essentieel hierbij is dat gerealiseerd wordt dat workflow in principe alleen mogelijk is als de betrokken processen zijn gestandaardiseerd. Door processen in samenhangende componenten op te delen, ontstaat een belangrijk methodisch kader voor het standaardiseren van processen. Het voorbeeld hieronder geeft een beeld van de decompositie van processen.⁴⁹

⁴⁸ Ibidem, pag. 61-63

⁴⁹ Cate ten, Timo; Groeneweg, Ronald - *Workflow en documentflow, Over mensen, organisaties, processen en documenten* - Digital Display, Deventer (2007); pag. 16 <http://www.ddisplay.nl/beheer/UserFiles/File/Seminar/Seminar%202006/workflow%20en%20documentflow.pdf>



Figuur 12: hiërarchische opbouw procesarchitectuur

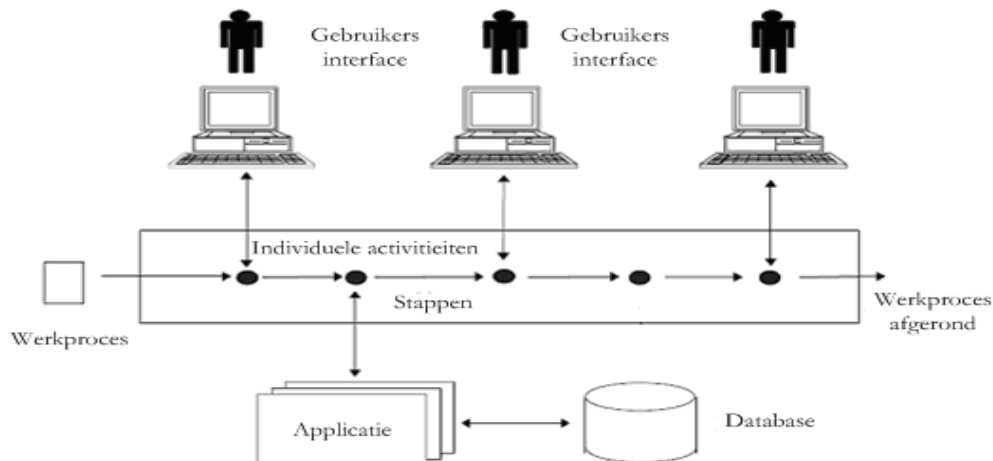
Door een werkproces te beschrijven wordt het mogelijk om het proces vorm te geven, te modelleren en op grafische wijze weer te geven. Een procesbeschrijving dient minimaal die informatie te bevatten die het workflow managementsysteem nodig heeft om te kunnen functioneren. “This includes information about its starting and completion conditions, constituent activities and rules for navigation between them, user tasks to be undertaken, references to applications which may to be invoked, definition of any workflow relevant data which may need to be referenced”.⁵⁰

Workflow wordt gebruikt om processen te formaliseren. Nadat een werkproces met behulp van een workflow-definitie is vastgelegd, kunnen gebruikers de definitie steeds opnieuw gebruiken bij de uitvoering van hun werksproces. Workflow is met name handig voor het bewaken van termijnen, zoals afhandelingstermijnen of expiratietermijnen. Door workflow wordt grip gekregen op het procesverloop en wordt het mogelijk te sturen op de stappen die in een proces moeten worden genomen. Daarnaast kan workflow worden ingezet voor het routeren en verdelen van werkvoorraden. Het kan dan gaan om eenvoudige lineaire routing, maar ook om complexe processen met parallelle stromen.⁵¹ De informatie die door middel van de procesbeschrijvingen aan het workflow-systeem is toegekend zorgt dat het systeem in staat is gegevens bij te houden over werkvoorraden, overschrijding van de gewenste doorlooptijden en productiviteit. Voor een goed begrip van procesarchitectuur is het van belang dat er eenduidigheid bestaat over de gebruikte termen (zie voor een overzicht bijlage 1).

⁵⁰ Hollingsworth, David - *The Workflow Reference Model* - Workflow Management Coalition (1995); pag. 12
<http://www.wfmc.org/standards/docs/tc003v11.pdf>

⁵¹ Workflows kenmerken zich door de volgende patronen:

- Serieel (activiteiten volgen elkaar op)
- Parallel (activiteiten worden tegelijkertijd uitgevoerd)
- Een combinatie van serieel en parallel (sommige activiteiten volgen elkaar op, terwijl andere gelijktijdig worden uitgevoerd)
- Cyclisch (wanneer het nodig is terug te kunnen gaan in een proces)



Figuur 13: schematische weergave workflow-systeem

Een workflow managementsysteem beschikt over het algemeen over de volgende vier functies:

- Routing: controle op de distributie van documenten, informatie of taken van gebruiker tot gebruiker.
- Wachtijdbeheer: hiermee wordt de prioriteitstelling van de taken geregeld en de status van het werk weergegeven. Mocht het noodzakelijk zijn, dan kunnen wachtrijen worden vrijgegeven en taken worden toegewezen aan andere gebruikers met als doel de doorlooptijd van het werkproces niet onnodig te verlengen.
- Statistieken: op basis van statistieken die door het workflow-systeem worden geproduceerd, kan het management inzicht krijgen in het verloop van de werkprocessen. Dit heeft betrekking op de nog uit te voeren taken, doorlooptijden en productiviteit.
- Simulatietool: instrument dat de impact van wijzigingsvoorstellen op werkprocessen in een virtuele omgeving analyseert.⁵²

Een workflow managementsysteem is dus een informatiesysteem dat op geautomatiseerde wijze de besturing, beheersing en bewaking van werkstromen uitvoert. Door (een deel van) de workflow te automatiseren worden gebruikers verplicht om te werken volgens de vooraf opgestelde procedures die het werkproces beschrijven. In het workflow-systeem kunnen bijvoorbeeld precondities (voorwaarden) worden vastgelegd. Een preconditie kan zijn dat in het werkproces processtap 2 alleen mag plaatsvinden als processtap 1 volledig is afgerond en document x is gecreëerd. De precondities van een proces kunnen de procesgang automatisch opstarten, onderbreken of beëindigen.⁵³

Van belang is om te realiseren dat ook allerlei stappen kunnen worden gedigitaliseerd zonder dat er documenten aan vastzitten. Dat is het essentiële verschil tussen workflow en documentflow.

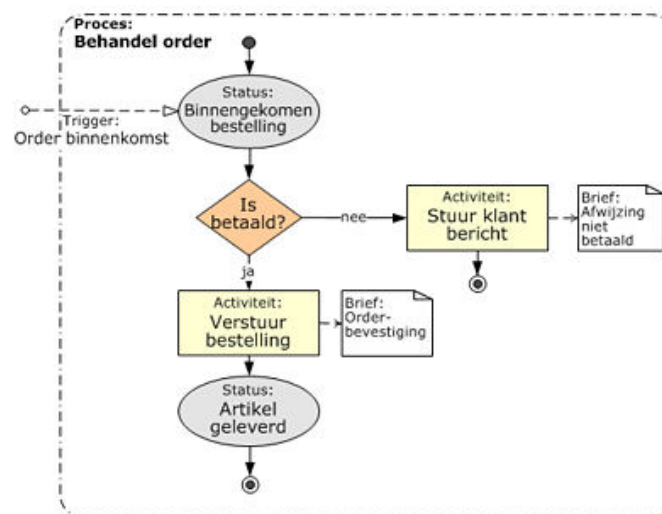
⁵² Damen, J.; Leeuwen, S. van - *Document beheer in de praktijk: Trends, voordelen en voorwaarden bij een document management software implementatie* - Management Kennisbank Vakspecialisten (2006); pag. 8 <http://rubeo.eu/files/Document%20beheer%20in%20de%20praktijk.pdf>

⁵³ Er kan binnen een workflow systeem op verschillende manieren met precondities worden omgegaan:

- Geen controle op precondities van processtappen. Buiten het systeem om vindt de controle op precondities plaats zodat binnen het systeem het proces altijd door kan gaan.
- Menselijke controle op precondities (checklist als document). Buiten het systeem om vindt de controle op precondities plaats dit wordt binnen het systeem aangegeven. Echter binnen het systeem kan het proces altijd doorgaan.
- Het systeem controleert op precondities. Alle voorwaarden worden als metagegevens aan het proces gekoppeld, het systeem controleert deze gegevens alvorens een vervolgstap, al dan niet automatisch, wordt gestart.

De laatste gaat altijd in eerste instantie over het sturen van documenten. De een creëert een document, stuurt het naar een ander, die doet er iets mee, net zolang tot de behandeling helemaal af is.⁵⁴

Business Process Modelling (BPM) en workflow zijn nauw met elkaar verbonden. Zo resulteert het digitaliseringstraject van processen vaak in een kritische beschouwing van het betreffende proces. Hiermee komen we op het terrein van Business Process Modelling. BPM is een methode om de werkprocessen in kaart te brengen. Door zowel de situatie van het moment als de gewenste situatie schematisch te representeren wordt het mogelijk het proces te analyseren en te verbeteren. De modelmatige beschrijvingen van werkprocessen kunnen op meerdere detailniveaus opgesteld worden. Eerst kan het globale proces met de relevante stappen op hoofdniveau beschreven worden. Vervolgens kunnen de subprocessen tot in detail worden uitgewerkt. Voorbeeld van een modelmatige representatie van een werkproces:



Figuur 14: modelmatige representatie van een werkproces

Werkprocessen beschrijven en modelleren kan tijdrovend zijn en daardoor een kostbare activiteit. Maar door administratieve handelingen en doorlooptijd te beperken, kunnen veel kosten worden bespaard. Ook kan BPM een kwalitatief betere aansturing, beter toezicht en minder fouten en overschrijdingen van deadlines opleveren.

3.3.3. Content Management

Een Content Management Systeem (CMS) is een applicatie waarmee beheer, presentatie en publicatie van zowel digitale als analoge content worden geautomatiseerd. Onder het begrip CMS valt weer een breed scala aan applicaties. Zo bestaan er eenvoudige oplossingen voor het beheer van relatief eenvoudige websites, maar ook zeer geavanceerde systemen voor complexe toepassingen. Eigenlijk is het gemeenschappelijke doel van alle CMS-en contentbeheer.

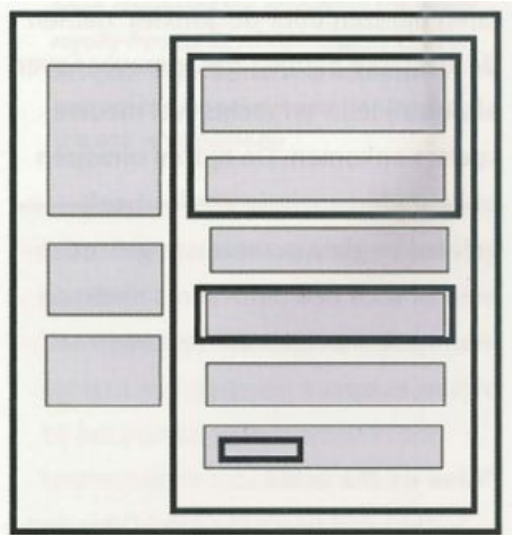
⁵⁴Cate ten, Timo; Groeneweg, Ronald - *Workflow en documentflow, Over mensen, organisaties, processen en documenten* - pag. 62

In deze scriptie ga ik uit van de volgende definitie van een CMS: ‘Systeem om het content management (automatisch) te beheren. Een CMS voldoet minimaal aan de volgende drie eigenschappen:

- het is een geautomatiseerd publicatiesysteem;
- het maakt een scheiding tussen content, opmaak en structuur (en rollen);
- het biedt de mogelijkheid om dynamisch informatie te kunnen publiceren uit een repository van herbruikbare content-elementen.⁵⁵

Op het eerste gezicht vertonen content management systemen veel gelijkenissen met document management systemen. Beide systemen ondersteunen de administratieve organisatie van creatie, gebruik en opslag van ongestructureerde informatie. Daarnaast spelen bij zowel een CMS als een DMS metadata een cruciale rol bij het toegankelijk maken en houden van informatie. Het verschil tussen de twee is dat document management systemen werken op het niveau van documenten en content management systemen een niveau lager opereren, op het content niveau.

Een document is zoals hierboven gedefinieerd ‘een betekenisvolle, samenhangende groep gegevens die als eenheid een functie vervullen ter overdracht van informatie⁵⁶, content is het niveau onder het document, hetgeen waaruit een document is opgebouwd. Content is een ‘informatie-item binnen een bepaalde context, bijvoorbeeld tekst, afbeeldingen, geluid, video en animaties.⁵⁷ Anders gezegd, een document is opgebouwd uit verschillende content elementen die gezamenlijk het document vormen. Het grote verschil tussen een content management systeem en document management systeem is dan ook de mate van granulariteit. Hiermee bedoel ik de mogelijkheid om documenten op te delen in losse blokken. In tegenstelling tot een DMS is een CMS is staat om documenten op te delen in individuele informatie-elementen als hoofdstukken, alinea’s en zelfs woorden en getallen.⁵⁸



Figuur 15: CMS granulariteit

‘Content is a compromise between usefulness of data and the richness of information. Content is rich information that you wrap in simple data. The data that surround the information (meta data) is a simplified version of the context and meaning of the information⁵⁹

Er kunnen verschillende categorieën content worden onderscheiden. Content met specifieke kenmerken wordt wel een *content-type* genoemd. Voorbeelden hiervan zijn rapporten, formulieren en brochures. Een content-type kan weer bestaan uit verschillende *items*. Als een content-type bijvoorbeeld een brochure is, maar er worden verschillende brochures uitgegeven, dan zijn de

⁵⁵ http://content.hartman-communicatie.nl/Over_CMS/begrippenlijst.html

⁵⁶ Bussel, G.J.; e.a - *Softwarespecificaties voor Record Management Applicaties* - pag. 17-22

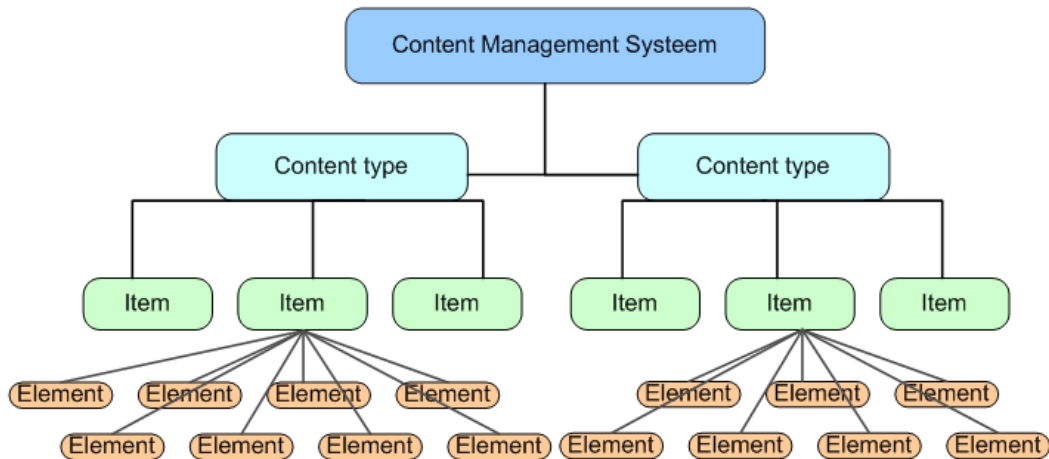
⁵⁷ http://content.hartman-communicatie.nl/Over_CMS/begrippenlijst.html

⁵⁸ Hartman, Erik M. - *Content management: een groeimarkt* - In: VIP vakblad voor documentmanagement, nummer 3 jaarboek (2004) pag. 14 – 17, alhier pag. 15

http://content.hartman-communicatie.nl/assets/binaries/Documenten/cm_ een_groeimarkt_x_hartman.pdf

⁵⁹ Boiko, Bob - *Content Management Bible (2nd edition)* - Indianapolis (2005); pag. 12

individuele brochures content-items. Ieder content item is op te delen in *elementen*. Elementen kunnen op verschillende niveaus worden onderscheiden. Zo kunnen hoofdstukken, paragrafen, zinnen en zelfs woorden als elementen worden gedefinieerd. Content heeft niet alleen betrekking op tekst, ook plaatjes kunnen als elementen worden gedefinieerd. Zo kan een logo een element zijn van een formulier of brief.⁶⁰ Schematisch ziet de relatie tussen CMS, content-type, content item en content-element er als volgt uit:



Figuur 16: opbouw van content management systeem

Hergebruik van content

Door documenten niet integraal op te slaan, maar ze in content-elementen op te delen, wordt het mogelijk de content-elementen te hergebruiken bij de creatie van nieuwe documenten. Het hergebruiken van content is op zich niet nieuw. Op dit moment gebeurt dat voornamelijk door bij het creëren van nieuwe documenten stukken tekst uit bestaande documenten te ‘knippen’ en ‘plakken’. Dit lijkt een goede manier, maar er ontstaan problemen op het moment dat de hergebruikte content moet worden aangepast. Dan kan het een tijdrovende opdracht worden om alle plaatsen te achterhalen waar de specifieke content is hergebruikt en de content daar aan te passen. Ook bestaat het risico dat niet alle plaatsen worden gevonden waar de content is hergebruikt, of dat plaatsen over het hoofd worden gezien. Dit heeft tot gevolg dat het content-element in twee verschillende vormen bestaat (de oude versie en de aangepaste versie) en dat documenten hierdoor inconsistent of zelfs inaccuraat worden. Daarbij is het mogelijk dat zowel het verouderde content-element als het vernieuwde element bij het ‘knippen en plakken’ op een later moment opnieuw worden hergebruikt, waardoor de inconsistentie zich opstapelt en er uiteindelijk twee verschillende content-elementen ontstaan.⁶¹

Hergebruik van content in een CMS werkt dan ook volgens een ander principe. Een CMS werkt volgens het proces van ‘linken’ van content. Content-elementen worden eenmalig opgeslagen in een repository. Wordt een element hergebruikt, dan wordt de content wel weergegeven in het nieuwe document, maar dat is dan slechts een link naar de originele content in de repository die het document weergeeft. Deze manier van werken is vergelijkbaar met het invoegen van een plaatje in Microsoft Word (Insert>Picture>From file>Insert and Link). Voordeel is dat een content-element dat in de repository wordt aangepast, automatisch wordt aangepast op alle plaatsen waar het betreffende element door hergebruik voorkomt.⁶²

In een CMS worden inhoud, vormgeving en structuur van content los van elkaar gecreëerd en opgeslagen. Zo kan dezelfde content worden hergebruikt voor verschillende publicaties met deels

⁶⁰ Boiko, Bob - *Content Management Bible* - pag. 22-23

⁶¹ Rockley, Ann; e.a. - *Managing enterprise content, a unified content strategy* - pag. 24

⁶² Ibidem, pag. 24

verschillende inhoud en verschillende vormgeving. Zo verschillen bijvoorbeeld de inhoud, structuur en vormgeving van een brochure en een persbericht maar kan in beide publicaties gebruik worden gemaakt van (gedeeltelijk) dezelfde content-elementen. Er kunnen twee soorten van content hergebruik worden onderscheiden, namelijk opportunistisch en systematisch hergebruik.

We spreken van *opportunistisch hergebruik* op het moment dat een auteur er bewust voor kiest een bestaand content-element op te zoeken en te hergebruiken. Voorwaarden voor opportunistisch hergebruik zijn dat de auteur op de hoogte is van de mogelijkheid tot hergebruik en tevens bereid is op zoek te gaan naar content die herbruikbaar is. Deze werkwijze verschaft de auteur optimale flexibiliteit en vrijheid. De auteur zélf bepaalt of content hergebruikt wordt en zo ja, welke content daarvoor geschikt is. Tegelijkertijd vormt dit een nadeel, omdat de zoeklasten volledig bij de auteur liggen.

Systematisch hergebruik houdt in dat content hergebruik automatisch plaatsvindt. Hierbij wordt specifieke content als herbruikbaar geïdentificeerd en door het CMS automatisch op de juiste locatie in een document geplaatst. Systematisch hergebruik gebeurt op basis van vooraf gedefinieerde gedetailleerde informatiemodellen. Door een content-type, zoals een rapport, telkens volgens hetzelfde vooraf gedefinieerde informatiemodel op te stellen wordt het mogelijk om elementen automatisch te hergebruiken. Content-typen definiëren met de daarbij behorende informatiemodellen vormt daarmee de basis voor het principe van systematisch content hergebruik. Bij systematisch hergebruik is het niet de auteur maar het systeem dat (op basis van informatiemodellen) bepaalt welke content op welke locatie wordt hergebruikt.⁶³

Content aanpassen of vervangen

Content kan na verloop van tijd zijn relevantie verliezen, waardoor het nodig is deze aan te passen of te vervangen. Het kan dan gaan om de vormgeving, de inhoudelijke informatie, de structuur of een combinatie daarvan. Content kan in bestaande documenten of publicaties (gecreëerd in een CMS) op drie manieren worden aangepast of vervangen:

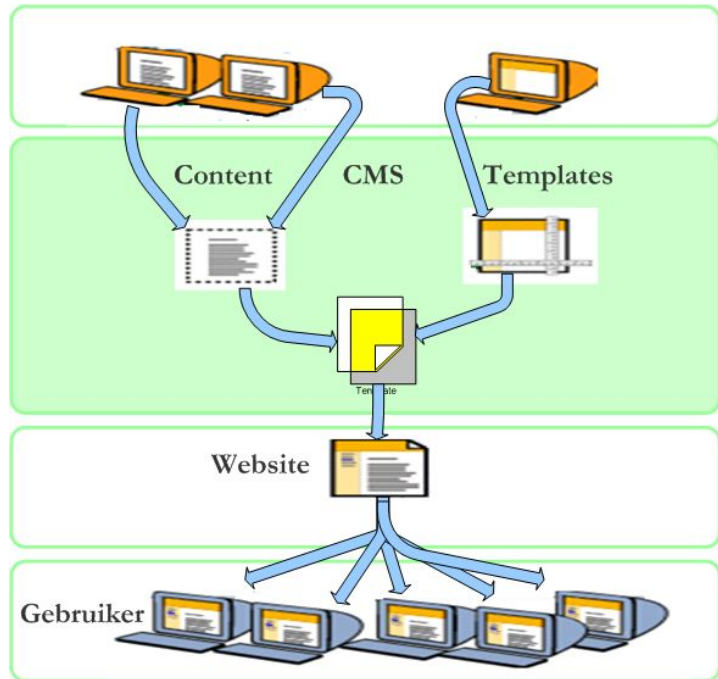
- *Automatisch aanpassen*: dit houdt in dat hergebruikte content automatisch wordt aangepast (op alle plaatsen waar deze zich bevindt) op het moment dat de 'originele' content wordt veranderd. Het aanpassen van hergebruikte content gebeurt zonder dat de auteur, die gebruik heeft gemaakt van de 'originele' content, daarvan op de hoogte wordt gesteld. Voordeel is dat de auteur geen tijd kwijt is aan het veranderen van content. Nadeel kan zijn dat de verandering niet relevant of correct hoeft te zijn op alle plaatsen waar de content is hergebruikt. Automatische verandering van hergebruikte content kan dus leiden tot foutieve informatie.
- *Optioneel aanpassen*: hierbij wordt de auteur genotificeerd op het moment dat content die hij hergebruikt, wordt veranderd. De auteur kan vervolgens bepalen of de verandering relevant is en of deze verandering dus geaccepteerd of geweigerd moet worden.
- *Niet aanpassen*: als de originele content wordt veranderd, verandert de hergebruikte content niet mee en hebben auteurs ook niet in de gaten dat de originele content is gewijzigd. Aangezien de hergebruikte content niet wordt aangepast, ontstaan er van elkaar verschillende content elementen (vergelijkbaar met de 'knippen en plakken'-problematiek). Om die reden is deze optie dan ook niet aan te raden.⁶⁴

Een CMS dient niet alleen om content te hergebruiken maar ook om deze te kunnen gebruiken in verschillende publicatievormen, zoals brochures, websites en papieren publicaties. Om dezelfde content in verschillende publicatievormen te kunnen gebruiken is het nodig content (inhoud) en

⁶³ Ibidem, pag. 24-42

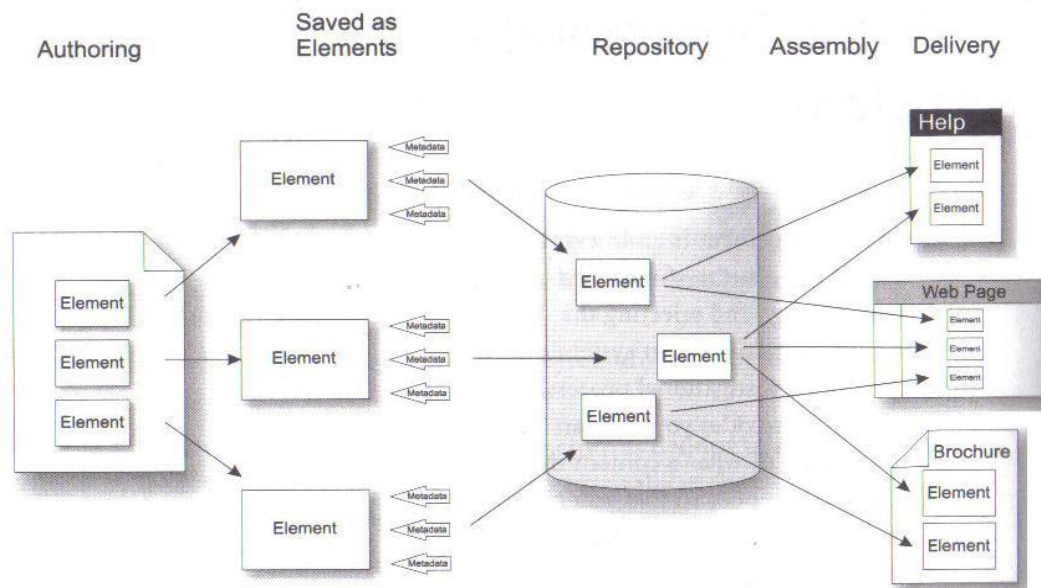
⁶⁴ Ibidem, pag. 319

opmaak (de manier waarop de content aan de gebruiker wordt gepresenteerd) van elkaar te scheiden. Scheiding tussen enerzijds de content en anderzijds de presentatie van de content is dan ook een van de belangrijke kenmerken van een CMS. De opmaak van de site of publicatie wordt in een CMS in *sjablonen* of *templates* vastgelegd, los van de contentinhoud. Per content-type en per publicatiekanaal (internet, intranet of papier) worden een of meerdere templates gedefinieerd. Het kunnen meerdere zijn omdat een content-type gepubliceerd kan worden in meerdere vormen, die elk een eigen vormgeving kennen, bijvoorbeeld HTML, XML en PDF. Het voordeel van templates is dat de content die via het CMS wordt aangeleverd, door het gebruik ervan automatisch op de juiste plaats en in de juiste vorm op de site wordt gepubliceerd. Hoewel een CMS een complexe applicatie is, valt haar basisprincipe redelijk simpel weer te geven.



Figuur 17: CMS, scheiding content en opmaak

Content wordt gecreëerd in een 'authoring tool' en vervolgens als individuele elementen opgeslagen in het CMS (repository). Hierbij zijn twee zaken van belang. Ten eerste wordt de content opgedeeld in elementen van een dusdanig formaat dat ze voldoen aan de eisen die contenthergebruik stellen. Ten tweede moeten aan de elementen metadata worden gekoppeld die het mogelijk maken de elementen terug te vinden en te hergebruiken. In de repository worden de elementen opgeslagen en, wanneer deze worden hergebruikt, naar de betreffende media gestuurd.⁶⁵ Schematisch ziet dit er als volgt uit:



Figuur 18: werking van content management systeem⁶⁶

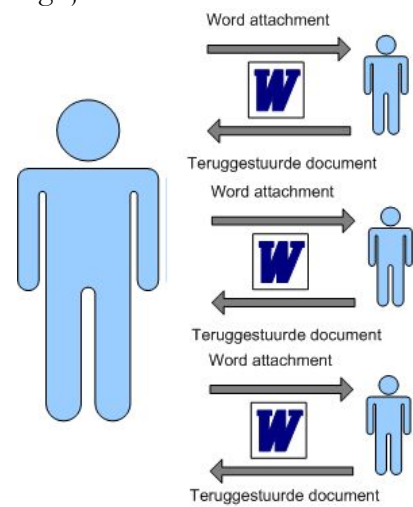
⁶⁵ Ibidem, pag. 312

⁶⁶ Bron: Ibidem, pag. 312

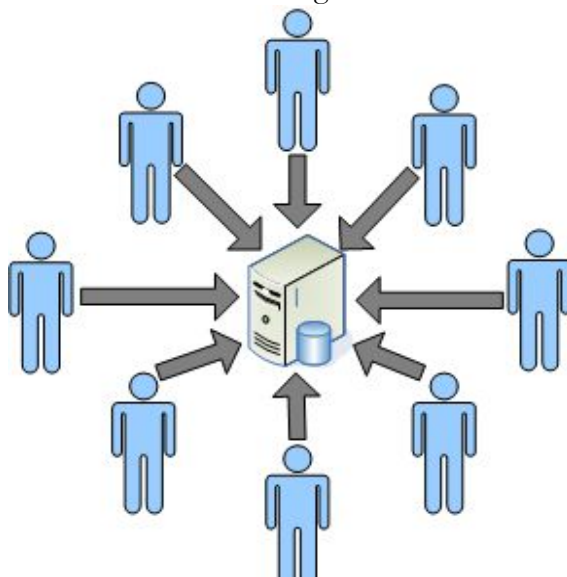
3.3.4. Collaborative Software

Wat ik in deze scriptie collaborative software noem, wordt ook wel aangeduid als *Group Support Systems*, *Groupware* of *Team Collaborative Applications*. Hoe men het ook noemt, het gaat om “software die de significantie van groepen binnen kantoren erkent door het aanbieden van functies en diensten die de samenwerkende activiteiten van werkgroepen bevorderen.”⁶⁷

Gebruikmakend van de interactieve mogelijkheden van het internet/netwerk heeft collaborative software tot doel de interactiviteit tussen deelnemers aan een proces te verhogen. Hierbij staan twee gedachten centraal. De eerste is dat kennis delen de basis vormt voor het gebruik van de software. De tweede is dat de locatie waar de gebruikers zich bevinden geen belemmering mag zijn voor deelname aan het proces. *Collaboration tools* bieden mensen, waar ze zich ook bevinden, de mogelijkheid om gezamenlijk aan projecten te werken en ideeën uit te wisselen. De tools om dit te bereiken variëren van virtuele werkruimtes en software voor gezamenlijke brainstormsessies tot berichtenservice (messaging) en conferentieapplicaties. Een belangrijk verschil tussen collaboration tools en ad hoc samenwerking via bijvoorbeeld e-mail is dat collaboration tools uitgaan van op repository gebaseerde samenwerking. Ad hoc samenwerking is over het algemeen zeer inefficiënt, omdat samenwerking via de email in de basis ongeordend en ongeorganiseerd is. Laten we uitgaan van een scenario waarin een auteur een document als attachment meestuur met als doel dit door andere teamleden van commentaar te laten voorzien. Het is mogelijk om een document aan meerdere teamleden tegelijk te sturen. Er ontstaan echter problemen op het moment dat meerdere teamleden aanpassingen maken en het document vervolgens terugsturen. Afhankelijk van de groepsomvang kan de auteur een groot aantal versies van het document (ieder met andere aanpassingen) teruggestuurd krijgen. Het samenvoegen van alle aanpassingen kan een gecompliceerde en tijdrovende taak zijn. Veel efficiënter is het om gebruik te maken van collaborative



Figuur 19: ongeordende samenwerking



Figuur 20: repository gebaseerde samenwerking

software die uitgaat van op repository gebaseerde samenwerking. Uitgangspunt hierbij is dat documenten niet naar individuele teamleden worden verstuurd maar dat ze wordt opgeslagen in een repository, waarna de teamleden er toegang toe krijgen. Een document in de repository kan vervolgens door de teamleden worden beoordeeld en waar nodig worden aangepast. Door het document op één plaats op te slaan en toegankelijk te maken voor alle teamleden wordt voorkomen dat er meerdere versies van een document ontstaan die op een later moment moeten worden samengevoegd. Ook wordt zo voorkomen dat verwarring ontstaat doordat er meerdere versies van een document in omloop zijn en niet duidelijk is wat de meest recente versie is.⁶⁸

⁶⁷ Laudon, Kenneth C.; Laudon, Jane P. - *Management Information Systems, New Approaches to Organization and Technology* - Vijfde editie, (1998)

⁶⁸ Jenkins, Tom - *Enterprise Content Management Technology* - Ontario (2006) pag. 170-171

Voorbeelden van collaboration tools:

Chat verzorgt online en real time (zonder vertraging) tekstuele communicatie. De chatfunctionaliteit maakt gebruik van een eenvoudige teksteditor die het mogelijk maakt dat mensen direct (tekstueel) met elkaar kunnen communiceren. Na beëindiging kan de chatsessie worden opgeslagen in een database zodat zij op een later moment kan worden gereproduceerd.

Whiteboarding maakt het mogelijk om real time tekeningen, schetsen en handschriften uit te wisselen. Het is een schoolbord dat voor alle partijen zichtbaar is op een scherm, waarop iedereen kan meeschrijven.

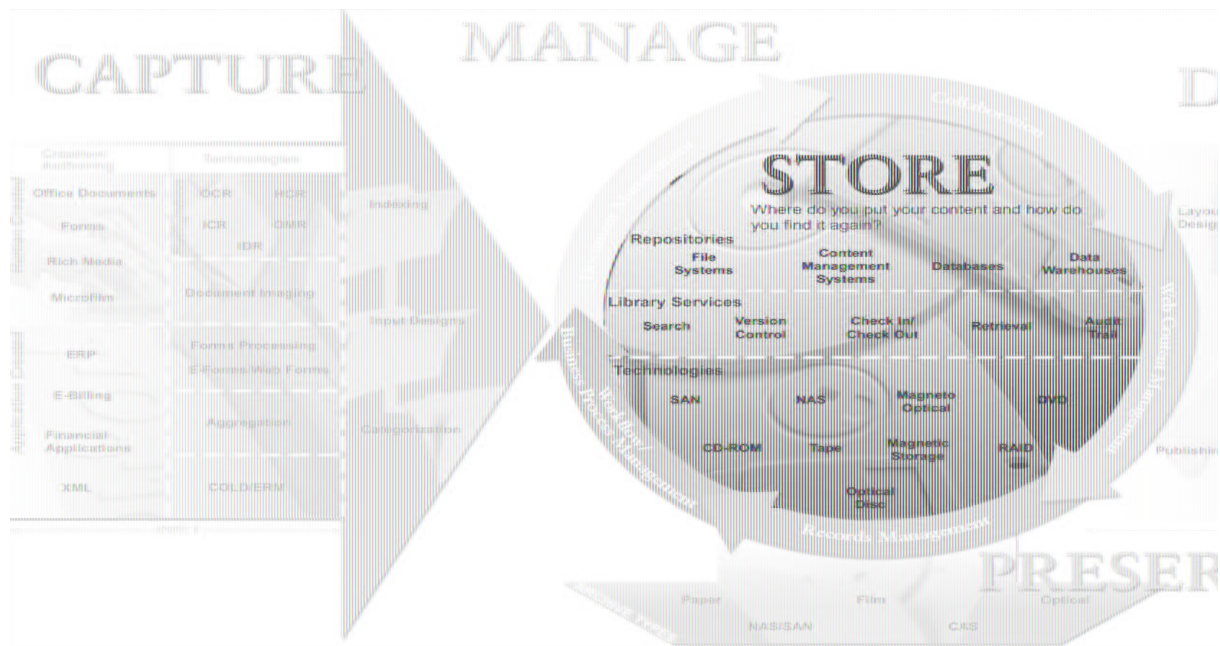
Een *wiki* is een website waarop bezoekers zelf op een eenvoudige manier informatie kunnen toevoegen of aanpassen. Om content aan te kunnen passen is geen autorisatie of toegangscode nodig.

Video conferencing maakt het mogelijk om op afstand realtime te communiceren via beeld en geluid. Let wel: bij video conferencing kan de beschikbare bandbreedte van het netwerk een beperkende factor zijn.

Messaging is berichtenverkeer dat lijkt op e-mail. Belangrijkste verschil is dat de message box alleen berichten bevat die met het werkproces te maken hebben waarnaast de messagefunctionaliteit geplaatst is. Het grote voordeel van messaging is dat de belangrijke berichten over een specifiek werkproces daar te vinden zijn waar ze relevant zijn.⁶⁹

⁶⁹ Meester, G. J. - *Collaboration Tools* - (2002)
<http://www.betterbe.com/docs/Colabtools.pdf>

3.4. De derde component: Store



Bij de Store-component draait het om de vraag 'Where do you put your content and how do you find it again?' De store-component, zoals door de AIIM is gedefinieerd, kan hiervoor in drie categorieën worden onderverdeeld:

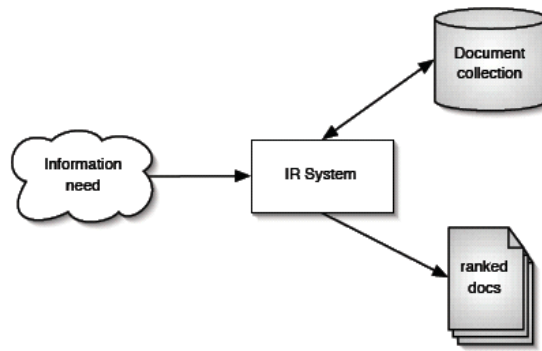
1. De repositories die als opslaglocatie voor de informatie fungeren. Het gaat hier o.a. om filesystemen, databases en datawarehouses. Omdat ik de repositories en hun functies in hoofdstuk 4 bespreek, ga ik er hier niet verder op in.
2. Opslagtechnologieën als SAN (storage area network), NAS (network attached storage), DVD en optische disks. Dit zijn technologieën die wel onderdeel uitmaken van de store-component, maar ook betrekking hebben op de preserve-component. Om die reden zal ik ze daar bespreken.
3. De 'library services', functionaliteiten die fungeren als administratieve componenten en de toegang tot de informatie in de repositories regelen. Library services hebben o.a. betrekking op functionaliteiten als versiebeheer, check-in/check-out en zoekmogelijkheden. Zowel versiebeheer als check-in/check-out zijn in de bespreking van de manage-component al aan bod gekomen. Bij de beschrijving van de store-component richt ik me daarom volledig op de zoekmogelijkheden.⁷⁰

3.4.1. Information retrieval

Om digitaal opgeslagen informatie binnen een bedrijfsdomein na oplag weer te kunnen lokaliseren en hergebruiken zal deze moeten worden ontsloten. Dit kan door middel van een vooraf gedefinieerde ordening, een zoekmachine of een combinatie daarvan.

⁷⁰ Kampffmeyer, Ulrich - *ECM enterprise content management* - pag. 54- 57

In een vereenvoudigde vorm komt information retrieval neer op het volgende: er is een informatievraag en er is een collectie documenten. Het information retrieval systeem heeft tot doel de informatievraag te beantwoorden door de meest relevante documenten uit de collectie te presenteren. De meest bekende information retrieval systemen als Google en Yahoo zijn ondanks hun populariteit om verschillende redenen



Figuur 21: concept information retrieval systeem

niet direct geschikt voor gebruik binnen bedrijven. Dit type zoekmachine is op zich makkelijk in het gebruik maar het kan frustrerend zijn wanneer de ingevoerde zoekterm 20.000 zoekresultaten oplevert - wat niet ondenkbaar is binnen een grote organisatie. Internetzoekmachines zijn met name geschikt voor 'ad hoc' zoekacties. Een voorbeeld daarvan is de vraag 'Wie ontdekte Amerika?'. Uit welk document (op het internet) het antwoord op de informatievraag komt, maakt voor de gebruiker niet uit zolang het antwoord maar correct is.

In een bedrijfsomgeving is het echter van belang dat de vraag wordt beantwoord door informatie uit een vooraf bekend document, zoals een bepaalde factuur, een ingekomen brief of een rapport. Dit type zoekactie wordt een *known item search* genoemd. De gebruiker weet dat een specifiek document bestaat en wil dan ook alleen dat document en de juiste versie ervan vinden. Daarbij komt dat er bij de 'google' manier van zoeken naar informatie van uitgegaan wordt dat de betekenis van een document volledig en uitsluitend kan worden afgeleid uit het document zelf. Deze veronderstelling wordt het *book on hand*-principe genoemd. Het is echter zeer de vraag of dit wel het geval is.

“Eenvoudig gezegd komt dit neer op het nobele streven na te gaan (en te representeren) wat de auteur heeft bedoeld. Het komt me echter voor dat de betekenis van een document mede wordt bepaald door het doel waarvoor het document wordt gebruikt, c.q. de kennisbehoefte op grond waarvan het document wordt gezocht. In de praktijk heeft het daarom meer zin om na te gaan wat de gebruiker beoogt, en het document te interpreteren op basis van zijn of haar vocabulaire en wereldbeeld. De analyse en representatie van de betekenis van een document moet daarom steeds contextueel zijn, letterlijk 'in termen van' de organisatie, het individu of zelfs de individuele vraagstelling. Dit betekent dat (automatische) indexing gebaseerd moet zijn op kennis van de organisatie, en van de probleemcontext van het individu op het moment dat de zoekvraag wordt gesteld. Dit stelt twee voorwaarden aan het retrievalssysteem, namelijk dat het in staat is zich een beeld te vormen van de probleemcontext van de gebruiker, en dat het in staat is om, in 'real time', de betekenis van documenten te analyseren in relatie tot die probleemcontext.”⁷¹

De documenten die worden opgesteld of ontvangen binnen een bedrijfsomgeving, zijn geen losstaande entiteiten maar zijn opgesteld/ontvangen met een doel. Zij maken deel uit van een werkproces. De document moet dan ook in deze context⁷² worden geïnterpreteerd. De ideale

⁷¹ Mackenzie Owen, John - *Kennis in veelvoud* - Rede uitgesproken bij de aanvaarding van het ambt van hoogleeraar in de Documentaire Informatiewetenschap aan de Universiteit van Amsterdam op woensdag 24 juni 1998
<http://igitur-archive.library.uu.nl/DARLIN/2005-0130-130700/ORATIE02.pdf>

⁷² Context: 'Het geheel van relevante entiteiten in de omgeving van archiefbescheiden die de betekenis bepaalt die aan archiefbescheiden moet worden toegekend. Voorbeelden van entiteiten zijn: werkprocessen, actoren,

situatie zoals hierboven geschetst is nog niet bereikt, maar een goede manier om context te creëren is documenten te archiveren op basis van het werkproces waar ze uit voortkomen. Dit wordt ook wel zaaksgewijs archiveren genoemd.⁷³ Hierbij worden documenten op een zodanige manier opgeslagen dat ze direct herkenbaar zijn als een samenhangende groep, die een bepaald voorval van een proces ondersteunt. Met andere woorden, de context waarbinnen de documenten zijn opgemaakt of ontvangen wordt hiermee vastgelegd.

Een bijkomend probleem van internetzoekmachines is dat deze ervan uitgaan dat alle informatie die niet is afgeschermd, openbaar is. Dit werkt goed in een omgeving waarin alle informatie vrij toegankelijk is, maar in een bedrijfsomgeving is dit veelal niet het geval. In een situatie waarin het voor een werknemer mogelijk is om zonder restricties binnen alle in het bedrijf aanwezige informatie te zoeken bestaat het risico dat personen die daar niet toe zijn geautoriseerd geheime of beperkt openbare informatie inzien.

Om dit te voorkomen is het nodig de toegang tot die informatie te structureren. Zoekmachines binnen het domein van een bedrijf functioneren dan ook anders dan die op het internet.⁷⁴ Er kunnen grofweg twee mogelijkheden worden onderscheiden die gebruikt worden voor het zoeken naar documenten binnen de databases van bedrijven, namelijk de *ordeningsstructuur* en het *(metadata)searchmodel*. Hierbij moet men zich ervan bewust zijn dat de modellen elkaar aanvullen en dus naast elkaar kunnen worden gebruikt, waarbij *fulltext search* in beide modellen tot de mogelijkheid hoort.

3.4.2. Ordeningsstructuur

Een ordeningsstructuur⁷⁵ bootst als het ware digitaal een fysiek archief in een slimme mappenstructuur na. De ordeningsstructuur wordt primair gebruikt om documenten eenvoudig terugvindbaar te maken en de rechten⁷⁶ binnen de documenthuishouding te verdelen. In deze structuur worden dossiers (zaken) geplaatst. Een ordeningsstructuur is alleen consistent wanneer daarop een bepaald stramien van toepassing is. Hierbij zijn verschillende invalshoeken mogelijk:

- ordening volgens het organisatieonderdeel;
- ordening volgens het onderwerp;
- ordening volgens de werkprocessen.

medewerkers, informatiesystemen.' Uit *Softwarespecificaties voor Record Management Applicaties voor de Nederlandse Overheid (ReMANO)* 2004.

⁷³ Een zaak wordt meestal gedefinieerd als een in de tijd begrensd complex handelingen betreffende een bepaald geval. Het geheel van stukken, ontvangen of opgemaakt bij de behandeling van één zaak heet dan ook de zaaksgewijze ordening of het dossierstelsel. Uit: Giesbers, Saskia - *Record Management Terminologie* - pag. 47-48

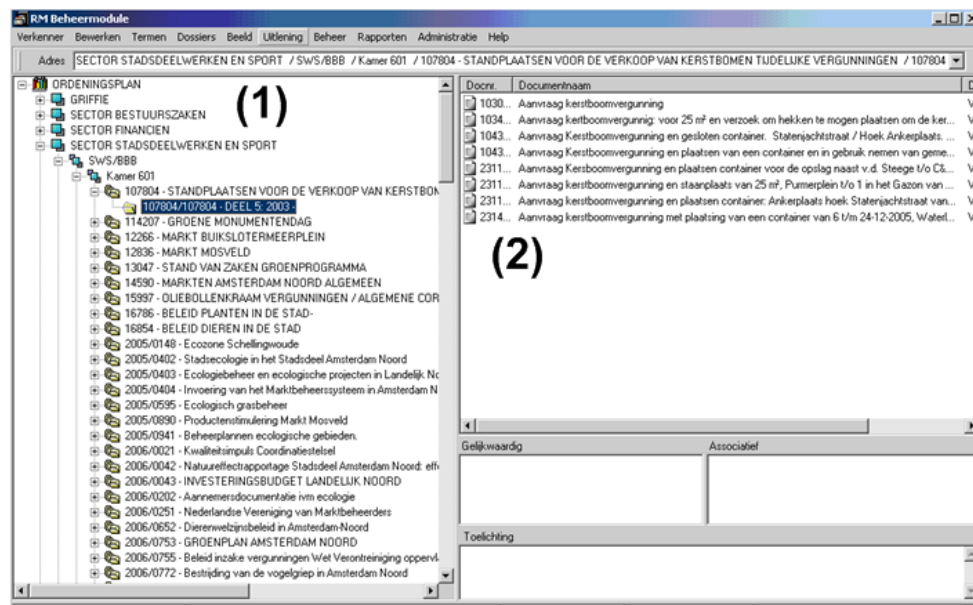
⁷⁴ Jenkins, Tom - *Enterprise Content Management Technology* - pag. 169-170

⁷⁵ In de discussie rond ordening wil nog er wel eens verwarring bestaan over de begrippen ordeningsmodel en ordeningsstructuur. In dit verband is het belangrijk het verschil tussen beide begrippen expliciet te maken. Het *ordeningsmodel* is de verzameling (minimale set van) van uitgangspunten, richtlijnen, metadata en relaties op basis waarvan elke structuur kan worden opgebouwd. Een model kan daarom generiek worden ontwikkeld. Een *ordeningsstructuur* is de visualisering van het ordeningsmodel binnen een specifieke context. Een structuur is afhankelijk van specifieke wensen en omstandigheden en kan daarom niet generiek worden ontwikkeld.

⁷⁶ Zo kan de inrichting bijvoorbeeld zo zijn dat de rechten die aan gebruikers zijn verleend voor een bepaald gedeelte van de structuur (mappen) door overerving gelden voor alle onderliggende subdossiers en documenten.

Bij digitale archivering verdient het de voorkeur dit op basis van het bedrijfsproces te doen.⁷⁷ Processen zijn, in tegenstelling tot organisatieonderdelen en onderwerpen, niet (zo sterk) aan verandering onderhevig en bieden hierdoor meer stabiliteit.

Hieronder laat ik zien hoe een ordeningsstructuur eruit kan zien. Aan de linkerkant (1) staat een ordeningstructuur die vergelijkbaar is met de mappenstructuur van de Windows verkenner. Door in de ordeningsstructuur op een map/dossiers te klikken wordt aan de rechterkant (2) getoond welke documenten zich in de map/het dossier bevinden.



Figuur 22: ordeningsstructuur

3.4.3. Fulltext retrieval

Het surfen op het web heeft de verwachtingen die gebruikers hebben bij het gebruik van informatie retrieval systemen sterk beïnvloed. Het gaat hierbij om de verwachting hoe de systemen eruit zien, zich gedragen en hoe de interactie tussen systeem en gebruiker verloopt. Een typisch voorbeeld van een functionaliteit van een traditioneel informatiesysteem is het gebruik van een gecontroleerd vocabulaire.

Op het internet wordt geen gebruik gemaakt van een gecontroleerd vocabulaire, wat tot gevolg heeft dat het gebruik van gecontroleerd vocabulaire als moeilijk en tegennatuurlijk wordt ervaren.

⁷⁷ Voordelen ordeningsstructuur op basis van werkprocessen:

- Door de procesmatige structurering is het eenvoudig om documenten snel te zoeken, te vinden en op te slaan. De organisatie gebruikt één systeem, waardoor iedereen weet waar hij of zij aan toe is en waar de documenten geplaatst moeten worden. De ordeningsstructuur biedt de mogelijkheid om het document 'top-down' te vinden: door de ordening te doorlopen wordt de correcte locatie en daarmee ook het document gevonden.
- Functioneert persoonsonafhankelijk en organisatieafhankelijk: de documenten zijn voor iedere betrokkene toegankelijk en overzichtelijk gerangschikt. Medewerkers en bedrijfsonderdelen kunnen de structuur niet naar eigen voorkeur en op eigen initiatief wijzigen. De ordeningsstructuur zorgt ervoor dat de locatie van het document adequaat, consequent en consistent gekozen wordt.
- Maakt selectiebeleid mogelijk: door de aan het document gekoppelde metadata, de bestandslocatie of een combinatie van beide zijn bewaar- en vernietigtermijnen direct na opslag bekend. De metadata kunnen door de ordeningsstructuur aan het document toegevoegd worden.
- Makkelijk te beheren en stabiel: de werkprocessen bepalen de plaatsing van de documenten, en die processen veranderen niet snel in de tijd. Als er organieke of personele wijzigingen plaatsvinden, dan zal dat geen invloed hebben op de ordeningsstructuur.

Of het nu gaat om het internet of information retrieval binnen organisaties, de gedachte is vaak dat fulltext retrieval veel handiger, sneller en nauwkeuriger is. Echter het gebruik van gecontroleerd vocabulaire is in het verleden juist ontwikkeld om de taalkloof tussen de gebruiker en het systeem te overbruggen en daarmee de gebruiker tot dienst te zijn.

Het belang van een gecontroleerd vocabulaire is te illustreren aan de hand van een lijst van in totaal meer dan 600 verschillende manieren waarop de zoekterm *Britney Spears* is ingevoerd door gebruikers van Google.com bij het zoeken naar informatie over *Britney Spears*.⁷⁸

Bij 75% van de zoekopdrachten werd de zoekterm correct ingevuld. Dit lijkt best veel, maar betekent wel dat dus 25% van de zoekopdrachten mislukt. Hierbij dient rekening te worden gehouden met het feit dat de gebruikers bij weinig of geen zoekresultaten de conclusie trekken dat er dus geen informatie beschikbaar is, en niet dat het geringe aantal treffers mogelijk het gevolg is van foutieve spelling.⁷⁹

Om zoekmachines binnen een bedrijfsdomein goed te laten functioneren is enig begrip van hoe zoekmachines werken van belang. Veel zoekmachines maken gebruik van *fulltext search*. Vaak wordt gedacht dat de computer hiervoor real time door de tekst zoekt tot de zoekterm is gevonden. Dit is niet het geval; om fulltext te kunnen zoeken wordt er door de computer eerst een index van het document gemaakt; een zogenoemde *inverted file*. Een inverted file is enigszins vergelijkbaar met een boekindex. Het belangrijkste verschil tussen een boekindex en een inverted file is de omvang. De eerste bestaat uit een beperkt aantal trefwoorden, de tweede uit (bijna⁸⁰) alle woorden uit de tekst. Een inverted file is dus een document waarin het originele document op woordniveau uiteengegrafeld en alfabetisch gerangschikt is. Wanneer een document fulltext wordt doorzocht, wordt niet het document zelf doorzocht, maar wordt de inverted file van het document doorzocht.⁸¹

Aan de hand van de onderstaande tabel is het makkelijker uit te leggen hoe een inverted file er uitziet en wat de voordelen ervan zijn.⁸² Om het overzichtelijk te houden is ervoor gekozen ieder document slechts uit één zin te laten bestaan (hoe groot het document namelijk ook is, het principe blijft hetzelfde).

```

488941 britney spears
40134 brittany spears
36315 brittney spears
24342 britany spears
7331 britny spears
6633 briteny spears
2696 britteny spears
1807 briney spears
1635 brittney spears
1479 brintey spears
1479 britanny spears
1338 britiny spears
1211 britnet spears
1096 britiney spears
991 britaney spears
991 britnay spears
811 brithney spears
811 brtiney spears
664 birtney spears
664 brintney spears
664 briteney spears
601 bitney spears
601 brinty spears
544 brittaney spears
544 brittnay spears
    
```

Figuur 23: zoekterm Britney Spears

Document	Tekst
1	Pease porridge hot, pease porridge cold
2	Pease porridge in the pot
3	Nine days old
4	Some like it hot, some like it cold

⁷⁸ Dit zijn de 25 meest voorkomende spellingen voor de complete lijst zie bijlage 2

⁷⁹ Griffiths, J. R.; Brophy, P. - *Students searching behaviour and the web* - in *Library Trends* 53(4) 2005: 539-554; alhier pag. 541

⁸⁰ Bijna de hele tekst wordt geïndexeerd, de meeste softwareprogramma's indexeren veel voorkomende woorden als *en*, *de* en *het* niet omdat deze zo veelvuldig in ieder document voorkomen dat deze om die reden onbruikbaar zijn als zoekterm

⁸¹ Jenkins, Tom - *Enterprise Content Management Technology* - pag. 59-60

⁸² Voorbeelden van inverted file ontleend aan presentatie tijdens college 'Full Text Information Retrieval' door Avi Arampatzis & Jaap Kamps. UVA Documentaire Informatiewetenschap Semester I, 2006/2007

5	Some like it in the pot
6	Nine days old

Om de documenten fulltext te kunnen doorzoeken wordt er een inverted file van gecreëerd. Die kan op verschillende manieren worden samengesteld. Een manier is om per woord aan te geven in welk document het betreffende woord voorkomt. In het onderstaande voorbeeld van een inverted file wordt duidelijk dat het woord ‘old’ in document 3 en 6 voorkomt. Als die zoekterm wordt ingevoerd, dan worden documenten 3 en 6 als relevante zoekresultaten gepresenteerd.

Originele documenten

Doc.	Tekst
1	Pease porridge hot, pease porridge cold
2	Pease porridge in the pot
3	Nine days old
4	Some like it hot, some like it cold
5	Some like it in the pot
6	Nine days old



Inverted file

Nummer	Tekst	Document
1	cold	1, 4
2	days	3, 6
3	hot	1, 4
4	in	2, 5
5	it	4, 5
6	like	4, 5
7	nine	3, 6
8	old	3, 6
9	pease	1, 2
10	porridge	1, 2
11	pot	2, 5
12	some	4, 5
13	the	2, 5

Figuur 24: inverted file

Het is natuurlijk handig te weten *dat* een zoekterm in een document voorkomt, maar als het een groot document is dan is het ook van belang te weten *waar* de betreffende term zich in dat document bevindt. Een meer geavanceerdere manier is dan ook om niet alleen aan te geven dat een woord in een bepaald document voorkomt, maar ook op welke plaats in het document je het woord kunt vinden. Dit is bijzonder handig in het geval dat een gezochte term alleen voorkomt op de laatste pagina van een honderd pagina's tellend document. Uit het onderstaande voorbeeld blijkt dat het woord ‘old’ zowel in document 3 als in document 6 het derde woord is.

Originele documenten

Doc.	Tekst
1	Pease porridge hot, pease porridge cold
2	Pease porridge in the pot
3	Nine days old
4	Some like it hot, some like it cold
5	Some like it in the pot
6	Nine days old



Inverted file

Nummer	Tekst	Document en locatie woord
1	cold	(1; 6), (4; 8)
2	days	(3; 2), (6; 2)
3	hot	(1; 3), (4; 4)
4	in	(2; 3), (5; 4)
5	it	(4; 3, 7), (5; 3)
6	like	(4; 2, 6), (5; 2)
7	nine	(3; 1), (6; 1)
8	old	(3; 3), (6; 3)
9	pease	(1; 1, 4), (2; 1)
10	porridge	(1; 2, 5), (2; 2)
11	pot	(2; 5), (5; 6)
12	some	(4; 1, 5), (5; 1)
13	the	(2; 4), (5; 5)

Figuur 25: inverted file

3.4.4. Metadata search

Fulltext zoekmogelijkheden zijn zeer handig, maar de nauwkeurigheid van de zoekresultaten kan worden verbeterd door bij het zoeken ook gebruik te maken van metadata. Op zich is dit niet nieuw, ook in papieren archiefsystemen wordt gebruik gemaakt van metadata. Papieren documenten zijn voor een groot deel zelf drager van metadata. Deze kunnen bij de creatie van het document zijn opgenomen en betreffen dan bijvoorbeeld naam van de auteur, geadresseerde en datum. Maar ze kunnen ook op een later moment aan het document worden toegevoegd en dan gaat het veelal om zaken als een paraaf of een stempel. Metadata van digitale documenten en papieren documenten zijn in de meeste gevallen van hetzelfde soort, maar in papieren documenten zijn ze fysiek aanwezig en in digitale documenten niet.⁸³

'In the digital environment, authoritative records are those accompanied by metadata defining their critical characteristics. These characteristics must be explicitly documented rather than being implicit, as in some paper-based processes. In the digital environment, it is essential to ensure that the creation and capture of record management metadata are implemented in systems that create, capture and manage records. Conversely, the digital environment present new opportunities for defining and creating metadata and ensuring the complete, contemporaneous capture of records. These records can be evidence of transactions or themselves be transactions.'⁸⁴

Zoals ik eerder heb aangegeven, dient men bij digitale archivering bij voorkeur het bedrijfsproces als uitgangspunt te nemen. Een bedrijfsproces is een complex van activiteiten, die worden verricht om een taak te vervullen en die een informatieproduct opleveren. Een bedrijfsproces kan meerdere stappen omvatten die een of meerdere personen moeten uitvoeren. Elke stap of substap die genomen wordt, levert niet alleen een informatieproduct op maar ook een verzameling metadata. Die kunnen betrekking hebben op de volgende zaken: welk bedrijfsproces betreft het, om welke stap binnen het bedrijfsproces gaat het, op welke datum en op welk tijdstip heeft het plaatsgevonden en wie zijn erbij betrokken geweest. Door middel van deze metadata worden documenten met elkaar in verband gebracht, bijvoorbeeld omdat ze deel uitmaken van hetzelfde dossier. Het geheel aan informatie, dus zowel de inhoud van het informatieproduct als de bijbehorende metadata, moet worden vastgelegd.⁸⁵ Door gebruik te maken van metadata kan het gezochte document nauwkeuriger worden geïdentificeerd (en gevonden). Hoe meer metadata er aan een document worden gekoppeld, des te meer mogelijkheden zijn er om het document op een later moment weer te zoeken. Elke organisatie zal op basis van de eisen en wensen die aan de informatiehuishouding worden gesteld zelf moeten bepalen welke registratievelden (metadata) er worden gebruikt.

⁸³ Bussel, G.J.; e.a - *Softwarespecificaties voor Record Management Applicaties* - pag. 92

⁸⁴ *NEN-ISO 23081-1:2006 information and documentation – record management processes – Metadata for records – part 1: Principles* - Nederlands Normalisatie-instituut (2006); pag. 2

⁸⁵ Arnold, A.G.; Glashouwer, B - *De moderne informatiehuishouding van de digitale overheid* - pag. 27

De voordelen van het gebruik van metadata bij het ontsluiten van digitale informatie kunnen als volgt worden samengevat:

- Het metadatamodel zorgt ervoor dat er kenmerken aan een document worden toegekend, waardoor altijd correct achterhaald kan worden wat voor document het is, waar het voor dient en uit welk proces het afkomstig is.
- Het metadatamodel biedt de mogelijkheid om het document bottom up te vinden. Door concrete zoekcriteria te specificeren die overeenkomen met de toegekende metadata, wordt het document vindbaar.⁸⁶

Hieronder geef ik een voorbeeld van het zoekscherm van een DMS. Elk zoekveld is in principe niets anders dan een zoekmogelijkheid naar een specifiek metadata-element. Men kan hierbij zowel op één zoekveld als op een combinatie van verschillende zoekvelden zoeken.

Figuur 26: zoekscherm DMS

⁸⁶ *Notitie Zoekmogelijkheden, Ordeningsmodel en Documentbeheer* - ten behoeve van Werkgroep DMS door: Stadsdeel Amsterdam-Centrum, (2005); pag. 3-5
http://digidiv.amsterdam.nl/handleidingen_en_instrumenten/documentaire_informatiesystemen/kenmerken_en_casus/zoeken_ordenen_en_beheren.doc

Naast de reeds beschreven ontsluitingsmethode van digitale informatie zijn er nog vele andere ontwikkelingen gaande. Van enkele relevante geef ik een korte omschrijving:

Federated search: de mogelijkheid om met één zoekvraag gelijktijdig in verschillende repositories te zoeken. Een zoekactie vindt bijvoorbeeld plaats binnen de repository van het DMS, maar tegelijkertijd ook op het intranet of internet. Grote voordeel van federated search is dat de gebruiker niet langer in de zoekstructuur van elke individuele database hoeft te verdiepen.

Dynamic clustering: technieken waarmee overeenkomstige zoekresultaten automatisch worden verzameld en gegroepeerd.

Document summarization: techniek waarbij op basis van statische gegevens automatisch een samenvatting van een document wordt gemaakt.⁸⁷

⁸⁷Benneker, Frank - *Quickscan naar mogelijkheden voor automatische metadatageneratie* - Digitale Universiteit (2006); pag 6
http://www.darenet.nl/nl/page/repository.item/show?saharaIdentifier=ou_11:oar:dspace.ou.nl:1820/803
Jenkins, Tom - *Enterprise Content Management Technology* - pag. 66

3.5 De vierde component: Deliver



De deliver-component draait om de vraag ‘How do you get the right content to the right audience and the right device?’ De deliver-component is hiervoor op te delen in drie onderdelen: *transformation technologies*, *security technologies* en *distribution technologies*. Het gaat te ver om op elk van de onderdelen in te gaan, dus ik beperk me tot de kernactiviteiten van de deliver-component:

- Transformatie en conversie
- Distributie en outputmanagement
- Samenstellen van informatie
- Personalisatie
- XML

3.5.1. Transformatie & conversie

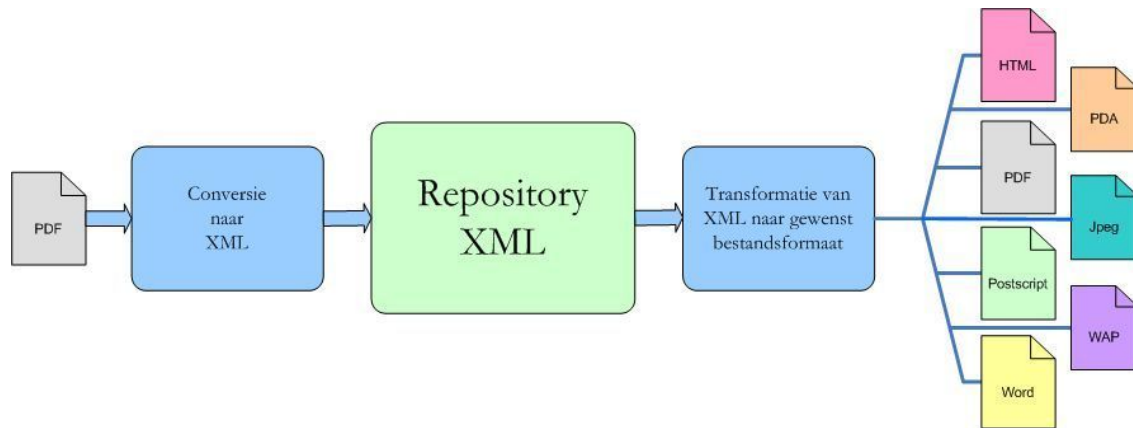
In het verleden was het zo dat informatie en de opmaak daarvan geschikt moesten worden gemaakt voor één specifiek publicatieformaat, bijvoorbeeld een gedrukt rapport. Tegenwoordig kunnen inhoud en vormgeving van elkaar worden gescheiden, zoals ik in de paragraaf over content management al heb aangegeven. Voordeel hiervan is dat content in verschillende formaten kan worden gepubliceerd zonder te hoeven worden aangepast.

Een van de kernactiviteiten van de deliver-component is de *transformatie* van content van het formaat waarin zij is opgeslagen naar het gewenste outputformaat, bijvoorbeeld PDF, PostScript⁸⁸, HTML, Microsoft Word of WAP.⁸⁹ Dit hangt nauw samen met het *conversie*proces. Conversie lijkt op transformatie, maar verschilt ervan omdat zij over het algemeen aan de

⁸⁸ PostScript is een pagina beschrijvingstaal (page description language), oftewel een programmeertaal om tekst en afbeeldingen op een apparaatonafhankelijke manier te beschrijven. Zo zou een PostScriptbestand in principe door iedere laserprinter hetzelfde geprint moeten worden.

⁸⁹ Formeel slaat de term WAP hoofdzakelijk op het protocol waarmee mobiele telefoons met het netwerk communiceren. Het aanbieden van content gaat via het HTTP-protocol en normale web servers, met als enige verschil dat de pagina's in WML of XHTML Mobile Profile zijn opgemaakt in plaats van in HTML.

inputkant plaatsvindt. Content wordt in een bepaald formaat aangeleverd en door het systeem geconverteerd naar een ander bestandsformaat voor opslag en verdere distributie. Een voorbeeld van dit proces is een document dat wordt aangeleverd in Word formaat en wordt geconverteerd naar XML. Nadat het document is geconverteerd in XML, kan de content door transformatie in verschillende formaten, zoals HTML of PDF, worden gepubliceerd.⁹⁰



Figuur 27: conversie & transformatie

3.5.2. Distributie & outputmanagement

Het is van belang informatie niet alleen bij de creatie maar gedurende de hele levenscyclus te beheren en te managen. Dit betekent dat informatie ook nadat zij als output door het deliverysysteem is afgeleverd, gemanaged moet worden. Als bijvoorbeeld een PDF-versie van een document wordt gecreëerd en gepubliceerd, dan moet er met behulp van outputmanagement voor gezorgd worden dat altijd alleen de complete, meest up-to-date versie van het document wordt gedistribueerd. Outputmanagement moet voorkomen dat er gelijktijdig verschillende versies van een document in omloop zijn. Idealiter wordt outputmanagement zoveel mogelijk automatisch gedaan door het deliverysysteem.⁹¹

3.5.3. Samenstellen van informatie

Zoals ik in paragraaf 3.3.3. heb gesteld, is het basisidee van een content management systeem dat informatie kan worden opgedeeld in losse elementen. De kern van het publicatiesysteem zijn de sjablonen, die de schakel vormen tussen de neutrale content in de repository en de behoefte aan vormgeving en structuur in de publicatieomgeving. Publicatiediensten zijn tools die het content management systeem ondersteunen bij de samenstelling van publicaties uit de losse content-elementen en metadata die zich in de repository bevinden. Naast de repository van het CMS kan het publicatiesysteem ook aan andere datasystemen gekoppeld zijn. Door de koppeling met externe gegevensbronnen kunnen data die zich niet in de repository van het CMS bevinden, toch gebruikt worden in publicaties. Met name voor publicatie op internet wordt gebruik gemaakt van de mogelijkheid om dynamisch informatie samen te stellen. De content wordt op basis van de zoekvraag samengesteld en aan de gebruiker gepresenteerd. Dynamische content is dus content die wordt samengesteld op het moment dat ernaar wordt gevraagd. De content die wordt getoond, bestaat voorafgaand aan de vraag niet als zodanig in de database.⁹²

⁹⁰ Rockley, Ann; e.a. - *Managing enterprise content* - pag. 353-356

⁹¹ Ibidem, pag. 356-357

⁹² Ibidem, pag. 357

3.5.4. Personalisatie

Dynamische content vormt de basis van *personalisatie*. Personalisatietechnieken maken het bijvoorbeeld mogelijk dat alleen informatie (content-elementen) die relevant is voor de gebruiker, wordt gepubliceerd op de portal.⁹³ Personalisatietechnieken zijn gebaseerd op metadata over de gebruiker. Grofweg zijn er twee typen personalisatie te onderscheiden, namelijk *expliciete* personalisatie en *impliciete* personalisatie. Die twee vormen verschillen van elkaar in de manier waarop de benodigde metadata worden gewonnen.

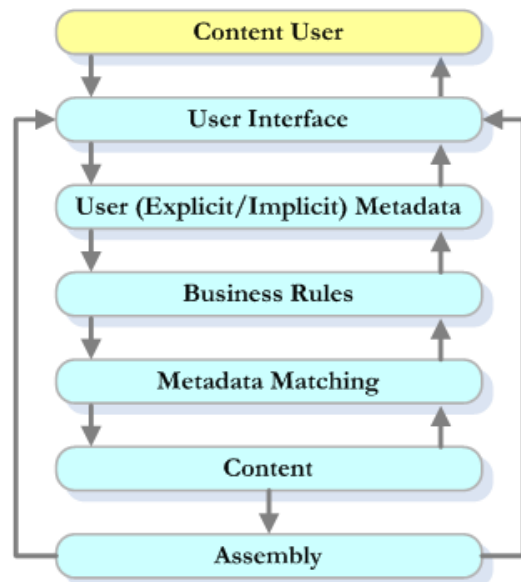
Expliciete personalisatie gaat uit van het actief toevoegen van personalisatiemetadata, bijvoorbeeld door middel van gebruikersprofielen, waarin per persoon of afdeling wordt vastgelegd wat voor de specifieke groep relevant is. Ook kan gebruikers de mogelijkheid worden geboden zelf aan te geven waarin zij geïnteresseerd zijn en wat voor hen relevant is.

Impliciete personalisatie is een techniek waarbij analytische software wordt ingezet om gebruikers relevante content te kunnen bieden.

Die software probeert als het ware te begrijpen welke artikelen de gebruiker leest, welke producten hij of zij koopt of welke secties van een website worden bezocht. Met deze informatie als vertrekpunt worden bij toekomstig sitebezoek de pagina's of artikelen gepresenteerd waarvan de software meent dat ze interessant zijn voor de gebruiker.

Personalisatie werkt over het algemeen als volgt. De personalisatiemetagegevens, waarbij het niet uitmaakt of die impliciet of expliciet zijn gewonnen, worden geïnterpreteerd op basis van de geldende *business rules*. Deze 'regels' bepalen welke content op welk moment geschikt is voor welke gebruiker en wie geautoriseerd is wat te zien. Na de interpretatie door middel van de business rules wordt een match gemaakt en de content 'uit het CMS gehaald'. Die content wordt vervolgens samengesteld en aan de

gebruiker getoond via de user interface.⁹⁴ Om content te transformeren en te personaliseren zijn twee elementen van groot belang; metadata en XML. Metadata heb ik al besproken en laat ik in hoofdstuk 4 nog uitgebreid aan de orde komen. Eerst zal ik ingaan op XML.



Figuur 28: personalisatie

3.5.5. eXtended Markup Language

Een van de grootste problemen bij digitale communicatie vormt de uitwisseling van gegevens tussen verschillende systemen. Een van de oplossingen hiervoor is afkomstig uit de bibliotheek wereld: SGML ofwel *Standard Generalized Markup Language*. Uitgangspunt van SGML is dat gegevens uiteindelijk in alle toepassingen dezelfde moeten zijn. De titel van een boek is overal dezelfde, of deze nu wordt bekeken in een browser, Microsoft Word of een specifieke bedrijfsapplicatie. Om die eenheid te bereiken worden er in SGML *tags* geplaatst om

⁹³ Een portal is een web-pagina die dienst doet als "toegangspoort" tot een reeks andere websites of specifieke content, die over hetzelfde onderwerp gaan. Een portal fungeert als vertrekpunt en overzichtstabel voor verdere navigatie binnen een onderwerp.

⁹⁴ Rockley, Ann; e.a. - *Managing enterprise content* - pag. 358-359

gegevenselementen waarmee de inhoud van het element wordt weergegeven. Hierbij gaat het om elementen als de titel van een boek, naam van de auteur of het jaar van uitgave. SGML heeft echter nadelen, waarvan het belangrijkste is dat het bijzonder ingewikkeld is. Een tweede nadeel is dat SGML zich richt op inhoud en niet op de weergave van gegevens. Met de komst van de eerste webbrowsers in de jaren negentig van de vorige eeuw werd juist de weergave van gegevens een belangrijk onderwerp.

Om de weergave van gegevens door middel van tags te definiëren werd HTML ontwikkeld. HTML (*Hyper Text Markup Language*) is een afgeleide van SGML. Het nadeel van HTML is echter dat het alleen de weergave van gegevens beschrijft en weer niet de inhoud. De oplossing voor de problemen van zowel SGML als HTML is XML (*eXtended Markup Language*), eveneens een afgeleide van SGML.⁹⁵

XML is ontwikkeld met als hoofddoel een standaard te creëren voor de beschrijving en uitwisseling van elektronische gegevens. Ook XML werkt met tags, maar die liggen niet vast zoals bij SGML en HTML. In tegenstelling tot de tags van HTML, bevatten die van XML geen informatie over opmaak van het element. Een belangrijk verschil tussen XML en HTML en SGML is dat de gebruikers hun eigen tags kunnen ontwikkelen.



```

XML document

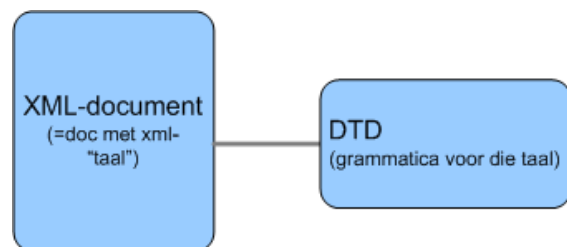
<bedrijf>
  <naam>Van Bussel Document Services </naam>
  <adres>Hof Bruheze 421 </adres>
  <pc>5704 NR</pc>
  <plaats>Helmond</plaats>
  <tel>0492-533 236</tel>
</bedrijf>

```

Figuur 29: XML-document

De 'inhoud' van een document wordt door middel van XML-tags gestructureerd. Elk onderdeel van het document, elk element, wordt door middel van tags gescheiden van de rest van het document. Deze techniek, waarbij documenten worden opgedeeld in losse elementen, maakt het mogelijk de losse elementen - in plaats van het hele document - op te slaan, te doorzoeken en voor andere publicaties te gebruiken. In het voorbeeld hiernaast staan de adresgegevens van een bedrijf zoals die in een klantendatabase zouden kunnen voorkomen.

De losse elementen, zoals naam en adres, worden omschreven door middel van de tags <naam> en <adres>. Dit XML-document bevat alleen elementen, maar om de relatie en inhoud van de elementen te definiëren wordt gebruik gemaakt van een DTD (Document Type Definition). Een DTD bevat een lijst met alle elementen, attributen, notaties en entiteiten die in een XML document kunnen voorkomen. Er wordt niet alleen in aangegeven welke elementen in een document mogen worden gebruikt, maar ook in welke volgorde ze moeten staan, hoe vaak ze mogen of moeten voorkomen, wat het type relatie is dat ze met elkaar hebben en welk soort inhoud (bijvoorbeeld cijfers, letters of beide) ze mogen hebben.



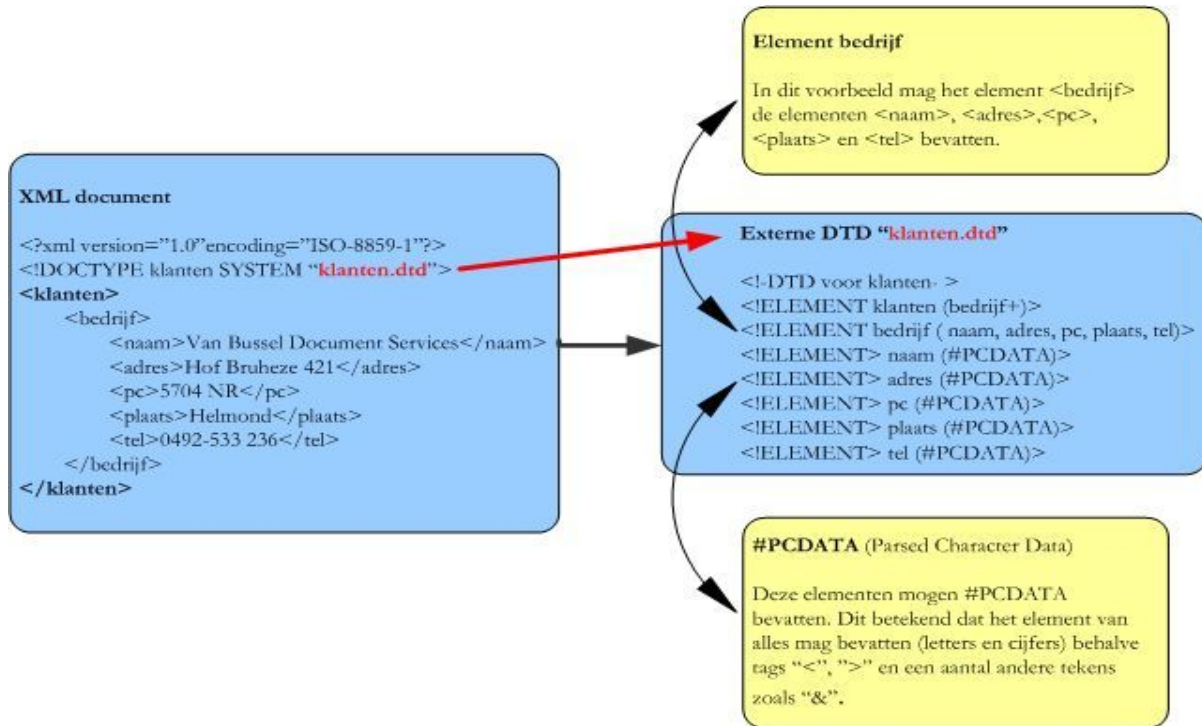
Figuur 30: XML - DTD

In een DTD wordt ook vastgelegd waar in een document extra informatie opgenomen kan worden, zoals verwijzingen naar externe documentdelen, figuren, beeld en geluid, hypertext links en het gebruik van bijzondere schriftsoorten.

Eigenlijk wordt in een DTD bepaald wat de afspraken zijn binnen het betreffende XML-document. Een DTD kan zowel in het XML-document zelf *embedded* zijn, als in een extern document waarnaar gelinkt wordt. De tweede optie houdt in dat de informatie in een extern bestand staat waarnaar in het XML-bestand wordt verwezen. Het voordeel van een externe DTD

⁹⁵ Heijkoop, Harry - *Basicursus XML (herziene editie)* - Den Haag (2005); pag 1-2

is dat deze voor meerdere documenten kan worden gebruikt. Moet de DTD worden aangepast, dan hoeft dit slechts eenmaal te gebeuren door de externe DTD aan te passen. Wordt eenzelfde DTD *embedded* in het document opgenomen, dan zal die in alle documenten waarin ze is gebruikt, moeten worden aangepast. In de DTD wordt van ieder element de inhoud gedefinieerd.⁹⁶

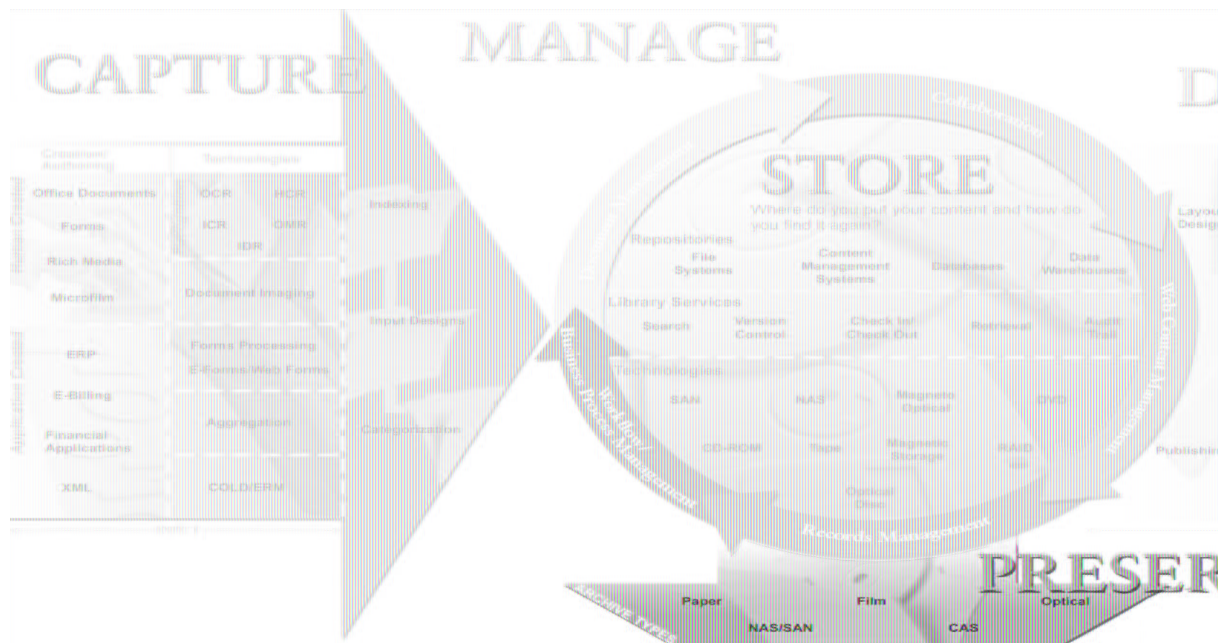


Figuur 31: XML-document met DTD

Voor de opmaak van een XML-document kan gebruik worden gemaakt van *Cascading Stylesheets* (CSS) of *eXtensible Stylesheet Language* (XSL). CSS is een techniek voor de vormgeving waarbij de informatie over de vormgeving wordt toegevoegd aan het XML document. Een stylesheet biedt de mogelijkheid om de inhoud en vormgeving van een document van elkaar te scheiden en op die manier een consistente vormgeving over meerdere documenten te bereiken. CSS worden vooral gebruikt om HTML-documenten op te maken. Een van de sterke kanten van XSL is dat het als het ware in twee stukken is verdeeld. Het eerste deel zorgt voor de traditionele opmaak als lettertype, tekengrootte en kleur (net als CSS). Het tweede deel richt zich op het omzetten van de data. Hierdoor is het mogelijk om door middel van een XSL-stylesheet de inhoud van één XML-document op meerdere manieren vorm te geven. Met behulp van XSL kan de inhoud van een document worden gesorteerd en is de opmaak afhankelijk van de inhoud van een element. Denk hierbij bijvoorbeeld aan prijslijsten gesorteerd van hoog naar laag of juist omgekeerd. Met andere woorden: het is met XSL mogelijk een selectie toe te passen.

⁹⁶ Heijkoop, Harry - *Basiscursus XML* - pag. 59-61

3.6. De vijfde component: Preserve

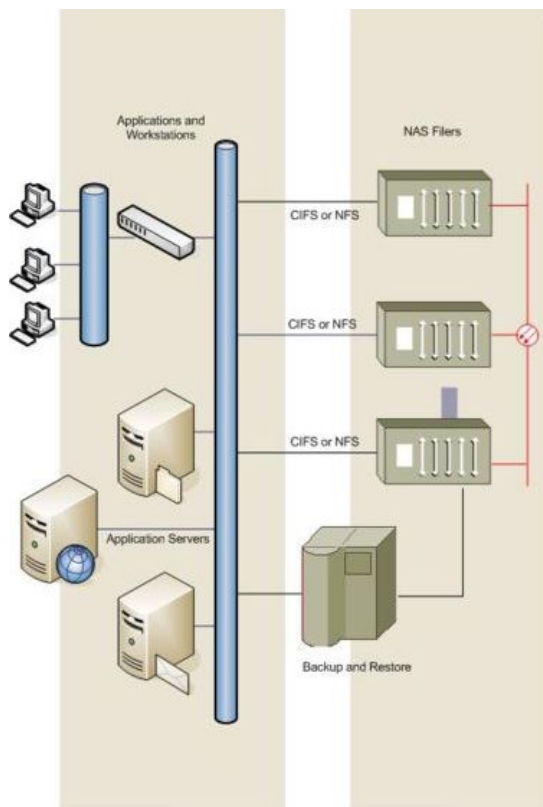


De preserve-component draait om de vraag ‘What are your options for long-term archival and storage of your company’s essential content?’ De preserve-component wordt gebruikt voor zowel de langetermijnopslag, veilige opslag en back-up van statische informatie. De technologieën die hiervoor gebruikt worden, komen voor een deel overeen met die van de store-component. Van de belangrijkste technologieën, te weten NAS, SAN en WORM (write once read many), zal ik een korte beschrijving geven.⁹⁷ Langetermijnopslag van digitale informatie is echter niet alleen een kwestie van die technologieën.

3.6.1. Opslagtechnologieën

Network Attached Storage (NAS) staat voor een opslagmedium dat zoals de naam al zegt direct gekoppeld is aan het netwerk. NAS-apparaten zijn in feite volwaardige file servers. Bij NAS wordt het bestandssysteem beheerd vanuit het NAS-systeem zelf. Dit in tegenstelling tot SAN (Storage Area Network), waarbij het bestandssysteem beheerd wordt door servers. Elke machine die op het netwerk aangesloten is, kan connectie maken met NAS-systemen. Dit voordeel heeft als nadeel dat de prestaties niet zo hoog liggen, juist omdat alle data over het netwerk moeten.

⁹⁷ Kampffmeyer, Ulrich - *ECM enterprise content management* - pag. 62-64



Figuur 32: Network Attached Storage

Nog een nadeel is dat de prestaties van het Local Area Network (LAN) hierdoor ook kunnen afnemen. Voor de opslag van data kunnen NAS-systemen gebruikmaken van meerdere harde schijven die vaak in RAID staan. RAID is een afkorting van Redundant Arrays of Independent Disks. In een RAID-systeem worden meerdere harde schijven gecombineerd tot een groep van schijven (disk array). De in het systeem ingebouwde software zorgt ervoor dat de computer logisch gezien maar één schijf ziet. Op de achtergrond verdeelt de controller de data over de betrokken schijven.

Het RAID-systeem kan uit een willekeurig aantal schijven bestaan. Door deze combinatie van schijven kan een systeem sneller en veiliger werken dan een enkele schijf.⁹⁸ Een *Storage Area Network* (SAN) is een opslagmedium dat extern via een netwerk op een computer is aangesloten. SAN is een architectuur die het mogelijk maakt om externe opslagmedia als RAID, tapes en *optical jukeboxes* op een dusdanige wijze aan servers te koppelen dat het besturingssysteem de media als lokaal aangesloten beschouwt. De aansturing van een SAN gebeurt op

blokniveau, de computer regelt dus zelf de indeling van de via SAN aangeboden schijven. Dit in tegenstelling tot NAS, waar het externe systeem al een complete bestandsstructuur aanbiedt. Het voordeel van de opslag op blokniveau is dat het mogelijk is om maar een deel van een bestand te bewerken zonder dat het hele bestand over het netwerk verstuurd hoeft te worden. Deze techniek maakt het mogelijk om complexe databasesystemen op een SAN onder te brengen, iets wat met NAS vaak niet mogelijk is.⁹⁹

Write Once Read Many (WORM) is een opslagmedium dat zoals de naam al doet vermoeden slechts één keer beschreven kan worden en vervolgens meerdere keren kan worden gelezen. Er zijn twee type WORM-opslagmedia.

Een *True Worm* is een medium waarbij het schrijfproces een onomkeerbare fysieke wijziging veroorzaakt. De data zijn na het schrijfproces permanent vastgelegd en niet meer te wijzigen. De informatie op een True Worm kan alleen verwijderd worden door het medium fysiek te beschadigen. Voorbeelden hiervan zijn een CD-R en een DVD-R.

Het tweede type is een *Soft WORM*. Bij dit type WORM is de 'Write Once' softwarematig en niet tastbaar. Met behulp van softwarematige beveiliging wordt voorkomen dat wijzigingen in opgeslagen data kunnen worden aangebracht.¹⁰⁰

⁹⁸ http://en.wikipedia.org/wiki/Network_attached_storage

http://nl.wikipedia.org/wiki/Network_Attached_Storage

http://nl.wikipedia.org/wiki/Redundant_Array_of_Independent_Disks

⁹⁹ http://en.wikipedia.org/wiki/Storage_area_network

http://nl.wikipedia.org/wiki/Storage_Area_Network

¹⁰⁰ <http://www.sneurope.com/cms/publish/content/showpage.asp?pageid=569>

http://en.wikipedia.org/wiki/Write_Once_Read_Many

3.6.2. Digitale duurzaamheid

Een onderwerp dat onlosmakelijk met de langdurige opslag van digitale bestanden is verbonden, is digitale duurzaamheid. De waan van de dag en de gebruikelijke zorgen hebben tot gevolg dat organisaties nauwelijks nadenken over digitale archivering op de lange termijn.

Toch is de continuïteit van het digitale archief van cruciaal belang en helaas nog lang niet gegarandeerd. Digitale duurzaamheid vormt dan ook een enorme uitdaging voor de opslag van informatie op de lange termijn.

Als een doos met papieren documenten tien jaar in de hoek van een kantoor staat, dan zal de kwaliteit van het papier waarschijnlijk wat zijn afgenomen. Maar de documenten zullen nog goed leesbaar zijn. Brand een Cd-rom met informatie en leg die tien jaar in een la, dan is de kans dat de informatie nog leesbaar is erg klein. Dat digitale informatie na verloop van tijd niet meer leesbaar is, heeft verschillende oorzaken.

Conservering van digitale documenten is niet te vergelijken met die van papieren documenten. Een fundamenteel verschil is dat de conservering van papieren documenten zich richt op de drager (papier) en conservering van digitale documenten op de informatie van een document. De drager van informatie speelt bij digitale conservering een ondergeschikte rol.¹⁰¹ Een groot probleem voor digitale duurzaamheid is dat de ontwikkelingen op ICT-gebied elkaar in hoog tempo opvolgen. Doordat opslagformaten door de jaren heen verschillen, applicaties elkaar snel opvolgen en dragers al snel niet meer gebruikt worden, kan het zijn dat informatie die een aantal jaren geleden is geproduceerd, nu al niet meer leesbaar is. Een digitale werkomgeving zal dus te maken krijgen met het probleem van de garantie van duurzaamheid van digitale documenten. Verlies van digitaal geheugen betekent verlies van organisatiegeheugen. Het archief dient in een organisatie als kennisbron, verantwoording, bewijs en mogelijk in een later stadium als cultureel erfgoed. Het onleesbaar worden van digitale documenten kan vergeleken worden met het wegrotten van de heipalen onder een huis; het zakt weg en stort uiteindelijk in.¹⁰²

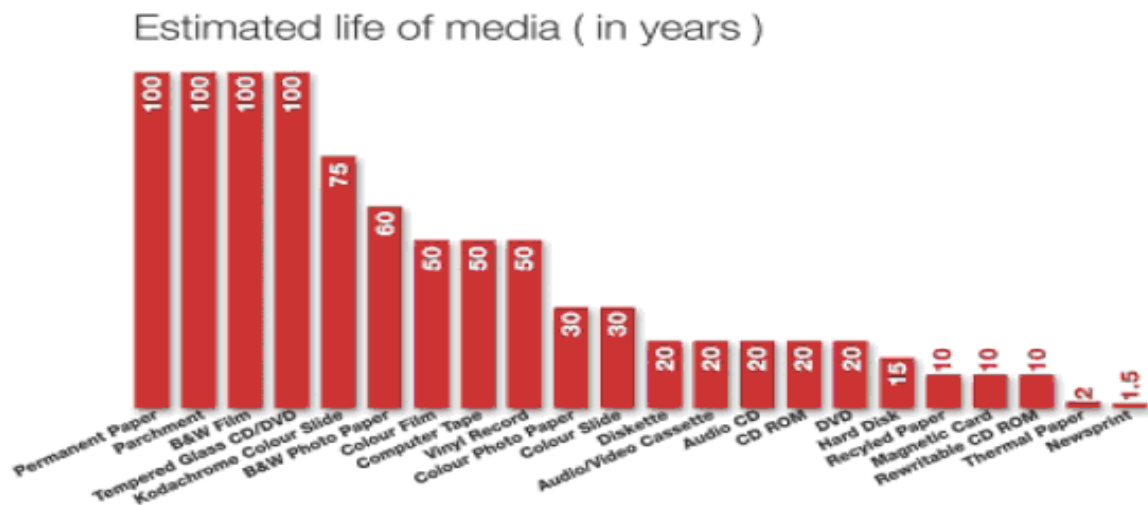
Door voortdurende veranderingen is het moeilijk standaarden te ontwikkelen voor het gebruik van soft- en hardware. Op dit moment staan op de website van whatis.com maar liefst 3520 verschillende bestandsformaten beschreven - om een indicatie te geven van het aantal bestandsformaten dat er in omloop is.¹⁰³ De consequentie hiervan is dat als het al mogelijk is om een standaard te ontwikkelen, deze vaak van een laag niveau is, zoals ASCII, of dat er sprake is van een fabriekseigen standaard als Pdf of Word. De realiteit is dat in de kantoorautomatisering een grote diversiteit aan hardware, software, databases, opslagformaten en gegevensdragers wordt gebruikt en dat het maar zeer de vraag is hoe lang die nog geraadpleegd kunnen worden. Dat dit niet een probleem voor de toekomst is, maar al op dit moment volop speelt, blijkt uit bekende hard- en software voorbeelden als Commodore 64, 5 ¼ floppy's en WP 5.1. Velen hebben hier nog mee gewerkt, maar door de razendsnelle ontwikkelingen zullen zij grote problemen ondervinden als zij die bestanden nog willen raadplegen.¹⁰⁴

¹⁰¹ Horsman, Peter - *Archiveren. Een inleiding* - pag. 78

¹⁰² Starre, Jan van der - *Omvallende bitjes: ons digitaal geheugen bedreig* - Informatie professional 2002 (6); p 30-33 alhier; 31

¹⁰³ <http://whatis.techtarget.com/file-extension-list-A/0,289933,sid9,00.html> (laatste update op 27-12-2006)

¹⁰⁴ Hofman, Hans - *De digitale archivaris: een nieuwe wereld. De invloed van de informatietechnologie op het archiefvak* - In: Horsman, Peter J.; Ketelaar, Frederick C.J.; Thomassen, Theo H.P.M. (red.) - Naar een nieuw paradigma in de archivistiek. Jaarboek 1999 Stichting Archiefpublicaties ('s-Gravenhage 1999) - pag. 211-224. alhier pag. 221



Figuur 33: duurzaamheid informatiedragers¹⁰⁵

Het is een groot probleem voor de archivering dat langdurige bewaring van digitale bestanden nog niet is gegarandeerd. Omdat in de archiefwet geen onderscheid wordt gemaakt tussen analoge en digitale documenten, ligt de lat voor overheidsorganisaties nog hoger dan voor niet-overheidsorganisaties. ‘Het Archiefbesluit 1995 draagt in artikel 11, eerste lid, iedere zorgdrager op zodanige voorzieningen te treffen ten aanzien van de door hem opgemaakte archiefbescheiden die ingevolge een voor hem geldende selectielijst voor bewaring in aanmerking komen, dat bij het raadplegen van die archiefbescheiden na ten minste honderd jaar geen noemenswaardige achteruitgang zal zijn te constateren’. Ondanks dat deze eis in het archiefbesluit van 1995 wordt gesteld, bevat de regeling duurzaamheid archiefbescheiden ten aanzien van digitale archiefbescheiden slechts een zeer beperkt aantal voorschriften over hoe het behoud van de archiefbescheiden moeten worden garanderen.¹⁰⁶ Om tot een oplossing voor het probleem van de digitale duurzaamheid te komen wordt er in verschillende richtingen gedacht, maar dit is een onderwerp dat te complex en te veelomvattend is om hier verder op in te gaan.

¹⁰⁵ Bron: *In paper we trust, why paper endured the digital Age: an investigation into the business, psychological and technological aspects of paper use* - Lexmark international (2002); pag.19

¹⁰⁶ *Regeling duurzaamheid archiefbescheiden* - Uit: Staatscourant 18 september 2001, nr. 180 / pag. 10
http://www.nationaalarchief.nl/images/3_2562.pdf

4. ECM: voorwaarden en meerwaarde

4.1. Inleiding

In het voorgaande hoofdstuk zijn op hoofdlijnen de belangrijkste technologieën waaruit ECM bestaat besproken. Dit zijn in principe allemaal bestaande technologieën, dit hoofdstuk gaat een stap verder en richt zich op het totaal van deze technologieën: ECM als strategie. Eén van uitgangspunten van de ECM-strategie is het elimineren van de meervoudige opslag van identieke informatie. Het credo hierbij is ‘enkelvoudig opslag, meervoudig gebruik’. Dit klinkt logisch maar is in veel organisaties op dit moment, om verschillende redenen, nog niet de praktijk. De realiteit is vaak dat informatie is opgeslagen in verschillende systemen die onderling niet met elkaar kunnen communiceren en dus geen informatie uitwisselen. De informatie zit bij wijze van spreken gevangen in de applicaties. Die worden daarom ook wel informatiesilo's genoemd. Vaak is de enige oplossing om de informatie uit de silo's te halen om identieke informatie in verschillende applicaties op te slaan. Waarom deze situatie onwenselijk is en hoe ECM hier een oplossing voor kan bieden komt in dit hoofdstuk uitgebreid aan de orde. Om te beginnen worden de twee belangrijkste voorwaarden voor het verwezenlijken van de ECM-strategie besproken: *Enterprise Application Integration (EAI)* en *standaardisering van de metadata*. Vervolgens zal worden ingegaan op de vraag wat nu eigenlijk het voordeel is van het koppelen van alle informatie, en de daarbij behorende applicaties. Hiervoor wordt nader ingegaan op het belang van het afbreken van informatiesilo's en het hoe en waarom van Information Lifecycle Management (ILM).

4.2. Voorwaarde voor ECM: *applicatie-integratie*

4.2.1. Eilandautomatisering

In hoofdstuk 2 is ECM gedefinieerd als een strategie voor het integraal gebruik en beheer van ongestructureerde informatie. Om deze strategie te kunnen realiseren is een van de basisgedachten van ECM het integreren van alle binnen de organisatie aanwezige ongestructureerde informatie. Vanuit een technische optiek wordt ECM dan ook beschouwd als een verzameling onderling verbonden applicaties voor het vastleggen, verwerken, beheren, gebruiken en distribueren van ongestructureerde informatie. Zoals uit het voorgaande hoofdstuk duidelijk is geworden vormen de applicaties uit de manage-component de ruggengraat voor het beheer van ongestructureerde informatie. Hierbij kan grofweg de volgende rolverdeling worden aangehouden: document management, document-centric collaboration en web content management richten zich op het dynamische gedeelte van het informatiebeheer. Record management op de statische fase, waarin aan de opgeslagen informatie geen veranderingen meer (mogen) worden aangebracht. En workflow en business proces management vormen de bindende link tussen de componenten. Deze applicaties bestaan ieder op zich uit twee componenten: 1) een database voor het vastleggen en opvragen van informatie en 2) een systeem voor de autorisatie van de toegang tot de informatie. Het doel van ECM is om deze twee componenten slechts éénmaal vast te leggen en vervolgens op alle ‘applicaties’ van toepassing te laten zijn.¹⁰⁷

¹⁰⁷ Kampffmeyer, Ulrich - *ECM enterprise content management* - pag. 10, 38-40, 52

Helaas kunnen veel van de gebruikte systemen niet direct met elkaar communiceren en dus geen informatie uitwisselen. Dat leidt ertoe dat dezelfde data vaak noodzakelijkerwijs in verschillende systemen worden opgeslagen. Om die reden spreekt men ook wel van verticale applicaties, eilandautomatisering en informatiesilo's. Het niet koppelen van systemen is niet alleen inefficiënt voor wat betreft dataopslag, het heeft ook tot gevolg dat sommige processen niet geautomatiseerd kunnen worden, simpelweg omdat deze processen vereisen dat verschillende applicaties aan elkaar worden gekoppeld. De techniek om de verschillende applicaties toch aan elkaar te kunnen verbinden heet Enterprise Application Integration (EAI).¹⁰⁸

4.2.2. Enterprise Application Integration

Enterprise Application Integration (EAI) is een verzamelterm voor de integratietechnologieën waarmee softwaresystemen aan elkaar kunnen worden geknoopt. Doel van EAI is het integreren van de verschillende systemen tot één virtueel systeem. In de ideale situatie merkt de gebruiker niet of er gewerkt wordt met één groot systeem, dan wel een verzameling van met elkaar verbonden kleine systemen.¹⁰⁹ De Gartner Group definieert het als volgt: 'EAI is the unrestricted sharing of data and business processes among any connected application or data sources in the enterprise.'¹¹⁰ Een wat meer technische uitleg luidt 'Enterprise Application Integration is a relatively new research area and many issues related with it are still under investigation. EAI combines a variety of integration technologies such as message brokers, adapters and application servers, to build a centralised integration infrastructure'.¹¹¹

Om de applicaties met elkaar te kunnen laten communiceren, met als doel informatie-uitwisseling mogelijk te maken, moeten deze met elkaar verbonden worden. Let wel, als we spreken over het integreren van systemen dan gaat het niet alleen om het kunnen lezen van gegevens uit een ander systeem, maar ook om ze weer weg te kunnen schrijven in de verschillende systemen. Dit kan gedaan worden door de applicaties rechtstreeks aan elkaar te verbinden door middel van interfaces, maar dit kan ook door middel van EAI. Het verschil tussen de twee methoden is dat bij de eerste methode er een interface wordt gebouwd om twee specifieke applicaties met elkaar te koppelen. EAI daarentegen koppelt applicaties aan een soort tussenlaag van waaruit weer koppelingen zijn gemaakt met de andere applicaties. De communicatie tussen twee systemen verloopt bij EAI hierdoor nooit direct maar altijd via de tussenlaag. Nadeel van het rechtstreeks koppelen van systemen is dat er binnen de kortste keren een warboel van verbindingen tussen de verschillende applicaties ontstaat, ook wel spaghettiarhitectuur genoemd. Het aantal connecties dat nodig is voor een volledige integratie door middel van directe koppeling is de uitkomst van de formule $\frac{n(n-1)}{2}$. Dit betekent dat voor de volledige integratie van 10 applicaties $\frac{(10)(9)}{2}$ of te wel 45 connecties nodig zijn.¹¹²

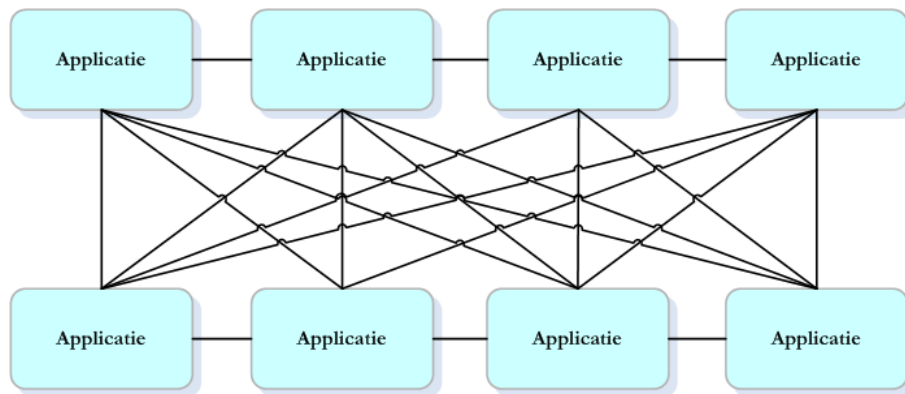
¹⁰⁸Themistocleous, Marinos; Irani, Zahir; Kuljis, Jasna; Love, Peter E.D. - *Extending the Information System Lifecycle through Enterprise Application Integration: A Case Study Experience* - Proceedings of the 37th Hawaii International Conference on System Sciences (2004); pag. 4- 5

¹⁰⁹ Kampffmeyer, Ulrich - *ECM enterprise content management* - pag. 11

¹¹⁰ Zoals geciteerd door Gable, Julie in: - *Enterprise application integration* - Information Management Journal. (March/April 2002)

¹¹¹ Themistocleous, Marinos - *Extending the Information System Lifecycle* - pag. 3

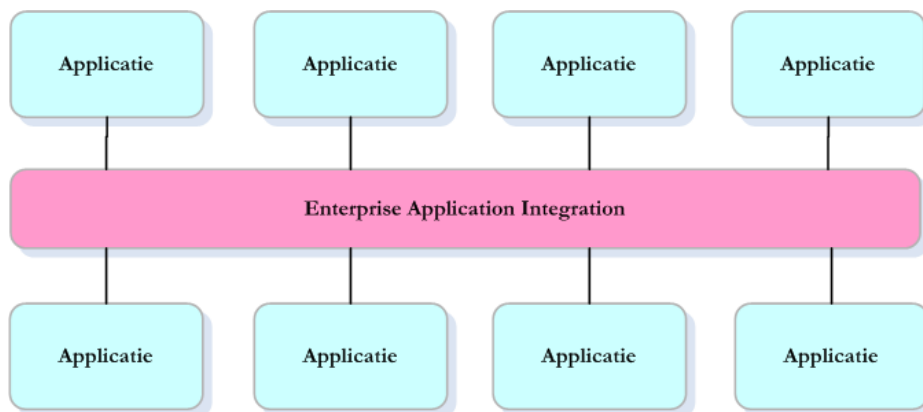
¹¹² http://en.wikipedia.org/wiki/Enterprise_application_integration



Figuur 34: spaghettiarchitectuur

Bijkomend probleem met het direct koppelen van systemen is dat het onderhoud moeilijk is. Zodra er iets in een applicatie verandert, door bijvoorbeeld een upgrade, moeten alle interfaces naar de andere applicatie worden aangepast.

Een praktischere manier om applicaties aan elkaar te koppelen is EAI. Hierbij worden applicaties niet direct aan elkaar gekoppeld maar door middel van een tussenlaag aan elkaar verbonden. In de technische systeemarchitectuur zit EAI dan ook tussen de verschillende applicaties in. Het aantal connecties dat moet worden gemaakt om de systemen met elkaar te verbinden neemt hiermee drastisch af.



Figuur 35: Enterprise Application Integration

De verbinding tussen de applicaties en de EAI gebeurt door middel van connectors. Dit zijn standaardinterfaces tussen een applicatie en een EAI-pakket. De applicaties kunnen door middel van de connectors berichten met elkaar uitwisselen. Hiervoor moeten wel eerst de berichten uit de applicatie worden vertaald naar een standaardtaal, zoals bijvoorbeeld XML, zodat alle aangesloten applicaties de berichten kunnen lezen. In de praktijk zou de interactie tussen de verschillende applicaties er als volgt kunnen uitzien. Er wordt door een gebruiker in applicatie X een aanvraag gedaan. De webserver zet de aanvraag om in een boodschap (bijvoorbeeld in XML) en stuurt deze naar de EAI-server. Deze server bekijkt de boodschap, vertaalt deze indien nodig en stuurt het vervolgens naar de betreffende database. De database server stuurt een antwoord terug naar de EAI, en deze weer naar de webserver die dan het resultaat laat zien.

De grote uitdaging van EAI is, dat het een verbinding moet vormen tussen systemen die op verschillende platforms draaien, verschillende typen databases gebruiken en gebruik maken van verschillende programmeertalen. Er zijn twee verschillende benaderingen van EAI

- **Mediation.** Hierbij fungeert het EAI systeem als een ‘message broker’, die gegevens routeert en transformeert tussen de verschillende applicaties. Op het moment dat bijvoorbeeld nieuwe informatie wordt toegevoegd in een applicatie wordt dit doorgegeven aan de integratiemodule in het EAI systeem. De module verspreidt vervolgens de nieuwe informatie naar de andere relevante applicaties.
- **Federation,** bij deze benadering fungeert het EAI-systeem als een soort balie die is gepositioneerd voor de verschillende applicaties. De interactie tussen gebruiker en applicatie verloopt hierbij standaard via het EAI systeem. In deze benadering is het alleen het EAI systeem (en niet de gebruiker zelf) die interactie heeft met de achterliggende applicaties om de gewenste informatie op te zoeken.¹¹³

4.2.3. ECM suite

Ondanks dat ECM in deze scriptie als een strategie en niet als een software pakket is gedefinieerd bestaat er een snel groeiende markt voor ECM pakketten. Doel is simpel gezegd om met één pakket alle ongestructureerde informatie te beheren en te ontsluiten. Zo is er een ontwikkeling gaande waarbij veel van de ‘manage’ applicaties niet alleen individueel maar ook als geïntegreerde ‘ECM suite’ worden aangeboden. In veel gevallen is hierbij een ‘store’ component in de ‘manage’ component geïntegreerd.¹¹⁴ Belangrijk voordeel hiervan is dat de technische integratie van de verschillende applicaties in een ‘suite’ door de leverancier wordt gedaan. Maar ook de grote spelers op de markt kunnen niet alle applicaties zelf ontwikkelen. Er is de afgelopen jaren dan ook een ontwikkeling gaande waarbij door middel van het opkopen van gespecialiseerde applicaties grote leveranciers functionaliteiten toevoegen aan hun producten. De pakketten worden daarmee krachtiger, maar ook complexer en duurder. De toename in functionaliteit binnen deze ECM suite pakketten is vooral waarneembaar op de volgende drie gebieden:

- Content management - web content management, document management, record management, etc;
- Portal framework - framework om via één uniforme interface verschillende applicaties en gegevens gepersonaliseerd aan te kunnen bieden;
- Collaboration tools - functionaliteit voor samenwerken en delen van informatie, denk aan instant messaging, on-line vergaderingen, chats, wiki, RSS, blog, etc.¹¹⁵

Maar de termen ‘ECM’ en ‘ECM suite’ zijn hip en worden door software leveranciers nog wel eens ten onrechte gebruikt. Een beetje DMS pakket met een workflow-component wordt al een

¹¹³ Themistocleous, Marinos - *Extending the Information System Lifecycle* - pag. 3 - 6

http://en.wikipedia.org/wiki/Enterprise_application_integration

http://nl.wikipedia.org/wiki/Enterprise_Application_Integration

¹¹⁴ Kampffmeyer, Ulrich - *ECM enterprise content management* - pag. 38-40, 52

¹¹⁵ Leer, Paul de - *ECM-markt in beweging* - Verdonck, Klooster en Associates (2004) http://www.jokietje.be/school/mmPublishing/Extra_Informatie/Extra_Informatie/040728_VIP_ECM_PL.pdf

ECM suite genoemd. We spreken echter pas over een ECM Suite als het pakket voldoet aan de volgende voorwaarden:

1. 'Compatibility of all components with all other components.
2. Modular and component expandability.
3. 'The use of shared administration components and resources for authorization, security, access, classes, metadata, storage addresses, communication etc.
4. Uniform use of all information by all components of the suite.
5. Platform conformity, i.e. all components are executable on all platforms of the suite.
6. Uniform use of all information by all components of the suite.
7. Shared development plan which ensures the compatibility of the individual components as well as of the suite.
8. Completeness, i.e. inclusion of all components normally required for the target functionality.
9. Compliance with standards to facilitate incorporation of third-party components as effectively as the native suite components.
10. Shared and consistent sales and marketing concept for both the suite and its constituent components, including pricing models for suite and components.'¹¹⁶

Het idee van een ECM suite klinkt echter mooier dan de realiteit vaak is. Een probleem waar de meeste ECM-softwareleveranciers mee te maken hebben, is dat de aangekochte producten bijna nooit in dezelfde 'stijl' zijn ontwikkeld als de al aanwezige componenten. Zo kunnen producten die deel uitmaken van de zelfde ECM suite bijvoorbeeld een totaal andere *look & feel* hebben. Ook kunnen de gebruikte technologieën verschillend zijn. De integratie tussen de verschillende applicaties blijkt ook bij ECM systemen die als suite worden aangeboden dan ook vaak nog lang niet altijd optimaal te zijn.¹¹⁷

'Also, the various function-point applications within some vendor suites remain far from integrated. In the product reviews below, you'll see suites whose individual components run natively on separate operating systems, support different application servers, and carry divergent repository models. This kind of platform inconsistency should come as little surprise given the turgid pace of mergers and acquisitions and the relative dearth of pure R&D resources at many vendors' disposals to stitch different tools together.'¹¹⁸

¹¹⁶ Kampffmeyer, Ulrich - *ECM enterprise content management* - pag. 75-76

¹¹⁷ Hartman Erik, M. - *Content management: een groeimarkt* - pag. 16-17

Klein, Johan - *Enterprise Content Management, Trends voor 2005 - Lost Boys* (2004); pag 6-8

¹¹⁸ *The ECM Suites Report 2008, Comprehensive Product Evaluations (sample edition version 2.0 - CMS Watch* (2008); pag. 13

4.3. Voorwaarde voor ECM: standaardisering van metadata

4.3.1. Waarom standaardisering

Zoals ik hierboven heb uitgelegd gaat de ECM-strategie uit van koppeling van verschillende applicaties en databases. Maar hoe vind je snel en efficiënt de bruikbare informatie in die enorme verzameling? Welke informatie is aan elkaar verwant? Wat voor soort informatie is het eigenlijk en wat is de status ervan? Een zoekmachine die alles *fulltext* doorzoekt is ongetwijfeld bruikbaar, maar schiet bij dergelijke vragen toch al snel tekort. Een zeer belangrijk aspect van informatie- en archiefmanagement, dat in het bijzonder geldt voor de uitwisseling van informatie, is dan ook het gebruik van metadata.

Het bepalen van metadata is altijd een belangrijk onderdeel geweest van informatie- en documentbeheer. Het gebruik van metadata is dan ook niet nieuw; ook in de traditionele archiefsystemen wordt er gebruik van gemaakt. Het gaat daarbij om informatie als de naam van de auteur, de datum van ontvangst en de aanduiding van het type document. Papieren documenten zijn voor een groot deel zelf de drager van deze metadata. Deze kunnen zowel bij creatie van het document zijn opgenomen als op een later moment aan het document worden toegevoegd. Een belangrijk kenmerk van papieren archiefbescheiden is dat deze vaak impliciete metadata bevatten, die door de gebruiker van de stukken kunnen worden afgeleid. Zo wordt de context van archiefstukken deels impliciet bepaald door bijvoorbeeld de fysieke locatie en de plaatsing samen met andere documenten in een dossier. Context en archiefbescheiden zijn onlosmakelijk met elkaar verbonden, zonder context verliezen archiefbescheiden hun waarde.¹¹⁹ Digitale documenten hebben grotendeels dezelfde soort metadata als papieren documenten, maar in tegenstelling tot papieren documenten zijn ze in de regel niet fysiek op het document te vinden. Een digitale omgeving stelt dan ook andere en hogere eisen aan de vaststelling, toekenning, het beheer en gebruik van metadata.

‘Context en archiefbescheiden zijn onverbrekelijk met elkaar verbonden.

Archiefbescheiden in een digitale omgeving worden ‘gefixeerd’ door metadata, zowel wat vorm, structuur en inhoud als wat tijd en plaats betreft. Tijd en plaats worden vastgelegd in gegevens over de context. Deze contextgegevens dienen betekenis te geven aan die archiefbescheiden, zodat deze ook voor niet direct betrokkenen begrijpelijk zijn, en nemen in de tijd in omvang toe. Oorzaken daarvan zijn het gebruik van de archiefbescheiden in steeds weer nieuwe omgevingen, en het beheer dat voor hun instandhouding moet zorg dragen.’¹²⁰

Dat metadata in de digitale omgeving niet fysiek op het document te vinden zijn, maakt het noodzakelijk dat ze in applicaties en procedures worden vastgelegd. Zoals ik eerder heb aangegeven, wordt voor de digitale archivering en de daarbij behorende metadata het werkproces als uitgangspunt genomen. Een werkproces is een complex van activiteiten dat ter vervulling van een taak wordt verricht en een informatieproduct oplevert. Elke stap of substap die genomen wordt, levert niet alleen een informatieproduct op, maar ook een verzameling metadata. Het gaat hierbij om metadata die bijvoorbeeld betrekking hebben op het betreffende werkproces, de stap

¹¹⁹ Bussel, G.J.; e.a - *Softwarespecificaties voor Record Management Applicaties* - pag. 92-93

Logisch metadataamodel 'Functionaliteiten kernmodel Interlab' Versie 1.0 - Werkgroep Metadata Interlab (2004); pag 6-7

http://www.rijks-interlab.nl/Images/Kernmodel-LogischMetadataamodel-DI_tcm134-82198.pdf

¹²⁰ Hofman, Hans - *Een uitdijend beelal? Context van archiefbescheiden* - In: Context, interpretatiekaders in de archivistiek, stichting archiefpublicaties jaarboek (2000); pag, 45-65, alhier pag. 60

binnen dat werkproces, de datum en het tijdstip waarop het heeft plaatsgevonden en wie er bij betrokken zijn geweest. Het geheel aan informatie, dus zowel de inhoud van het informatieproduct als de bijbehorende metadata, dient gedurende het werkproces te worden vastgelegd.¹²¹ Zoals uit het bovenstaande al blijkt hebben metadata betrekking op uiteenlopende zaken en kunnen ze voor meerdere doeleinden worden gebruikt. Vanuit de vraag waartoe metadata dienen worden in de ReMANO¹²² de volgende typeringen gebruikt:

- Metadata die documenten identificeren: o.a. uniek registratienummer, datum opmaak, verzending of ontvangst, verwijzing naar een dossier of andere aggregatievorm, titel, auteur;
- Metadata die een juiste interpretatie van documenten mogelijk maken: o.a. classificatienummer, verwijzing naar dossier, verwijzing naar proces en procedure;
- Metadata die de authenticiteit van documenten aantonen: o.a. datum opmaak, datum verzending of ontvangst, auteur, relatie met andere documenten, wijze van verzending, wijze van authenticatie (elektronische handtekening);
- Metadata die de documenten makkelijker vindbaar maken: onderwerp, auteur, titel, datum, classificatiecode;
- Metadata die correcte (re)presentatie mogelijk maken: o.a. software en hardware omgeving, fysieke opslaggegevens, compressiegegevens;
- Metadata die beheershandelingen sturen of mogelijk maken: o.a. toegangsautorisatie, datum vernietiging of overdracht, datum conversie of migratie; hiertoe kunnen ook logistieke data gerekend worden (verblijfplaats);
- Metadata die uitgevoerde beheershandelingen vastleggen: o.a. wanneer door wie gebruikt, datum conversie, conversie of migratie *log file*, datum en wijze van overdracht, datum vernietiging.¹²³

Bij de inrichting van ECM moet men zich realiseren dat een flink aantal metadata door de systemen zelf kan worden vastgelegd. Zo kan vooraf per werkproces vastlegt worden wat de kenmerken van de ontvangen en gecreëerde documenten zijn, bijvoorbeeld bewaartermijnen en eigenaren. Dit maakt mogelijk dat de registratie van een document wordt gereduceerd tot de identificatie van het documenttype (bijvoorbeeld de aanvraag voor een evenementenvergunning of een factuur). De applicatie vult dan de meeste registratievelden automatisch in op basis van de vooraf in het systeem vastgelegde informatie.

Een andere mogelijkheid is dat de 'actor'¹²⁴ bij het systeem bekend is als degene die is ingelogd en dat de autorisatie van deze actor hierdoor automatisch bekend is. Ook metadata als datum, tijd en de versie/applicatie/platform waarmee wordt gewerkt, zijn automatisch aan het document te koppelen. Typ bijvoorbeeld maar eens <CTRL>+D in Adobe's acroread. Alle metadata die over het document worden getoond, zijn automatisch gegenereerd. Uitgangspunt bij het toevoegen van metadata moet dan ook zijn dat de medewerker bij de uitvoering van het bedrijfsproces er zo

¹²¹ Bussel, G.J.; e.a. - *Softwarespecificaties voor Record Management Applicaties* - pag. 92

Arnold, A.G.; Glashouwer, B - *De moderne informatiehuishouding van de digitale overheid* - pag. 27

¹²² ReMANO staat voor de software specificaties voor Record Management Applicaties voor de Nederlandse Overheid (ReMANO). Doel van de ReMANO 'is een set specificaties waarmee overheidsinstellingen in Nederland kunnen communiceren met producenten van software. Het document vormt samen met het Metadata model één geheel dat als een ontwerp- en/of selectiedocument voor software is te gebruiken. (ReMANO pag. 13)

¹²³ Bussel, G.J.; e.a. - *Softwarespecificaties voor Record Management Applicaties* - pag. 93

¹²⁴ Een *actor* is een overheidsorgaan, een particuliere organisatie of een persoon die een rol speelt of werkzaam is op een bepaald beleidsterrein. <http://www.vhic.nl/default.asp?A1PID=2238PDGH&A1SID=321142200757>

min mogelijk last van ondervindt. Het op geautomatiseerde wijze toekennen van metadata heeft nog een ander belangrijk voordeel: het verkleint de kans op fouten.¹²⁵

Omdat ECM uitgaat van de koppeling van verschillende systemen is het van groot belang de gebruikte metadata te standaardiseren. De invoering van een uniforme metadata-set kan echter een moeilijke en gecompliceerde operatie zijn, omdat er organisatiebreed afspraken moeten worden gemaakt over de definiëring en het gebruik ervan. Complexiteit mag echter nooit een reden zijn om de metadata niet te standaardiseren, want een uniforme metadata-set heeft tot doel:

- Het vergroten van de uitwisselbaarheid (interoperabiliteit) van informatie. Standaardisering van metadata is essentieel voor de ontsluiting en uitwisseling van documenten, dossiers en/of procesinformatie. Uitwisselbaarheid van informatie kan in veel gevallen betekenen dat er doelmatiger gebruik van beschikbare informatie kan worden gemaakt.
- Het vergroten van de bereikbaarheid en toegankelijkheid van informatie. Met andere woorden: relevante doelgroepen in staat stellen om via verschillende ingangen relevante informatie te vinden. Gestandaardiseerde metadata dragen bij aan een eenduidige toegankelijkheid van de informatie.
- Het verhoogt de effectiviteit van de organisatie. Door in de verschillende informatiedomeinen en –systemen dezelfde gestandaardiseerde ontsluitingsmethodiek te gebruiken geeft dit structuur en houvast aan de gebruikers.
- Het borgen van de duurzame ontsluiting van informatie. Juist met het oog op digitale duurzaamheid is de standaardisering van metadata van groot belang.¹²⁶

4.3.2. Metadata-norm: NEN-ISO 23081-1:2006

Welke metadata dienen er nu precies te worden vastgelegd en hoe dient dit te gebeuren? De NEN-ISO 23081-1¹²⁷ standaard biedt hiervoor een handvat. Het geeft een overzicht van de metadata-sets die nodig zijn om aan de eisen te voldoen die in de ISO 15489-1:2001¹²⁸ aan informatie- en archiefmanagement worden gesteld. De ISO 23081-1 definieert echter geen verplichte set van metadata. Dat is ook niet mogelijk aangezien deze set wordt bepaald door zowel de eisen van de organisatie zelf als door specifieke juridische eisen.¹²⁹ In de ISO 23081-1 wordt onderscheid gemaakt tussen twee fasen waarin metadata ontstaan en/of worden toegevoegd. In de eerste fase wordt het document gecreëerd/ontvangen. Deze metadata, die het best omschreven kunnen worden als *ontstaansmetadata*, leggen zowel het bedrijfsproces vast waarin het document is gecreëerd of ontvangen (context) als de inhoud, structuur en vorm van het document. In de tweede fase wordt het document vervolgens gebruikt. Deze metadata kunnen

¹²⁵ Koenen, Kees; Baak, Paul - *Het gebeugen als actieve kracht, de archieffunctie binnen de digitale overheid* -(1999); pag. 26 <http://www.rmconventie.nl/uploads/GeheugenAlsActieveKracht.pdf>

Damme, Maarten van - *Business Case implementatietraject DMS stadsdeel Amsterdam-Centrum* - stadsdeel Amsterdam Centrum (2005); pag. 32

¹²⁶ *Logisch metadata-model Functionaliteiten kernmodel Interlab* - pag. 6, 28

¹²⁷ NEN-ISO 23081-1:2006 *information and documentation – record management processes – Metadata for records – part 1: Principles*

¹²⁸ De ‘NEN-ISO 15489-1: Informatie- en archiefmanagement’ is een norm die een algemeen geaccepteerd kader biedt, waarop organisaties hun informatiehuishouding kunnen baseren. Gezien de huidige ontwikkelingen rond bijvoorbeeld de elektronische overheid is een dergelijk kader noodzakelijk en vormt de norm een belangrijk houvast bij de inrichting van de documentaire informatievoorziening. In de norm wordt een stelsel basisregels beschreven (door middel van het gebruik van genormeerde werkwijzen en procedures) om de vereiste aandacht en bescherming te geven aan documentaire informatie (waaronder de archiefbescheiden) en om de (bewijs)informatie, die ze bevatten efficiënter en effectiever terug te kunnen vinden.

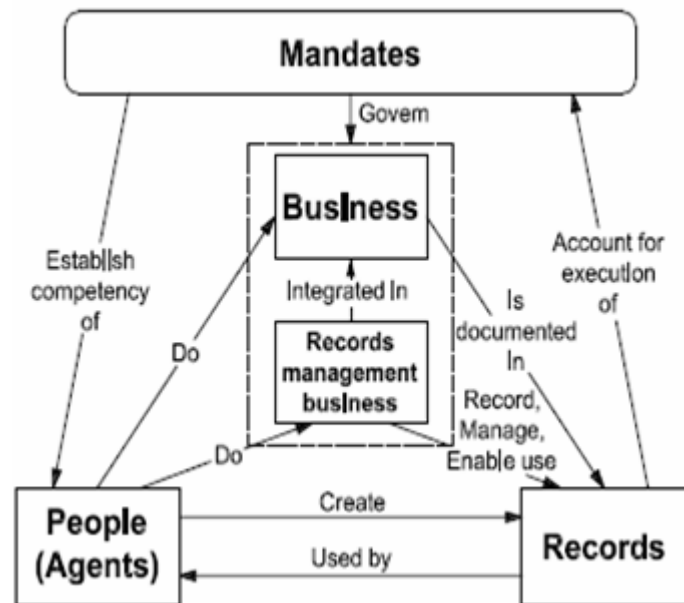
¹²⁹ NEN-ISO 23081-1:2006 *part 1: Principles* - pag. V

het best worden omschreven als *procesmetadata*. Zij leggen vast in welke werkprocessen het document vervolgens wordt gebruikt. Ook iedere verandering aan inhoud, structuur en vorm wordt vastgelegd.¹³⁰

De ISO 23081 bevat een apart hoofdstuk waarin uiteengezet wordt welke metadata dienen te worden vastgelegd om te voldoen aan de eisen die in de ISO 15489-1 worden gesteld.

Men onderscheidt vijf categorieën van metadata, waarbij voor elke categorie een tweedeling wordt gemaakt tussen ontstaansmetadata en procesmetadata. De onderlinge verhoudingen tussen de vijf categorieën kunnen schematisch als volgt worden weergegeven:

1. Metadata over het archiefstuk zelf
2. Metadata over bedrijfsregels, politiek en mandaten
3. Metadata over actoren
4. Metadata over bedrijfsactiviteiten of -processen
5. Metadata over records management processen/ archiefbeheer



Figuur 36: schematische weergave verhouding metadata NEN-ISO 23081-1: 2006

Hieronder zal ik de kenmerken van de vijf bovengenoemde categorieën kort bespreken.

1. Metadata over het archiefstuk zelf.

Het gaat hierbij om metadata als datum, tijdstip van ontstaan en identificatie van personen die betrokken zijn bij het ontstaan van het document. Maar ook om structurele en opslagmetadata, zoals het formaat en de benodigde technische omkadering die nodig zijn om de (digitale) duurzaamheid van het document te kunnen garanderen. Deze metadata hebben onder meer betrekking op de technische karakteristieken, de relatie tussen de gegevens of formatelementen waaruit het archiefstuk is opgebouwd en de technische gegevens die noodzakelijk zijn voor migratie of emulatie. Deze categorie bevat nog twee subcategorieën, namelijk metadata die nodig zijn om de archiefstukken weer terug te kunnen vinden en metadata die betrekking hebben op de beveiliging van archiefstukken. Voor het terugvindbaar maken van documenten dienen er onder andere metadata aan het document te worden gekoppeld waaruit blijkt tot welke aggregatie (bijvoorbeeld dossier of reeks) het document behoort en welke relaties het document heeft met andere documenten, actoren en processen. Voor de beveiliging van documenten dienen er onder andere metadata te worden toegekend die betrekking hebben op zowel de autorisatie als het opvragen en inzien van documenten en metadata.

¹³⁰ Ibidem, pag. 3

2. Metadata over bedrijfsregels, -politiek en mandaten

Archiefdocumenten komen voort uit werkprocessen. Zowel op het moment van opname in het archiefsysteem als gedurende het verdere gebruik dienen metadata te worden vastgelegd waaruit blijkt dat de archiefbescheiden zijn gecreëerd en gebruikt in overeenstemming met de vooraf opgestelde bedrijfsregels, -politiek en mandaten. Daarnaast dienen ook gegevens die betrekking hebben op het beheer van archiefstukken te worden vastgelegd. Zo moeten onder meer de regels en controlesystemen die de creatie, het beheer, de rechten en de raadpleging van zowel de archiefstukken als de metadata hebben beïnvloed, worden gedocumenteerd.

3. Metadata over actoren

Met actoren (in de norm wordt de term *agent* gebruikt) worden zowel individuen als organisaties bedoeld. In de metadata over deze actoren dient te worden vastgelegd wie verantwoordelijk of betrokken zijn bij creatie en beheer van zowel archiefstukken als metadata. Tevens moet in de metadata vastgelegd worden wie geautoriseerd is om welk document in te kijken en/of te veranderen.

4. Metadata over bedrijfsactiviteiten of -processen

Metadata over bedrijfsactiviteiten of -processen zijn nodig om de context van archiefstukken duidelijk te maken. Op basis hiervan kunnen archiefstukken ter verantwoording en bewijs worden gebruikt. Om de context van archiefstukken duidelijk te maken dienen metadata, die de relaties aangeven tussen de archiefstukken, actoren en de bedrijfsfuncties, activiteiten en transacties waarop zij betrekking hebben, te worden vastgelegd. Zo moeten niet alleen de tijd en datum van de creatie van het archiefstuk worden vastgelegd, maar bijvoorbeeld ook de actoren die bij het bedrijfsproces betrokken zijn geweest.

5. Metadata over record managementprocessen/archiefbeheer

Hiermee worden o.a. metadata bedoeld die betrekking hebben op het archiefbeheer. In de metadata dient informatie te worden vastgelegd die betrekking heeft op ordeningsschema's, rechten van actoren om archiefstukken aan te maken, te wijzigen en te vernietigen, e.d. Daarnaast dienen metadata te worden opgenomen die een verband leggen tussen de archiefstukken onderling en tussen de archiefstukken, actoren en werkprocessen waaruit ze voorkomen.¹³¹

4.3.3. De inhoud van metadata

Met alleen een metadata-standaard zijn we er nog niet. De verplichte aanwezigheid van metadata-element om bijvoorbeeld de ontvangstdatum van een document vast te kunnen leggen is slechts één aspect. Hoe de datum vervolgens moet worden vastgelegd is weer een ander aspect. Het is dan ook niet alleen van belang dat bepaald wordt welke metadata worden gebruikt, maar ook dat de registratie ervan op een gestandaardiseerde wijze gebeurt.

De standaardisering van metadata bestaat dan ook uit twee elementen. Ten eerste moeten er afspraken worden gemaakt waarbij per informatietype gedetailleerd wordt beschreven welke metadata-elementen moeten en/of mogen worden gebruikt. Ten tweede moet per metadata-element bepaald worden wat de toegestane waarden zijn. Worden hier geen duidelijke afspraken over gemaakt, dan zal de inhoud van de metadata-velden snel vervuilen. Want wonen we nu in Amsterdam, A'dam of gewoon in Mokum? En is het Jan Peter Balkenende, Balkenende, Jan Peter, of J.P. Balkenende? Voor gewone mensen zijn dit triviale verschillen - we weten immers toch dat we het over de hoofdstad en de premier hebben - maar voor digitale systemen en het gebruik ervan zijn ze een ramp.

¹³¹Ibidem, pag. 11-19

Poppe, Jeroen - *Metadata: ISO 23081 en andere standaarden (2)* - In OD 2006 nr. 10 pag. 18-19

Afspraken over de toegestane waarden van metadata-elementen worden gemaakt in *encoding schemes*. Door middel van encoding schemes wordt contextuele informatie aan de data gegeven, zodat deze geïnterpreteerd kunnen worden. Die contextuele informatie kan worden toegevoegd door bijvoorbeeld het gebruik van gecontroleerd vocabulaire of *parsing rules*. Met andere woorden, door middel van de encoding schemes wordt bepaald welke waarde(n) het element kan of mag bevatten. Er zijn twee soorten encoding schemes: *vocabulary encoding schemes* en *syntax encoding schemes*.¹³²

- Syntax encoding schemes. Dit zijn de ‘grammaticaregels’ voor de notatiewijze van de waarde van de elementen. De manier waarop een datum wordt genoteerd wordt hierin bijvoorbeeld vastgelegd, bijvoorbeeld 5-jan-2008, 5-1-2008, 05-01-2008 of toch 20080105.
- Vocabulary encoding scheme. Dit is een lijst met toegestane waarden (gecontroleerd vocabulaire). De door de VNG vastgestelde lijst met alle gemeentenamen in Nederland is een voorbeeld van een vocabulary encoding scheme.

Iedere organisatie moet voor zichzelf bepalen welke metadata van belang zijn en welke eisen er worden gesteld aan de in te voeren waarde van deze metadata. Voor een functionele inrichting dient er vervolgens van elk afzonderlijk registratieveld (in dit geval het veld *registratiedatum*) een beschrijving te worden gemaakt die er als volgt zou kunnen uitzien.

2. Registratiedatum		Basiskenmerken
Definitie	Datum van ontvangst van het document bij de DIV afdeling	
Veldlengte	10 posities	Datumveld
Verplicht veld	Ja	
Validatie	Ja	DD-MM-JJJJ
Automatisch gevuld	Ja	Systeemdatum
Default	Nee	
Overschrijfbaar	Nee	
Koppeling	Nee	
Raadpleegtabel	Nee	
Toelichting	Registratiedatum wordt bij registratie automatisch op systeemdatum geregistreerd.	

Figuur 37: beschrijving van registratieveld

- **Definitie:** definitie van de titel van het veld. Hier wordt duidelijk gemaakt wat de bedoeling van het registratieveld is.
- **Veldlengte:** minimaal aantal posities voor een volledige weergave van de informatie.
- **Verplicht veld:** verplichting van de medewerker om het veld in te vullen.
- **Validatie:** controle van het systeem of de ingegeven informatie overeenkomt met het verwachte formaat.
- **Automatisch gevuld:** of het systeem aan de hand van eerder ingevoerde informatie óf na keuze uit een raadpleegtabel het betreffende veld automatisch kan invullen met informatie.
- **Default:** de waarde die het veld weergeeft wanneer het systeem of de medewerker zelf geen informatie invoeren.

¹³² <http://dublincore.org/usage/documents/principles/>

Rooij, R. de - *Maak meer mogelijk met metadata, het DNA van een website* – Naar voren (2005)
<http://www.naarvoren.nl/artikel/metadata/>

- **Overschrijfbaar:** of het veld door iemand anders later overschreven mag worden (afhankelijk van autorisatieniveau).
- **Koppeling:** de medewerker of het systeem kan een externe tabel raadplegen voor het opzoeken en invoeren van de juiste informatie.
- **Raadpleegtabel:** tabellen worden geraadpleegd voor het opzoeken van NAW-gegevens, andere dossiers etc.
- **Toelichting:** informatie over het registratieveld die met de bovenstaande eigenschappen niet beschreven kan worden.¹³³

¹³³ Damme, Maarten van - *Functioneel Ontwerp Vervanging GRS Stadsdeel Amsterdam-Centrum Vierde concept* - stadsdeel Amsterdam Centrum (2006); pag. 4-5

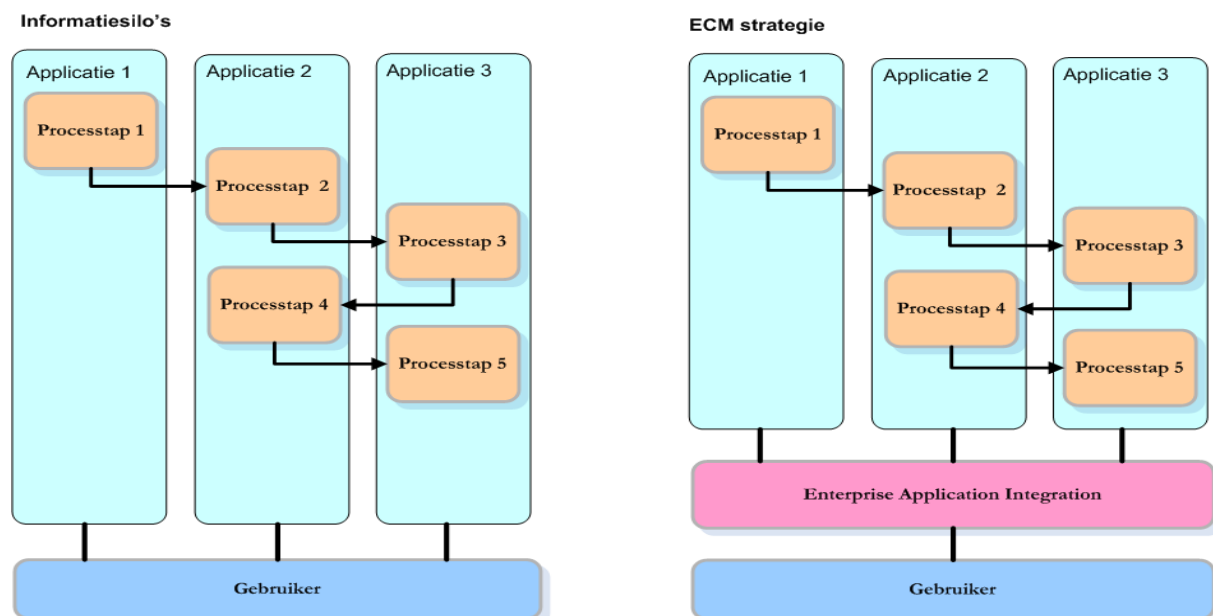
4.4. Meerwaarde van ECM: afbreken van informatiesilo's

4.4.1. Wat zijn informatiesilo's

In het voorgaande stuk is uitgebreid ingegaan op zowel de technische (EAI) als de inhoudelijk (metadata) voorwaarde voor een succesvolle ECM-strategie. Maar wat is het uiteindelijke doel van het koppelen van de verschillende systemen? Het antwoord hierop is eenvoudig: de informatiesilo's afbreken. Dit antwoord is misschien niet voor iedereen even verhelderend en ik zal hieronder dan ook uitleggen wat ik onder informatiesilo's versta en waarom het afbreken hiervan zo belangrijk is.

Met informatiesilo's worden de verschillende applicaties en databases bedoeld waarin de informatie zit opgeslagen. We spreken van informatiesilo's en verticale applicatie architectuur op het moment dat applicaties niet aan elkaar zijn gekoppeld en de informatie die erin is opgeslagen dus 'gevangen' zit in de verschillende databases. De informatievoorziening en de daarvoor gebruikte systemen zijn vaak gemodelleerd om procesgegevens binnen een afdeling of sector te kunnen verwerken en op te slaan. Kijken we bijvoorbeeld naar een gemeente dan zijn er voor vergunningen andere informatiesystemen dan voor belastingen en voor persoonsinformatie is weer een ander systeem ingericht. Het voordeel hiervan is dat per product een optimale oplossing kan bestaan. Een groot nadeel is dat veel gegevens dubbel voorkomen in de gemeente (bijvoorbeeld persoonsgegevens en adresgegevens).

Belangrijkste nadeel van informatiesilo's is dat de informatie alleen te gebruiken is binnen de bewuste applicatie en niet daarbuiten. In de praktijk zal echter voor het uitvoeren van een werkproces over het algemeen niet slechts gebruik worden gemaakt van één maar van meerder applicaties. Zijn deze applicaties niet aan elkaar gekoppeld dan zal de informatie uit processtap 1 (applicatie 1) niet beschikbaar zijn bij de uitvoering van processtap 2 (applicatie 2), zie voorbeeld hieronder.



Figuur 38: informatiesilo's versus ECM

Dat de aanwezige informatie zit opgesloten in applicaties kan in de praktijk veel gevolgen hebben. Om er een paar te noemen. Dezelfde informatie moet door een klant meerdere malen worden verstrekt om vervolgens in meerdere applicaties te worden ingevoerd. Medewerkers moeten in verschillende systemen zoeken naar de benodigde informatie. Of nog erger, informatie is niet

beschikbaar gedurende de uitvoering van het werkproces omdat medewerkers geen toegang hebben tot de bewuste database of zelfs niet op de hoogte zijn van het bestaan ervan. Het mag duidelijk zijn dat deze situatie verre van ideaal is. Essentieel onderdeel van de ECM-strategie is dan ook het afbreken van de informatiesilo's door het koppelen van applicaties. Zoals in het voorgaande stuk over EAI al is gesteld is het in de ideale situatie zo dat de gebruiker niet merkt of deze aan het werk is in één applicatie of in meerdere tegelijk. Doel is alle aanwezige informatie organisatiebreed te ontsluiten en toegankelijk te maken voor allen die daartoe geautoriseerd zijn.¹³⁴

4.4.2. De risico's van informatiesilo's

Het gevaar van informatiesilo's is niet alleen dat de benodigde informatie niet beschikbaar is, maar ook dat de aanwezige informatie inconsistent wordt.

“Too often, content is created by authors working in isolation from other authors within the organization. Walls are erected among content areas and even within content areas, which leads to content being created, and re-created, and re-created. Often with changes or differences at each iteration. We call this the content silo trap.”¹³⁵

Wanneer informatie op verschillende plekken door verschillende auteurs wordt gecreëerd, dan is het onvermijdelijk dat de content van elkaar zal verschillen. Dit leidt er in de praktijk toe dat er twee (of meer) verschillende vormen van dezelfde content bestaan en dat documenten hierdoor inconsistent of zelfs inaccuraat worden. Daarbij is het mogelijk dat beide content-elementen op een later moment opnieuw worden gebruikt, waardoor de inconsistentie zich opstapelt en er uiteindelijk twee verschillende content-elementen ontstaan.¹³⁶

“In general, the lack of application integration represented a problem throughout the content life cycle. For instance, the system for managing documents is poorly integrated with several production packages. While document copies could reside in several different systems, changes and updates were typically registered only in one of these.”¹³⁷

Het probleem is dus niet alleen dat informatie zit ‘opgesloten’ in verschillende applicaties, maar ook dat de gebruiker vaak niet (goed) weet welke informatie in welke applicatie zit. Dit maakt het voor de gebruiker uiterst ingewikkeld om toch alle informatie te vinden die nodig is om de taak uit te kunnen voeren. Dat dit een reëel probleem is, blijkt wel uit het onderstaande voorbeeld uit de praktijk. Hier wordt goed duidelijk dat in grote organisaties het aantal applicaties en databases (silo's) zo enorm kan oplopen dat het de gebruiker praktisch onmogelijk wordt gemaakt de juiste informatie te vinden.

“Typical of many large corporations, Statoil's IT architecture had evolved gradually into a broad portfolio of technologies with partly overlapping functionality and applications. As a result, the company's information was scattered across a number of different storage

¹³⁴ Kaderdocument, *Afbakening functionaliteitsgebieden regionale midoffice architectuur Drechtsteden Digitaal, Versie 1.2* - (2006); pag. 8

http://www.drechtstedendigitaal.nl/Dordrecht/up/ZiayqInHiB_Kaderdocument_regionale_midoffice_concept_v1.2.pdf

¹³⁵ Rockley, Ann; e.a.- *Managing enterprise content* - pag. 5

¹³⁶ Ibidem, 24

Munkvold Bjørn, e.a - *Contemporary Issues of Enterprise Content Management: The Case of Statoil* - pag. 81

¹³⁷ Munkvold - *Contemporary Issues of Enterprise Content Management* - pag. 81

media and applications, with the total number of databases estimated to exceed 5,500. This again created major challenges related to information retrieval, version control and information quality across the enterprise. (...) Although these could be technically shared across the enterprise they were not logically shared, as people were more or less unaware of the existence, contents, and relevance of a great number of these.¹³⁸

Door het koppelen van applicaties kan een van de basisprincipes van ECM worden toegepast, namelijk 'enkelvoudige opslag, meervoudig gebruik'. Doel van dit principe is het reduceren van meervoudige bestanden die identieke gegevens bevatten. Door het koppelen van applicaties en door enkelvoudige opslag als uitgangspunt te nemen is het niet langer nodig om dezelfde informatie in verschillende databases op te slaan. 'The Enterprise Library does not require content to be moved from its existing locations, but is a metadata layer that indexes and enables information workers to access content wherever it lives. The Enterprise Library layer makes it easy tot retrieve information from multiple enterprise systems with a single search and to present the results in any interface, including Web-based portals and desktop applications. The Enterprise library is connected to long-term storage devices to enable the use of record management rules to archive content for compliance purposes.'¹³⁹

Enkelvoudige opslag heeft niet alleen een positief effect op de terugvindbaarheid van informatie, maar biedt nog een aantal andere zeer belangrijke voordelen:

- De inhoud van de informatie is consistent omdat er slechts één exemplaar van de informatie is. Iedereen gebruikt dus dezelfde informatie.
- De status/versie van een document is altijd helder. Voorkomen wordt dat in verschillende databases zich verschillende versies van een document bevinden, waardoor de status/versie van een document onduidelijk is.
- Voorkomt redundante opslag van informatie waardoor het totaal aan benodigde opslagcapaciteit afneemt.¹⁴⁰

Informatiesilo's en een verticale applicatie-architectuur zijn nogal abstracte termen. Het is daarom niet altijd duidelijk wat ze in de praktijk betekenen en wat dus de meerwaarde van ECM kan zijn. Om een en ander concreter te maken werk ik hieronder een voorbeeld uit aan de hand van zaaksgewijs digitaal werken. Let wel, dit is een ordeningsprincipe dat op zich geen onderdeel uitmaakt van de ECM-strategie, maar dat ik alleen heb gekozen om de toepassing van ECM te concretiseren. Er zijn naast dit ordeningsprincipe vele andere mogelijkheden denkbaar.

Om het voorbeeld te kunnen begrijpen, is het van belang eerst kort uit te leggen wat zaaksgewijs archiveren precies inhoudt. Zoals ik al eerder aangaf, is het beter om bij digitale archivering het bedrijfsproces als uitgangspunt te nemen voor de ordening van de informatiehuishouding. Tijdens de uitvoering van de processen vindt de ordening van gegevens en documenten hier namelijk plaats. En groot voordeel hiervan is dat het bedrijfsproces zowel tijdens als na afloop van het proces zorgt voor de samenhang en context van de gegevens en het document. In de praktijk werkt dit als volgt. Een werkproces zoals hierboven beschreven wordt een zaaktype genoemd. Een zaak is een unieke, eenmalige uitvoering van een zaaktype. Alle documenten die betrekking hebben op de uitvoering van een zaaktype worden verzameld in een zaakdossier. Per zaaktype wordt vooraf bepaald welke documenten in het zaakdossier moeten en

¹³⁸ Ibidem, pag. 73

¹³⁹ Jenkins, Tom - *Enterprise Content Management Technology* - pag. 274

¹⁴⁰ Rockley, Ann; e.a. - *Managing enterprise content* - pag. 24 - 25

mogen. Het basisprincipe hierbij is dat elke keer dat een werkproces (zaaktype) wordt uitgevoerd, dit een uniek zaakdossier oplevert.¹⁴¹

In het voorbeeld ga ik uit van de gemeentelijke gegevensstandaard voor zaken, het Gemeenschappelijk Functioneel Ontwerp Zaken (GFO Zaken) van de Vereniging van Nederlandse Gemeenten (VNG).¹⁴² In het GFO Zaken wordt een minimale set van gegevens gedefinieerd die als standaard gelden om centraal de basiskenmerken van een 'zaak' te kunnen verzamelen en gegevens over de 'zaak' te kunnen halen uit verspreide informatiesystemen.¹⁴³ Op basis van deze basiskenmerken kunnen vervolgens zaakdossiers aan elkaar worden gekoppeld.¹⁴⁴



Figuur 39: GFO zaken model

Het zaakdossier is overigens niet de hoogste verzameleenheid en kan onderdeel uitmaken van een groter dossier. Een klantendossier bijvoorbeeld gaat verder. Het bundelt alle (historische en lopende) zaken van één klant. Eenvoudiger gezegd: een klantendossier bevat alle zaakdossiers van één klant. Door zaken te koppelen aan basiskenmerken zoals persoon, adres en kadastraal object ontstaat er een CRM-functionaliteit (Customer Relationship Management, ofwel klantbeheer). Alle zaken met betrekking tot één subject (klantdossier), één adres (adresdossier), één kadastraal object (gebouwdossier) worden op deze manier met elkaar in relatie gebracht.¹⁴⁵

Nu ik zaakgericht digitaal archiveren in vogelvlucht heb uitgelegd, kunnen we bekijken wat dit in de praktijk betekent in relatie tot losse applicaties en de ECM-strategie. Grofweg zijn er twee modellen te definiëren. De eerste gaat uit van een situatie waarbij de verschillende applicaties die

¹⁴¹ Cate ten, Timo; Groeneweg, Ronald - *De Zaak X...*, *zaakgevijs werken in theorie en digitale praktijk* - Digital display (2007); pag 11-13

<http://www.ddisplay.nl/beheer/UserFiles/File/De%20Zaak%20X.pdf>

¹⁴² Een zaak wordt in het GFO Zaken als volgt gedefinieerd 'samenhangende hoeveelheid werk met een gedefinieerde aanleiding en een gedefinieerd resultaat, waarvan kwaliteit en doorlooptijd bewaakt moeten worden.' *Zaken in zicht, GFO-ZAKEN* - VNG Uitgeverij, Den Haag, (2004) pag. 7

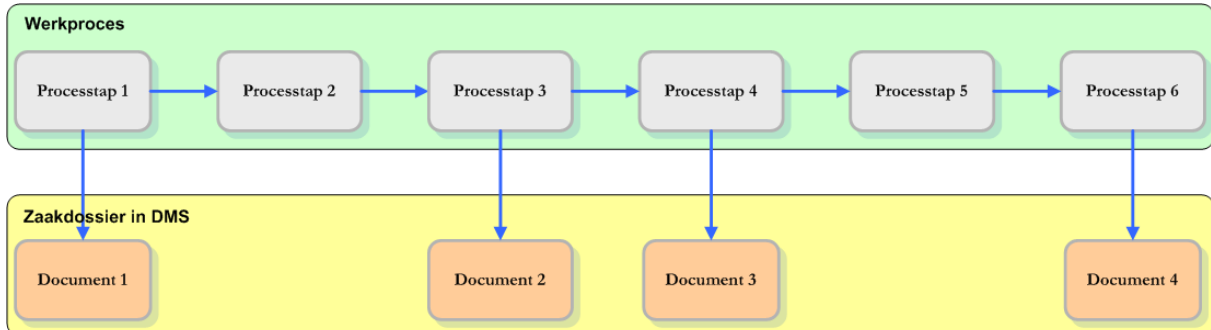
<http://egem-iteams.nl/system/files/gfozakeninzicht.pdf>

¹⁴³ Cate ten, Timo; Groeneweg, Ronald - *De Zaak X...*, *zaakgevijs werken in theorie en digitale praktijk* - pag 11

¹⁴⁴ Het GFO ZAKEN koppelt op een centraal punt basisinformatie over: lopende procedures en afgesloten procedures (ZAAK), de STATUS van de procedure, het SUBJECT dat het VERZOEK tot de zaak heeft gedaan of die de rol van BETROKKENE bij de zaak heeft, de ACTOR die verantwoordelijk is voor de behandeling (een MEDEWERKER of een ORGANISATORISCHE EENHEID), de voor de interne organisatie relevante STAP in de procedure, aan de procedure gekoppelde objecten (VERBLIJFSOBJECT, KADASTRAAL OBJECT) en het betrokken ADRES. (*Zaken in zicht, GFO-ZAKEN* - pag. 7)

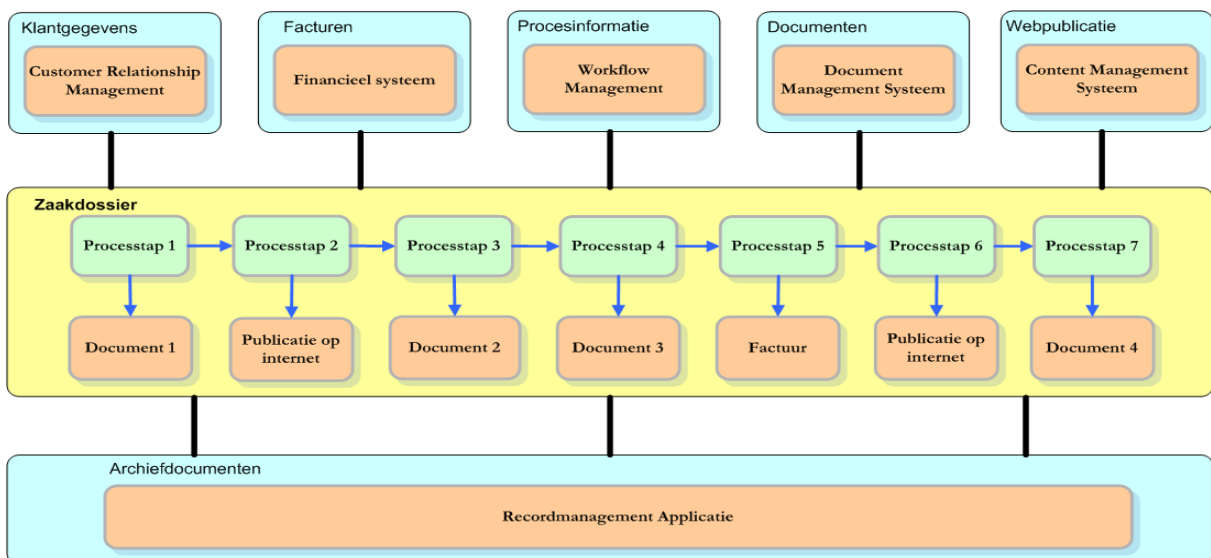
¹⁴⁵ *Zaken in zicht, GFO-ZAKEN* - pag. 15

gedurende het proces worden gebruikt niet zijn gekoppeld (informatiesilo's). De archivering van het zaakdossier vindt hierbij plaats in de verschillende applicaties. Zo worden bijvoorbeeld alle documenten die gedurende het proces worden opgemaakt of ontvangen, opgenomen in het DMS. Schematisch ziet dit er als volgt uit.



Figuur 40: zaakdossier DMS

Werkprocessen bestaan echter niet alleen uit documenten. Neem bijvoorbeeld de stap ‘verstuur een factuur’. Dit is een belangrijke stap binnen een proces. Het is bijvoorbeeld van belang om op een later moment te kunnen vaststellen wanneer een factuur is verstuurd en door wie? Maar noch de factuur zelf, noch de bijbehorende procesinformatie wordt over het algemeen vastgelegd in een DMS. Die informatie zit vaak in het financiële systeem (informatiesilo). Daarnaast is de procesinformatie, d.w.z. de informatie over hoe en wanneer een proces is uitgevoerd, opgeslagen in het workflow managementpakket (informatiesilo).¹⁴⁶ En niet te vergeten de content van de publicatie op het internet, deze is opgeslagen in het Content Management Systeem (informatiesilo). Door dit alles ontbreekt het de gebruiker aan een goed overzicht van alle zaakgerelateerde informatie. Dat brengt het risico met zich mee dat hij of zij belangrijke informatie met betrekking tot de zaak niet vindt en dus niet gebruikt. Door de ECM-strategie van koppeling van systemen en het daarmee afbreken van de informatiesilo's wordt niet slechts een gedeelte, maar wordt **alle** informatie die betrekking heeft op een uniek zaakdossier zichtbaar, mits de gebruiker daartoe is gemachtigd natuurlijk. Deze situatie ziet er schematisch als volgt uit.



Figuur 41: zaakdossier ECM

¹⁴⁶ Door middel van workflow worden documenten niet alleen door de organisatie gerouteerd, maar worden ook bijvoorbeeld alle overdrachtsmomenten, beslismomenten en behandeltermijnen vastgelegd en bewaakt.

4.5. Meerwaarde van ECM: Information Lifecycle Management

4.5.1. Wat is Information Lifecycle Management

Zoals ik in hoofdstuk 2 heb aangegeven, neemt de omvang van de digitale informatie die wereldwijd wordt geproduceerd en opgeslagen exponentieel toe. Over de omvang van de groei van datavolumes binnen bedrijfsomgevingen verschillen de meningen. Zo zijn er schattingen dat de omvang van ongestructureerde data binnen grote bedrijven zich iedere twee tot zes maanden verdubbelt.¹⁴⁷ Anderen gaan uit van een groei van ‘slechts’ 40% per jaar. Wat de exacte groeisnelheid ook moge zijn, de datavolumes van grote commerciële bedrijven groeien zo hard dat database met een omvang van een petabyte (10^{15} bytes) naar verwachting in 2009 een realiteit zullen zijn.¹⁴⁸

Misschien nog belangrijker dan vast te stellen dat er sprake is van een exponentiële groei is de simpele constatering dat al die data op een of andere manier gemanaged moet worden. En hoewel commerciële spelers op de digitale archiveringsmarkt al jarenlang het tegendeel promoten, blijkt ook in de digitale wereld het proces van waardering en selectie¹⁴⁹ noodzakelijk. De noodzaak tot selectie in de digitale wereld wordt krachtig verwoord in het advies “Informatie: grondstof met toekomstwaarde” opgesteld door Raad van Cultuur en de Raad voor het Openbaar Bestuur:

‘De Raden kunnen hier kort over zijn: selectie moet, juist in een digitale wereld. Selectie heeft immers niet alleen volumereductie tot doel, maar ook het waarborgen van privacy van de burger. Daarnaast spelen het garanderen van duurzame toegankelijkheid – documenten moeten over een lange periode vindbaar en leesbaar zijn – en het in de hand houden van kosten die verbonden zijn aan beheer en toegankelijkheid, een belangrijke rol.’¹⁵⁰

Alles archiveren blijkt niet alleen heel duur te zijn, maar vergroot ook aanzienlijk de zoektijd, ook al beschikt men over geautomatiseerde information retrieval systemen. De archiveringsstrategie waarbij het uitgangspunt is dat alle informatie zonder selectie of vernietigingsbeleid wordt opgeslagen, wordt ook wel de *big-dump approach* genoemd, ‘archiving everything and hoping for the best’. De praktijk wijst echter uit dat dit resulteert in enorme volumes slecht geïndexeerde digitale documenten en arbeidsintensieve zoekacties.¹⁵¹ De enorme toename van de datavolumes stelt ons voor een aantal problemen:

- De opslagkosten stijgen, ondanks dat de kosten van opslaghardware per GB dalen. De reden hiervoor is dat datavolumes in omvang exponentieel toenemen.
- De dataopslag gebeurt inefficiënt en niet kosteneffectief. Data die weinig tot nooit worden gebruikt leggen beslag op dure opslagcapaciteit. Zo is bijvoorbeeld gemiddeld slechts 20% van de informatie die op diskdrives is opgeslagen origineel; 80% van de disk capaciteit wordt gebruikt voor de opslag van kopieën.

¹⁴⁷Jenkins, Tom; e.a. - *Enterprise Content Management Solutions* - pag.19

¹⁴⁸Ying, Chen; Shauchi, Ong - *Holistic information management solutions* - IBM (2005); pag. 2

¹⁴⁹Waardering is het toepassen van criteria voor het bewaren of het vernietigen van documenten. De criteria waarop deze waardering is gebaseerd worden bepaald door het belang van documenten met betrekking tot verantwoording, bewijs, kennis in een later stadium cultureel erfgoed. Selectie is de administratieve verwerking van de waardering

¹⁵⁰*Informatie: grondstof met toekomstwaarde* - pag. 14

¹⁵¹Reier, D. - *I Have to Show Them What?! E-Mail and the process of electronic discovery* - In: Information storage and security journal, juni 2005

- Vanuit verantwoordingsoogpunt wordt het steeds belangrijker om digitale informatie authentiek, betrouwbaar, integer en bruikbaar op te slaan.
- Het garanderen van digitale duurzaamheid. Oude, maar nog steeds relevante data zitten gevangen in niet langer in gebruik zijnde platforms en bestandsformaten.¹⁵²

Om deze problemen aan te pakken is een 'holistische aanpak' van de informatiehuishouding nodig, ook wel Information Lifecycle Management (ILM) genoemd. ILM wordt in de literatuur echter op verschillende wijzen geïnterpreteerd. Zo is er de wat abstracte interpretatie 'Information Lifecycle Management is a business-centric strategy for proactive management of information throughout its life, from its creation and use to its ultimate disposal.'¹⁵³ Daarnaast is er een benadering waarbij ILM zich met name richt op het zo kosteneffectief mogelijke gebruik van opslagmedia. '(ILM) is a sustainable storage strategy that balances the cost of storing and managing information with its changing business value. Information lifecycle management provides a practical methodology for aligning storage costs with business priorities.'¹⁵⁴ Er kunnen grofweg drie benaderingen van ILM worden onderscheiden.

1. Inhoudelijk, waarbij de focus ligt op de verantwoordelijkheid voor, toegang tot, beveiliging, beheer en vernietiging van bedrijfsinformatie. Het doel hiervan is compliance door middel van informatie- en record management
2. Technisch, waarbij de nadruk ligt op de IT-infrastructuur en hardwarebenadering van de distributie van data, databaseapplicaties en opslagmanagement.
3. Klantgericht, waarbij de nadruk ligt op het beheer van grote hoeveelheden data in datawarehouses en het leveren van business intelligence-informatie aan de gebruikers.¹⁵⁵

Naar mijn mening is het niet één, maar juist de combinatie van deze benaderingen die ILM tot een meerwaarde maken. ILM is een strategie waarbij informatie op basis van de waarde die het heeft voor het bedrijfsproces op een zo kosteneffectief mogelijke wijze geautomatiseerd wordt opgeslagen en/of vernietigd. In mijn scriptie ga ik dan ook uit van onderstaande definitie.

'ILM is a catch-all concept, an approach to information management that seeks to take into account the ways in which information changes over time, as well as the degree to which automation is required to manage, retain and migrate information across the enterprise as its value changes.'¹⁵⁶

ILM heeft betrekking op zowel de verschillende niveaus van de techniek als de inhoudelijke waarde van informatie en is daarmee onderdeel van de holistische informatiehuishouding. In relatie tot ECM heeft ILM met name betrekking op de technische en inhoudelijke benadering.

¹⁵² *Strategic archiving, using information lifecycle management to archive data more efficiently and comply with new regulations*, Sun Microsystems (2005); pag 3 http://www.sun.com/storagetek/white-papers/index.xml?displayPage=9&sort=modify_date&order=false&product=null

Reiner, David; Press, Gil; Lenaghan, Mike; Barta, David; Urmston Rich - *Information lifecycle management: the EMC perspective* - in: Data Engineering, 2004. Proceedings of the 20th international conference on data engineering (ICDE'04); pag. 804- 807

¹⁵³ Reiner, David e.a. - *Information lifecycle management: the EMC perspective* - pag. 804- 807

¹⁵⁴ *Information lifecycle management maturity model* - Sun Microsystems (2005); pag. 3 http://www.sun.com/storagetek/white-papers/ILM_Maturity_Model.pdf

¹⁵⁵ Short, James E - *Information Lifecycle Management: An Analysis of End User Perspectives* -The Information Storage Industry Center Graduate School of International Relations and Pacific Studies University of California (2006); pag. 4, 13

http://www.storagenetworking.org/pdfs/ISIC_ILM_WP_06-01_1.pdf

¹⁵⁶ Short, James E - *Information Lifecycle Management: An Analysis of End User Perspectives* - pag. 4

Die benaderingen, met name de inhoudelijke, zal ik hieronder bespreken. De business intelligence benadering van ILM heeft vooral betrekking op gestructureerde informatie en valt om deze reden buiten het bereik van ECM.

4.5.2. De inhoudelijk benadering van ILM

Zoals gezegd is ILM een strategie voor het effectief managen van informatie gedurende de levenscyclus. Een praktijk die overigens niet nieuw is en in de papieren omgeving van oudsher werd gedaan door archivariissen en andere informatiebeheerders.

Archivering en archiveringsbeleid blijken echter een groot probleem in de digitale bedrijfsomgeving, alleen al omdat gebruikers zich vaak niet bewust zijn van het nut en de noodzaak ervan. Te vaak wordt ervan uitgegaan dat er zoveel kopieën van een document rondzwerven dat, mocht het nodig zijn, er altijd wel één gevonden kan worden. Dit is de zogenoemde *safety by numbers*-aanpak van archiveren. Dat dit een riskante en voornamelijk inefficiënte manier van archivering is behoeft weinig uitleg. Maar het brengt nog andere problemen met zich mee. Een exponentieel groeiende hoeveelheid opgeslagen informatie wordt namelijk niet of nauwelijks geraadpleegd. Tegelijkertijd is onduidelijk wat de waarde van die informatie is en kan er niet af nauwelijks bepaald worden of zij bewaard moet blijven.

Hoewel ILM in principe betrekking kan hebben op alle vormen van informatie en data wordt er over het algemeen het meest gebruik van gemaakt in relatie tot ongestructureerde informatie afkomstig uit de werkprocessen. Vanuit deze invalshoek maakt ILM dan ook onderdeel uit van de algemene ECM-strategie. De realiteit is echter dat informatie vaak helemaal niet wordt gemanaged op basis van de rol die het speelt in de bedrijfsactiviteiten. Over het algemeen wordt informatie gemanaged op basis van technische zaken als opslagcapaciteiten en bandbreedte, en niet op basis van de waarde die zij heeft in de werkprocessen.¹⁵⁷

Door de digitalisering is een kloof ontstaan tussen het grotendeels gedigitaliseerde bedrijfsproces en de archieffunctie (de neerslag van het bedrijfsproces). Het in toenemende mate gedigitaliseerde bedrijfsproces en de neerslag hiervan worden beschouwd als twee aparte zaken, met als gevolg dat de twee zaken uit elkaar zijn gegroeid.¹⁵⁸ Dit heeft weer tot gevolg dat het merendeel van de uit de werkprocessen afkomstige informatie niet op basis van een vastgesteld archiefbeleid wordt gemanaged, maar alleen actief gemanaged wordt door degene die het heeft gecreëerd. ILM is een strategie om informatie en bedrijfsproces weer met elkaar te verbinden.¹⁵⁹ Om te begrijpen hoe de kloof kan worden gedicht is het belangrijk om inzicht te hebben in het fundamentele verschil tussen archivering in de analoge en de digitale omgeving, namelijk:

Analoge archivering

Archivering in de papieren omgeving gaat uit van de opeenvolging van een aantal handelingen:

1. Creatie, verwerving, identificatie, registratie
2. Distributie en gebruiksstadium, bewerking
3. Opslag in het archiefsysteem en onderhoud (inclusief ordening en raadpleging)
4. Waardering en selectieve vernietiging (van materiaal wat zijn waarde heeft verloren)
5. Archivistisch onderhoud en overdracht naar statisch archief

¹⁵⁷ Chen, Ying; Ong, Shauchi - *Holistic information management solution* - IBM research report (2005); pag. 3

¹⁵⁸ Koenen, Kees; Baak, Paul - *Het gebeugen als actieve kracht, de archieffunctie binnen de digitale overheid* - pag. 16, 19

¹⁵⁹ Davenport, Thomas H.; Cohen, Don - *Solving the information management puzzle: A life cycle approach* - (2005); pag. 3, 5
http://forbes.bitpipe.com/detail/RES/1147100054_420.html



Figuur 42: informatiebeheer in papieren omgeving

Het model gaat uit van gescheiden fasen binnen de levenscyclus die de archiefbescheiden één voor één doorlopen. Elke fase heeft een eigen begin- en eindpunt met daaraan gekoppeld duidelijk afgebakende bevoegdheden voor de betrokken personen. Als eerste is er de dynamische fase, waarin documenten worden opgemaakt en dossiers worden gevormd en geraadpleegd. Deze fase wordt gevolgd door de semi-statische fase, waarin dossiers nog maar weinig worden geraadpleegd, maar nog niet het moment hebben bereikt waarop ze vernietigd kunnen worden. De laatste fase is de statische fase, waarin de archiefstukken met cultuurhistorische waarde worden overgebracht naar een archiefbewaarplaats.

Het onderscheid tussen de verschillende fasen komt ook tot uiting in de bijbehorende beroepsprofielen, d.w.z. die van DIV (Documentaire Informatie Voorziening) voor de dynamische en semi-statische fase en daarop volgend die van archivaris voor de statische fase. Beide beroepsgroepen houden er hun eigen archiveringsstrategieën op na.¹⁶⁰

Digitale archivering: Records Continuüm model

Digitale archivering gaat uit van een nauwe verwevenheid van de uitvoering van het bedrijfsproces en het archiveringsproces. Een theoretische benadering die bij digitale archivering past is het 'Records Continuüm Model' van de Records Continuüm Research Group. Dit model stelt een geïntegreerd archiefbeleid voorop en verwerpt de lineaire opeenvolging van fasen. De archivering van de papieren neerslag van werkprocessen verschilt dan ook wezenlijk van die van de digitale neerslag. Waar de analoge strategie ervan uitgaat dat de levenscyclus bestaat uit verschillende van elkaar gescheiden fasen, gaat het records continuüm model ervan uit dat in de elektronische wereld deze scheidslijnen niet (haarscherp) zijn aan te geven.¹⁶¹ Zie voor een uitgebreide uitleg over het records continuüm model bijlage 3.

Digitalisering heeft tot gevolg dat informatie plaatsonafhankelijk is en in veelvoud beschikbaar. Meerdere personen kunnen gegevens dus gelijktijdig inzien en gebruiken. Het feit dat digitale informatie plaatsonafhankelijk is, betekent dat het bedrijfsproces en de archieffunctie geïntegreerd kunnen worden. Daarom dient archiefvorming in de digitale omgeving direct bij aanvang van het bedrijfsproces te beginnen.¹⁶² Het proces van waardering en selectie van digitale documenten vindt dan ook niet plaats na afloop van het bedrijfsproces, zoals met papieren documenten het geval is, maar gebeurt al voorafgaand aan de creatie van de documenten. Digitale archivering en het bijbehorende Information Lifecycle Management vereisen een grondige kennis van de werkprocessen binnen een organisatie.¹⁶³

¹⁶⁰ Waeyenbergh, S. - *Standaarden voor dynamisch archiefbeheer* - In Bibliotheek- en archiefgids 80 (2004) 2; pag. 10-17, alhier pag. 11

<http://www.vvbad.be/files/690-standaarden.PDF>

¹⁶¹ Syllabus VHIC faculty - *Leergang aankomend records manager* - pag. 56-57

¹⁶² Koenen; e.a. - *Het gebeugen als actieve kracht* - pag. 22

Arnold; e.a. - *De moderne informatiehuishouding van de digitale overheid* - pag. 28

¹⁶³ Colard, Robin; Zutphen, Frank van; Székely, Tom - *Personele en organisatorische gevolgen digitalisering backoffice* - ECORYS-NEI Arbeid & Sociaal Beleid (2002); pag. 24

http://www.rmconventie.nl/ozo/p_en_o-gevolgen_digi_backoffice.pdf

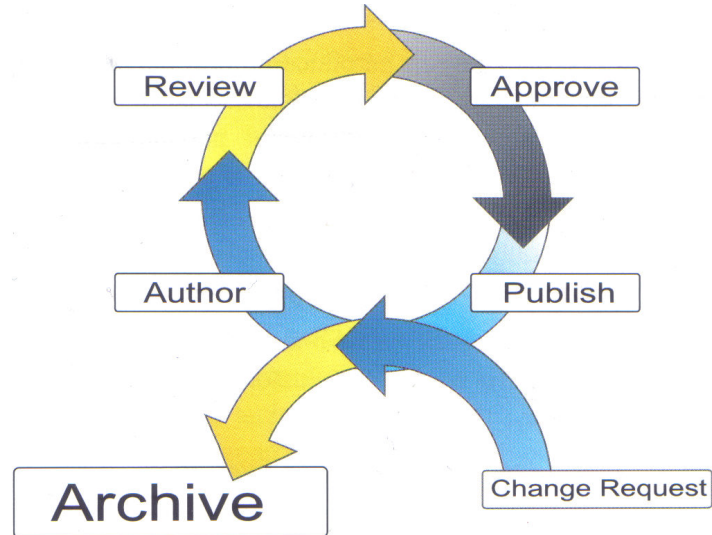
Een voorbeeld maakt het nogal abstracte idee van ILM wat inzichtelijker. Als we een sterk vereenvoudigde levensloop van een digitaal document nemen, dan kunnen over het algemeen de volgende stappen worden geïdentificeerd: creatie, beoordeling, goedkeuren, publicatie en vernietiging of conservatie.

Creatie: een document wordt gecreëerd, samengesteld en in samenwerking met anderen bijgesteld. Een document kan ook digitaal worden gecreëerd door het te scannen.

Beoordeling: op een zeker moment wordt het document ter beoordeling voorgelegd. Dit kan zich een aantal malen herhalen.

Goedkeuren/afkeuren.

Publicatie: het document wordt gepubliceerd of verzonden.



Figuur 43: levenscyclus van document

Archiveren/vernietigen: op een zeker moment wordt de bewaartermijn van een document geëvalueerd. Dan wordt het document of vernietigd of permanent opgeslagen.¹⁶⁴

ILM gaat uit van het principe dat ieder document een waarde heeft voor het bedrijfsproces waaraan het is verbonden en dat de waarde van een document gedurende zijn levensloop kan veranderen. Zo zijn er documenten die een lange periode van groot belang zijn, bijvoorbeeld langlopende contracten. Maar er zijn ook documenten die slechts een korte periode van belang zijn, zoals bestelbonnen. Op basis van de waarde die een document heeft in het bedrijfsproces en de periode dat een document deze waarde behoudt, kan er een bewaartermijn aan het documenttype worden gekoppeld. Die bewaartermijn moet natuurlijk in overeenstemming zijn met de wet- en regelgeving die op dat moment gelden.

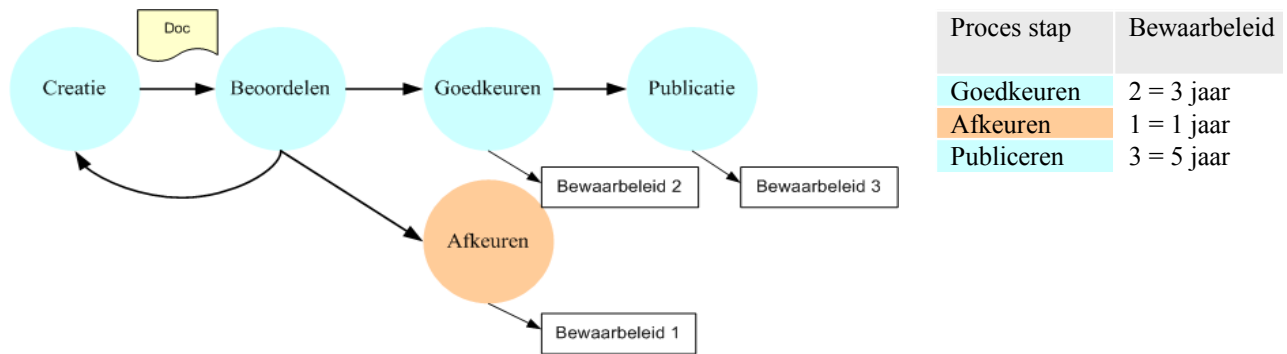
Om dit te kunnen realiseren moeten per bedrijfsproces de documenten worden geïdentificeerd die onderdeel uitmaken van het proces en moet per documenttype worden bepaald hoe lang deze bewaard moet worden. Op basis van de informatie uit de het workflow-systeem wordt vervolgens aan ieder document een bewaartermijn gekoppeld. In het volgende voorbeeld van een (vereenvoudigd) werkproces is bijvoorbeeld bepaald dat documenten die zijn goedgekeurd drie jaar moeten worden bewaard. Afgekeurde documenten hoeven maar een jaar te worden bewaard, gepubliceerde documenten daarentegen weer vijf jaar.¹⁶⁵

Orde op zaken, zaken op orde, rendement met informatiemanagement (versie 8) - Koenen Baak en Partners & VHIC (2003); pag. 14

http://digidiv.amsterdam.nl/div_nu_en_straks/elektronische_overheid/ozo.doc

¹⁶⁴ Jenkins, Tom - *Enterprise Content Management Technology* - pag. 107-114

¹⁶⁵ Ying Chen, e.a. - *Holistic information management solutions* - pag. 1, 6-7



Figuur 44: relatie werkproces bewaarbeleid

Door bij creatie aan alle documenten door middel van metadata een bewaartermijn te koppelen kunnen documenten automatisch gestructureerd en gecontroleerd worden vernietigd. Zo wordt voorkomen dat databases vollopen met documenten waarvan onduidelijk is of ze bewaard of vernietigd moeten worden.

4.5.3. De technische benadering van ILM

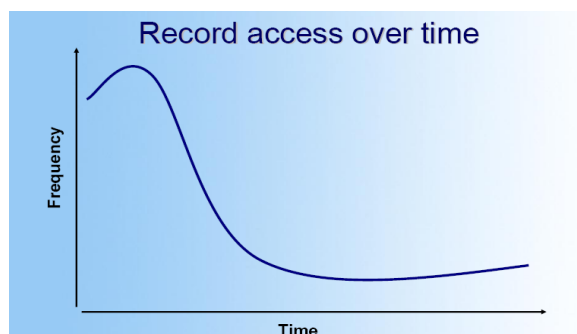
Zoals ik hierboven heb aangegeven, is het van belang om per document(type) een bewaartermijn vast te stellen. Gebeurt dit niet, wat nu vaak het geval is, dan ontstaat er een situatie waarin alle data gelijkwaardig worden behandeld, onafhankelijk van de waarde die ze hebben.¹⁶⁶

‘Industry surveys typically report that when IT managers are polled on how much of their primary disk storage has gone untouched within the last 6 months, many don’t know, and those who do know generally report that about 40% to 50% of their primary (tier 1) disk storage is being taken up by infrequently accessed data. The implication of this is two; storage costs are increased as firms must buy incremental capacity for new data, and IT is wasting resources by repeatedly backing up static data. Moreover, the alternative of manually relocating less-frequently accessed data from primary to secondary storage is considered prohibitively labor-intensive and error-rone.’¹⁶⁷

Van alle digitale data die worden gecreëerd, wordt ruim 90% opgeslagen. Hiervan wordt het overgrote deel na drie maanden niet of nauwelijks meer ingezien. Directe en snelle toegang tot die data is dan ook eigenlijk niet nodig.

Combineer dit gegeven met de toenemende trend om informatie voor langere tijd op te slaan en het moge duidelijk zijn dat een data-opslagstrategie noodzakelijk is om dit in goede banen te leiden.¹⁶⁸

Bij de technische benadering van ILM moeten dezelfde kernvragen gesteld worden als bij de inhoudelijke benadering. Namelijk, welke data moeten worden bewaard? Hoe lang moeten die data worden bewaard? En wat is de reden



Figuur 45: raadpleegfrequentie documenten

¹⁶⁶ Short James E. - *Information Lifecycle Management: An Analysis of End User Perspectives* - pag. 14

¹⁶⁷ Short James E. - *Information Lifecycle Management* - pag. 14

¹⁶⁸ *Information Lifecycle Management: Nieuw buzzword of doordachte strategie?* - In: VIP Vakblad voor documentmanagement, nr 5 juni (2004); pag 31-34, alhier pag. 31
<http://www.reflux-txt.nl/ILM.pdf>

dat we deze data bewaren? Vanuit een puur IT-technische benadering is het vrijwel onmogelijk om deze vragen goed te beantwoorden. De waarde van informatie wordt bepaald door de functie die ze vervult in een werkproces. IT-afdelingen zijn om die reden niet in staat om op een efficiënte en gebruiksvriendelijke manier vorm te geven aan een data-opslagstrategie. Het is voor hen vaak moeilijk te bepalen welke data thuishoren in primaire storage-voorzieningen en welke in goedkopere opslagvormen als nearline en offline.¹⁶⁹ Het resultaat is dat vrijwel altijd gekozen wordt voor de makkelijke weg, waarbij de oplossing voor meer data-opslag het bijplaatsen van meer schijven is. Maar schijven zijn duur in vergelijking met bijvoorbeeld tape. Daarnaast gaan data die op disk zijn opgeslagen meestal automatisch mee in de periodieke back-up, wat tijd en geld kost en bovendien onnodig is als die data niet of nauwelijks meer gebruikt wordt. Die data kunnen dan beter in een goedkoper opslagmedium zoals WORM-tape worden opgeslagen. Gegevens kunnen dus al na een korte periode afzakken in de digitale hiërarchie. De verdeling van data over de juiste opslagmedia is het uitgangspunt van de technische benadering van ILM.¹⁷⁰

‘Information lifecycle management is based on the idea of storing and managing data according to its value and purpose. It recognizes the inherent differences in the value and use of data sets. It makes use of tiers of storage to match archiving options with the appropriate cost, security, performance and accessibility requirements for particular data sets. Information lifecycle management takes a total systems approach that considers the full range of storage options — from disk to tape to optical.’¹⁷¹



Figuur 46: kostenvergelijking opslagmedia¹⁷²

Om dit probleem aan te pakken zal niet alleen de bewaartermijn, maar ook de opslaglocatie van documenten voorafgaand aan de creatie moeten worden bepaald. Dit betekent dat men vooraf moet vaststellen welke informatie toegankelijk moet zijn, op welke termijn zij toegankelijk moet zijn en voor hoe lang. Op basis hiervan worden de data weggeschreven naar disk of tape.

¹⁶⁹ Het gaat hierbij om verschillende opslagniveaus. *Nearline storage* onderscheidt zich van *online storage* door zijn lagere snelheid, maar kan de gegevens in tegenstelling tot *offline storage* zoals tape wel op ieder moment beschikbaar stellen.

¹⁷⁰ *Storage optimization, Information lifecycle management* - Sun Microsystems (2005); pag 1-4

http://www.sun.com/storage/tek/white-papers/Storage_Optimization_for_ILM.pdf

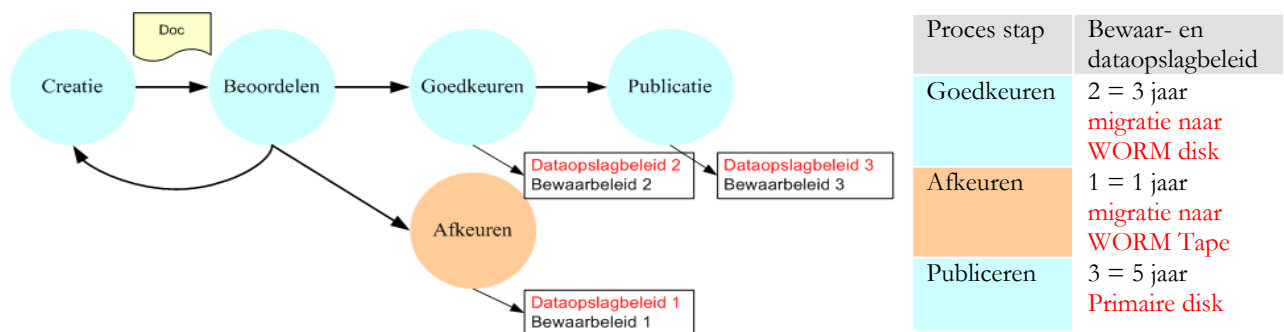
Information Lifecycle Management: Nieuw buzzword of doordachte strategie? - VIP Vakblad, pag. 31-33

¹⁷¹ *Strategic archiving, using information lifecycle management* - Sun Microsystems, pag 3

¹⁷² *Storage optimization, Information lifecycle management* - Sun, pag 3

Strategic archiving, using information lifecycle management - Sun, pag 6

Om deze theoretische beschrijving van ILM wat concreter te maken nemen we nogmaals het vereenvoudigde bedrijfsproces. Als ter beoordeling aangeboden documenten worden afgekeurd, dan hebben die waarschijnlijk nog maar weinig waarde voor het bedrijfsproces. Toch kan het om verantwoordingsredenen van belang zijn die documenten voor een bepaalde periode te bewaren. Er is echter geen reden om hiervoor dure opslagmedia te gebruiken. In de ILM-strategie kan bepaald worden dat documenten die afgekeurd zijn niet worden opgeslagen op relatief dure primaire of WORM-disks, maar op goedkopere WORM-tapes. De essentie van ILM is om voor alle datatypes te bepalen op welk moment in hun levenscyclus ze op welk type opslagmedium moeten worden opgeslagen.¹⁷³ Nemen we nog even het bedrijfsproces, dan kan per documenttype naast de bewaartermijn ook een data-opslagbeleid worden gedefinieerd.



Figuur 47: relatie werkproces ILM

4.5.4. ECM & ILM

Het concept van Information Lifecycle Management ligt in het verlengde van Enterprise Content Management en maakt daarmee deel uit van een ECM-totaalstrategie. Door de ECM-strategie van koppeling van de verschillende applicaties worden de informatiesilo's opgeheven. Het is hierdoor niet langer noodzakelijk om identieke data in verschillende losse applicaties op te slaan. Het ECM-principe van de enkelvoudige informatieopslag is dan ook essentieel voor een effectieve ILM-strategie. Op deze wijze wordt voorkomen dat identieke data meerdere malen wordt opgeslagen. En misschien nog wel belangrijker is dat data ook daadwerkelijk vernietigd kunnen worden. Door enkelvoudige opslag wordt voorkomen dat data na vernietiging nog op vijf andere locaties binnen de organisatie blijven bestaan. Het is de informatie uit de ECM-applicaties zoals document management, record management applicatie en workflow-systemen die nodig is voor een gefundeerd ILM-beleid.

¹⁷³ Ying, Chen; e.a - *Holistic information management solutions* - pag. 7-8

4.6. Conclusie theoretisch kader ECM

Het citaat hier onder, uit ‘The ECM Suites Report 2008, Comprehensive Product Evaluations’ van CMS Watch, is illustratief voor de redenen die over het algemeen worden gegeven voor het implementeren van een ECM oplossing.

‘The most common reasons that managers use to justify applying ECM technologies are:

- To bring order to the ever increasing volume of electronic documents
- To meet new legal or compliancy requirements regarding the management of information
- To reduce the amount of paper documentation within the organization
- To provide more standardized means of gathering and distributing data (e.g., via forms)
- To re-engineer business processes, and increase efficiencies
- To support business continuity requirements
- To obtain more value from costly investments in content
- To more consistently communicate to employees, partners, and customers.’¹⁷⁴

Hoe belangrijk ieder van de bovengenoemde reden ook mogen zijn, in mijn optiek zijn dit niet de juiste redenen voor het implementeren van een ECM-strategie. De redenen die hierboven worden genoemd zijn redenen die ook door middel van één of meerdere van de in hoofdstuk 3 besproken applicaties zoals document management, record management en content management gerealiseerd kunnen worden. Let wel, het gaat hier om losse applicaties zonder dat hiervoor een ECM-strategie nodig is. Zoals in dit hoofdstuk is aangegeven is het grote verschil tussen de losse applicaties en ECM dat de applicaties met elkaar zijn geïntegreerd. De toegevoegde waarde van ECM is dat hierdoor de informatiesilo’s worden afgebroken en ILM mogelijk wordt. Dit moet in mijn optiek dan ook de motivatie zijn voor het implementeren van een ECM-strategie.

Wat uiteindelijk de reden ook mag zijn dat er voor een ECM-strategie wordt gekozen, organisaties die de stap nemen moeten zich bewust zijn dat het gaat om de invoering van een ingrijpende en complexe strategie. Het beeld van ECM zoals dit in de scriptie is geschetst is gebaseerd op de literatuur. Het gaat hier om een theoretische benadering van wat ECM is en wat ECM zou kunnen betekenen voor een organisatie. Het is een ideaalschets en een ideaalschets gaat altijd uit van het principe ‘als ik het voor het zeggen had’. De realiteit blijkt zoals zo vaak weerbarstiger dan de theorie. Ondanks alle mooie verhalen over de mogelijkheden en voordelen van ECM blijkt de organisatiebrede adoptie van ECM tot heden nog beperkt:

‘To date, very few deployments of ECM technologies have spanned farther than a single department, be it the Marketing Department of a major retailer, the Research and Development group at a Pharmaceutical firm, or the Legal Department in a Fortune 500 firm. The typical ECM implementation targets groups of users who have very specific, document-centric needs to meet. To put it another way, few installations span the entire enterprise, whatever the ECM moniker might suggest.’¹⁷⁵

Ondanks deze kritische noot is dit zeker geen pleidooi tegen de ECM-strategie. Al was het alleen al om het feit dat niets doen vaak geen optie is. Zeker wanneer de bestaande systemen niet voldoende flexibiliteit bieden om aan de nieuwe vragen en eisen te voldoen. Om de complexe vraagstukken rondom het beheer van ongestructureerde informatie het hoofd te kunnen bieden is

¹⁷⁴ *The ECM Suites Report 2008 Comprehensive Product Evaluations* - CMS Watch, pag 3

¹⁷⁵ *The ECM Suites Report 2008* - pag. 4

het nodig om een duidelijke strategie te kiezen. ECM biedt deze. Integratie en metadata zijn daarbij de belangrijke steekwoorden. Alleen daarmee is de informatie-explosie te beheersen en is een productiviteitsverbetering te realiseren.

‘Nevertheless, most of the time enterprises turn to ECM to reduce costs and bring information overload under control. Indeed, with digital information mushrooming faster than most enterprises can manage it, ECM projects have become a kind of cost of doing business: Enterprises simply need to do something about information overload.’¹⁷⁶

¹⁷⁶ Ibidem, pag. 3

5. Casestudy

5.1. Inleiding

Na drie hoofdstukken waarin ik het theoretisch kader van ECM heb neergezet komen we nu toe aan het onderzoek. Het hoofdstuk is als volgt opgebouwd. Eerst zal ik uiteenzetten hoe het onderzoek is opgezet. Daarop volgt een korte beschrijving van de karakteristieken van ECM die zijn gedefinieerd. De beeldvorming rond ECM is daarmee compleet. Om een helder beeld te krijgen wat het handboek architectuur inhoudt zal ik hiervan een korte samenvatting geven. Dan volgt het daadwerkelijke onderzoek waarbij ik een vergelijking maak tussen de karakteristieken van ECM en het handboek architectuur. Per karakteristiek van ECM zal ik nagaan wat er in het handboek architectuur over wordt gezegd. Hierop volgt een reflectie waarin de onderzoeksresultaten in het perspectief van de vraagstelling worden geplaatst. Tot slot is er natuurlijk de conclusie waarin ik een antwoord op de onderzoeksvraag geef.

5.2. Onderzoeksopzet

In de voorgaande drie hoofdstukken heb ik een definitie gegeven van ECM, heb ik de belangrijkste technieken en applicaties besproken en zijn de voorwaarden en meerwaarde van ECM aan bod gekomen. Ik heb door middel van een literatuuronderzoek antwoord gegeven op de vraag wat ECM is. Uit dit theoretisch kader kan een aantal karakteristieken van ECM worden geabstraheerd. Het gaat hierbij om zowel *kenmerken* als *doelstellingen* van ECM. In totaal heb ik negen karakteristieken gedefinieerd op basis waarvan ik een vergelijking maak tussen ECM en het handboek architectuur van de Gemeente Amsterdam.

Het handboek architectuur is een door het college van B&W vastgestelde blauwdruk voor de gemeentelijke informatievoorziening van de Gemeente Amsterdam.¹⁷⁷ In de architectuur wordt op hoofdlijnen weergegeven hoe de verschillende componenten en initiatieven in de organisatie en de informatievoorziening in elkaar grijpen, zowel functioneel als technisch. Het handboek architectuur vormt het bestemmingsplan voor de toekomstige organisatie en informatievoorziening van Amsterdam. In dit plan wordt circa vier jaar vooruit gekeken.¹⁷⁸ ‘Het geeft via concrete grondslagen, modellen, standaarden en beheerprocedures een beschrijving van de inrichting van informatievoorziening en de “bouwkundige” kaders waarbinnen gewerkt moet worden (inrichten).’¹⁷⁹

Voor de vergelijking van de karakteristieken is de volgende standaard notatiemethode gebruikt.

¹⁷⁷ Onderdelen uit commissie flap (BD2007-004821) van 26 september 2007 :

‘De beschreven methodische aanpak van het Handboek Architectuur (met grondslagen, modellen, standaarden, werken onder architectuur) verplicht te stellen voor de informatievoorziening en ICT van diensten en bedrijven. Opdracht te geven aan diensten en bedrijven het Handboek Architectuur van toepassing te laten zijn op alle:

- A. informatievoorziening en ICT-projecten, zowel vernieuwing als vervanging (voor zover dit niet gepaard gaat met desinvesteringen);
- B. projecten die herontwerp van processen in het kader van Beter Presteren betreffen.

Afwijkingen door diensten en bedrijven op het Handboek Architectuur dienen beargumenteerd via de Stuurgroep Informatievoorziening aan de Commissie Informatievoorziening te worden voorgelegd. De Commissie Informatievoorziening geeft bij de afwijking een advies en legt dit ter besluitvorming voor aan het College.

[...]Het Handboek Architectuur (met grondslagen, modellen, standaarden, werken onder architectuur) als norm toe te voegen aan het vigerende Beleid grote ICT-projecten.’

(<http://biodata.asp4all.nl/andreas/2007/09012f978031930b/09012f978031930b.html>)

¹⁷⁸ *Handboek architectuur, de samenhang in organisatie en informatievoorziening van de Gemeente Amsterdam* - (1-2)

¹⁷⁹ *Handboek architectuur* - pag. 2-2

Karakteristiek van ECM	
Citaten en beschrijving van wat er over het betreffende ECM karakteristiek wordt geschreven in handboek architectuur	
Conclusie	Conclusie van vergelijking ECM en handboek architectuur

Bij het onderzoek heb ik gekeken op welke punten ECM en het handboek architectuur overeenkomen en op welke punten dit mogelijk niet het geval is. De hypothese is dat als ECM en het handboek architectuur verschillende termen voor hetzelfde principe zijn, de voor ECM gedefinieerde karakteristieken ook in het handboek tot uiting moeten komen. Op basis van deze vergelijking zal ik een reflectie geven waarin ik antwoord geef op de onderzoeksvraag.

We beginnen nu eerst met de negen karakteristieken van ECM waarbij per karakteristiek een korte beschrijving zal worden gegeven.

5.3. Karakteristieken van ECM

Kenmerken Enterprise Content Management

ECM is een strategie en niet een specifieke technologie.¹⁸⁰

Zowel in de definitie van Forrester als die van de AIIM wordt vooropgesteld dat ECM niet zozeer een specifieke technologie is maar een strategie voor het managen van ongestructureerde informatie. ‘While there are ECM technologies, more importantly, ECM is an ongoing and evolving strategy for maximizing how your content is to be used.’¹⁸¹ Dat ECM meer is dan een technologie wordt ook onderschreven door Dr. Ulrich Kampffmeyer¹⁸² ‘With all the comprehensive claims and all the countless components of Enterprise Content Management, at the end of the day ECM is a vision, a strategy, or even a new industry, but it is not a closed system solution or a distinct product.’¹⁸³

ECM heeft betrekking op het beheer van ongestructureerde informatie.¹⁸⁴

Dat ECM betrekking heeft op ongestructureerde informatie komt in alle in hoofdstuk 2 opgevoerde definities duidelijk naar voren. In deze scriptie ga ik uit van de definitie van de AIIM die ECM als volgt definieert: ‘Enterprise Content Management (ECM) is the technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists.’¹⁸⁵

ECM heeft als uitgangspunt ‘enkelvoudige opslag, meervoudig gebruik van informatie’.

¹⁸⁶ De ECM-strategie heeft als uitgangspunt dat content slechts éénmaal wordt opgeslagen, maar in verschillende applicaties wordt gebruikt. Dit is alleen mogelijk als de applicaties aan elkaar zijn gekoppeld. In de definitie die Forrester geeft van ECM wordt het als volgt verwoord: ‘ECM must be a strategy for:[...] how to integrate the many content repositories within an enterprise’¹⁸⁷ Doel van dit principe is het reduceren van meervoudige bestanden die identieke gegevens bevatten. Door applicaties te koppelen en enkelvoudige opslag als uitgangspunt te nemen is het niet langer nodig om dezelfde informatie in verschillende repositories op te slaan.

ECM gaat uit van het koppelen van de verschillende applicaties.¹⁸⁸

Doel is de verschillende systemen te integreren tot één virtueel systeem. Vanuit een technische optiek wordt ECM dan ook beschouwd als een verzameling onderling verbonden applicaties voor het opnemen, verwerken, beheren, gebruiken en distribueren van content. In de ideale situatie merkt de gebruiker niet of er gewerkt wordt met één systeem, dan wel een verzameling van met elkaar verbonden systemen. Het koppelen van de verschillende applicaties is een randvoorwaarde voor het hierboven genoemde uitgangspunt ‘enkelvoudige opslag, meervoudig gebruik van informatie’.

¹⁸⁰ Zie voor nadere uitleg hoofdstuk 2 ‘Een definitie van ECM’

¹⁸¹ <http://www.aiim.org/about-ecm.asp>

¹⁸² Dr. Ulrich Kampffmeyer is een voormalig lid van de raad van bestuur van AIIM Europa

¹⁸³ Kampffmeyer Ulrich - *ECM enterprise content management* - pag. 2

¹⁸⁴ Zie voor nadere uitleg hoofdstuk 2 ‘Een definitie van ECM’

¹⁸⁵ <http://www.aiim.org/about-ecm.asp>

¹⁸⁶ Zie voor nadere uitleg hoofdstuk 4 ‘ECM: voorwaarden en meerwaarde’

¹⁸⁷ From Forrester’s [Topic Overview: Enterprise Content Management](#)

¹⁸⁸ Zie voor nadere uitleg hoofdstuk 4 ‘ECM: voorwaarden en meerwaarde’

ECM gaat uit van generieke applicaties.¹⁸⁹

De belangrijkste ECM-applicaties, te weten document management, record management, content management, collaboration tools en workflow, zijn generieke applicaties. Deze applicaties voor de bewerking en het beheer van ongestructureerde informatie kunnen in verschillende werkprocessen voor verschillende afdelingen in een organisatie ingezet worden. Zo is een DMS in de basis een generieke applicatie die het mogelijk maakt om versies, varianten en kopieën van documenten beheersbaar te maken. Workflow is een generieke applicatie voor documentrotatie en processturing, in tegenstelling tot het beheer van gestructureerde informatie, die veelal is opgeslagen in databases en wordt bewerkt en beheerd met behulp van specifieke applicaties. Deze systemen zijn vaak ontwikkeld voor een specifieke taak en zijn eigenlijk alleen daarvoor geschikt.

ECM gaat uit van standaardiseren van metadata.¹⁹⁰

Omdat ECM uitgaat van de koppeling van verschillende systemen, is het van groot belang de gebruikte metadata te standaardiseren. Standaardisering van metadata is essentieel voor de ontsluiting, toegankelijkheid en uitwisseling van documenten, dossiers en/of procesinformatie. Uitwisselbaarheid van informatie kan in veel gevallen betekenen dat er doelmatiger gebruik gemaakt kan worden van de beschikbare informatie. Daarnaast is (standaardisering van) metadata van belang voor het borgen van de duurzame ontsluiting van informatie. Juist met het oog op digitale duurzaamheid is de standaardisering van metadata van groot belang

Doelstellingen Enterprise Content Management

ECM heeft als doel het openbreken van informatiesilo's.¹⁹¹

Met informatiesilo's worden de verschillende applicaties en databases bedoeld waarin de informatie zit opgeslagen. We spreken van informatiesilo's en verticale applicatie-architectuur op het moment dat applicaties niet aan elkaar zijn gekoppeld en de informatie die erin is opgeslagen dus 'gevangen' zit in de verschillende databases. Belangrijkste nadeel van informatiesilo's is dat de informatie alleen te gebruiken is binnen de bewuste applicatie en niet daarbuiten. In de praktijk zal echter voor het uitvoeren van een werkproces over het algemeen niet slechts gebruik worden gemaakt van één, maar van meerdere applicaties. Gevolg is dat dezelfde informatie in meerdere applicaties moet worden ingevoerd en opgeslagen. Doel van ECM is door het koppelen van de systemen de informatiesilo's open te breken en daardoor alle aanwezige informatie organisatiebreed te ontsluiten en toegankelijk te maken.

ECM heeft als doel gedurende de gehele levenscyclus grip te hebben op ongestructureerde informatie.¹⁹²

De exponentiële groei van de omvang van ongestructureerde informatie binnen organisaties stelt ons voor een aantal problemen. Om deze problemen aan te pakken is een 'holistische aanpak' nodig, ook wel Information Lifecycle Management (ILM) genoemd. ILM is een strategie waarbij informatie op basis van de waarde die het heeft voor het bedrijfsproces op een zo kosteneffectief

¹⁸⁹ Zie voor nadere uitleg hoofdstuk 3 'De vijf componenten van ECM'

Zie voor nadere uitleg hoofdstuk 2 'Een definitie van ECM'

¹⁹⁰ Zie voor nadere uitleg hoofdstuk 4 'ECM: voorwaarden en meerwaarde'

¹⁹¹ Zie voor nadere uitleg hoofdstuk 4 'ECM: voorwaarden en meerwaarde'

¹⁹² Zie voor nadere uitleg hoofdstuk 4 'ECM: voorwaarden en meerwaarde'

mogelijke wijze geautomatiseerd wordt opgeslagen en/of vernietigd. Of zoals de AIIM het verwoordt: 'Central to this strategy are the tools and technologies of ECM, which manage the complete lifecycle of content, birth to death.'¹⁹³

ECM gaat niet slechts om het 'managen' van ongestructureerde informatie, maar heeft tot doel het bedrijfsproces optimaal te ondersteunen.¹⁹⁴

Door de ECM-strategie van de koppeling van systemen en het daarmee openbreken van de informatiesilo's wordt niet slechts een deel, maar **alle** ongestructureerde informatie zichtbaar en bruikbaar. 'It's not enough to "manage" content. Of course, the ability to access the correct version of a document or record is important, but companies must go further. Content must be managed so that it is used to achieve business goals.'¹⁹⁵

¹⁹³ <http://www.aiim.org/about-ecm.asp> (18-03-2008)

¹⁹⁴ Zie voor nadere uitleg hoofdstuk 2 'Een definitie van ECM'
Zie voor nadere uitleg hoofdstuk 4 'ECM: voorwaarden en meerwaarde'

¹⁹⁵ <http://www.aiim.org/about-ecm.asp> (18-03-2008)

5.4. Samenvatting handboek architectuur Gemeente Amsterdam

Doel handboek architectuur

Het moge duidelijk zijn dat hier geen ruimte is voor een gedetailleerde beschrijving van het handboek architectuur. Ik zal dan ook volstaan met een samenvatting op hoofdlijnen van het in totaal 271 pagina's tellende werk.

De reden om het handboek op te stellen is volgens de auteurs helder. 'Omdat we zonder een gemeenschappelijke visie en een gedeelde blauwdruk van de gemeentelijke organisatie en informatievoorziening niet (meer) kunnen voldoen aan wat van ons verwacht wordt. Daarbij gaat het, van buiten naar binnen redenerend, om drie drijfveren.

- Burgers en bedrijven verlangen een overheid die niet naar de bekende weg vraagt, die klantgericht is, zich niet voor de gek laat houden, die weet waar ze het over heeft, die haar zaken op orde heeft en die niet meer uitgeeft dan nodig is.
- De gemeentelijke diensten en stadsdelen hebben behoefte aan overzicht, samenhang en toetsingskaders op het terrein van informatievoorziening en ICT.
- En ten slotte is, heel concreet, de concernbrede architectuur een van de randvoorwaarden voor de realisatie van het programma BasisRegistraties & ICT-infrastructuur.¹⁹⁶

Architectuur

Het gaat hier om het handboek architectuur, maar wat is nu eigenlijk precies een architectuur? 'Een architectuur geeft op hoofdlijnen weer hoe de verschillende componenten en initiatieven in de organisatie en de informatievoorziening in elkaar grijpen, zowel functioneel als technisch. Een architectuur beschrijft op samenhangende wijze een gewenste situatie. Dit Handboek Architectuur vormt daarmee het bestemmingsplan voor de toekomstige organisatie en informatievoorziening van Amsterdam. We kijken daarbij circa 4 jaar vooruit.'¹⁹⁷

De uitgangspunten van het handboek zijn: koppelen en delen, samenwerking en samenhang. 'We kunnen onze business-doelstellingen alleen halen als we met één loket naar buiten treden, processen delen, gegevens delen en middelen delen. En om alle organisaties op elkaar aan te sluiten zijn er voorzieningen "in het midden" nodig om de processen te regisseren en de gezamenlijke gegevens onderling uit te wisselen.'¹⁹⁸

Hieruit blijkt al dat het handboek architectuur verschillende niveaus onderkent (o.a. processen, gegevens en middelen). 'Om onderscheid te kunnen maken tussen de technische en de meer organisatiegerichte aspecten kent vrijwel elk architectuurmodel een indeling in lagen of niveaus. Het is vervolgens de kunst om de samenhang en de onderlinge afhankelijkheid van de verschillende niveaus zo goed mogelijk in beeld te brengen. Een uitgangspunt op het ene niveau moet op mogelijke consequenties voor de andere lagen beoordeeld kunnen worden.'¹⁹⁹

¹⁹⁶ *Handboek architectuur* - pag. management samenvatting -1

¹⁹⁷ *Ibidem*, pag. 1-2

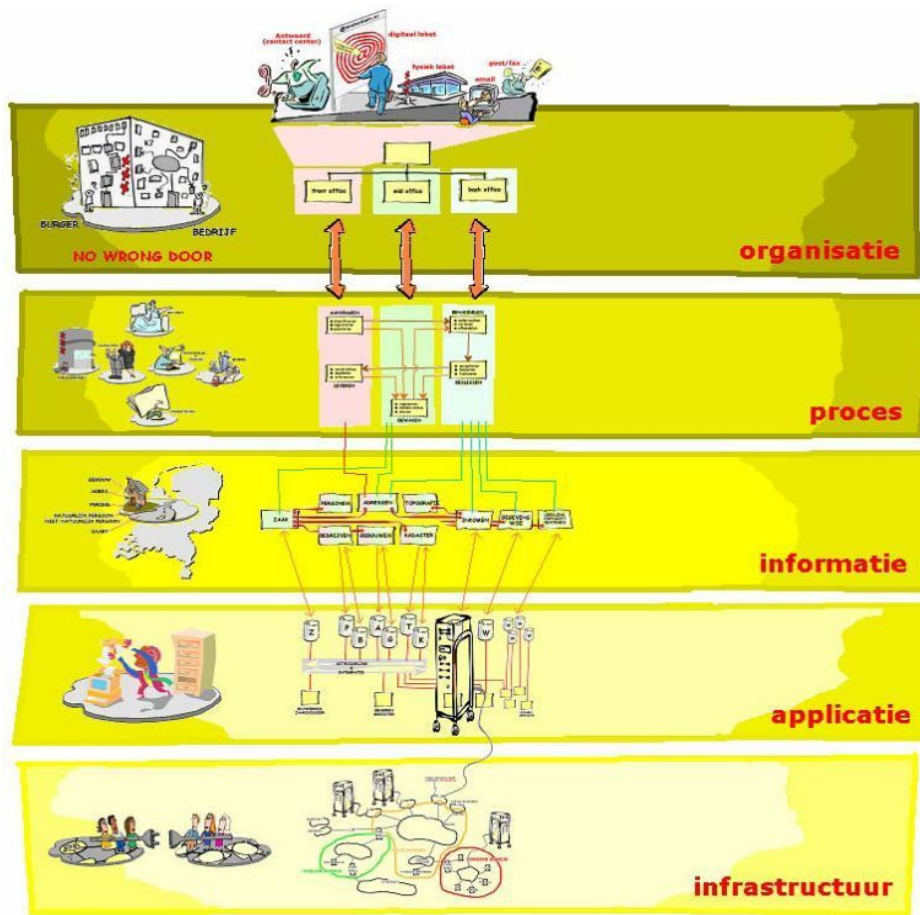
¹⁹⁸ *Ibidem*, pag. management samenvatting - 2

¹⁹⁹ *Ibidem*, pag. management samenvatting - 2

Een artist impression van de architectuur

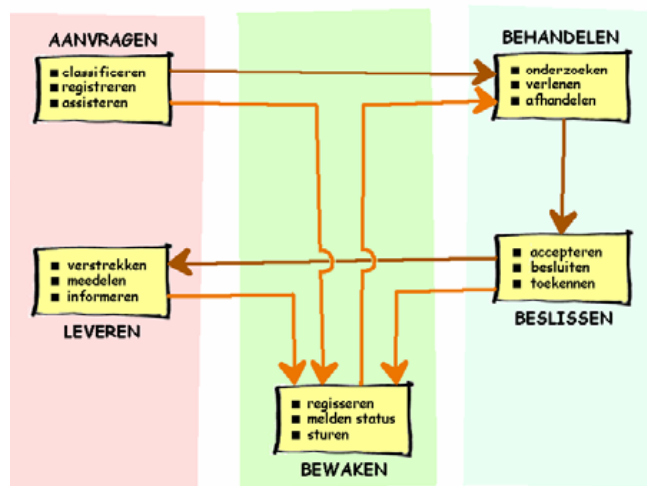
- laag 1: organisatie 'Het verhaal in grote lijnen. De redenering start, als vanzelfsprekend, bij de burger en de ondernemer. Het is onze uitdrukkelijke wens de klant niet meer van het kastje naar de muur te sturen maar een organisatie te zijn die open is en transparant. De klant weet ons te vinden, kan ons gemakkelijk bereiken, wordt vriendelijk te woord gestaan en professioneel geholpen. Hij is zelfredzaam, niet meer zelf op zoek naar het juiste loket: elk loket is juist. En als de dienstverlening ook digitaal zou kunnen, is die ook daadwerkelijk digitaal beschikbaar. Alleen nog naar het stadsdeelkantoor of het dienstloket als dat echt niet anders kan.
- laag 2: proces Voor onze processen ligt daar de uitdaging, we moeten beter presteren. Recentelijk is een indeling in een zestal gemeentelijke hoofdprocessen in zwang geraakt. Vier primaire processen, waaronder dienstverlening en handhaving, die elk bestuurd en ondersteund moeten worden. In aparte bijeenkomsten worden die processen opnieuw ontworpen.
- business alignment Er is een samenhang tussen de organisatielaag en de proceslaag. Op organisatieniveau is aan de orde wat we willen leveren en op welke manier we dat arrangeren, op procesniveau vragen we ons af hoe we die dienstverlening dan het beste kunnen inrichten. De bovenste twee lagen van het architectuurmodel zijn daarmee gekarakteriseerd. Voor het volledige beeld is het goed de redenering door te zetten naar de drie lagen die daaronder komen, om de verbinding te maken met de vereisten die deze organisatie- en procesdoelen stellen aan de informatievoorziening en de techniek van de ICT, kortom om de business alignment te maken.
- laag 3: informatie Bij het uitvoeren van de processen is informatie benodigd over de klant, de zaak en het product. Een aantal gegevens is bij vrijwel elk klantcontact vereist en daarom vastgelegd in zogenaamde basisregistraties. Er is op landelijk niveau een zestal gedefinieerd: persoonsgegevens, adressen, gebouwen, bedrijven, percelen en topografie.
- laag 4: applicatie De gegevens uit de basisregistraties zijn inmiddels goed gedefinieerd en gestandaardiseerd, maar moeten doorgevoerd worden in alle gemeentelijke applicaties en systemen. Dat is gemakkelijker gezegd dan gedaan. Zo moeten we in Amsterdam een slordige 150 persoonsadministraties terugbrengen tot één basisregistratie.
- laag 5: infra Op de onderste laag van het model, die van de technische infrastructuur, gaat het erom alle systemen fysiek met elkaar te verbinden. En open te stellen voor elkaar en, waar nuttig en noodzakelijk, voor de buitenwereld. Een en ander met de grootste zorg voor de beveiliging.
- de vijf lagen We kunnen de resultaten tot nu toe projecteren in het vijflagenmodel. [...]In het handboek worden deze abstracte beelden vervolgens vertaald in concrete modellen, diagrammen, definities, schema's en toelichtingen. De hoofdstukindeling volgt de vijf lagen van dit architectuurmodel.²⁰⁰

²⁰⁰ Ibidem, pag. management samenvatting - 5



vanuit het proces

‘We gaan één stap dieper in op het model. In principe moet een redenering op elke laag kunnen starten, ter illustratie wordt onderstaand de proceslaag als startpunt genomen. Dat begint met een inrichting op het hoogste abstractieniveau. Voor het proces dienstverlening kan dat er bijvoorbeeld als volgt uitzien.



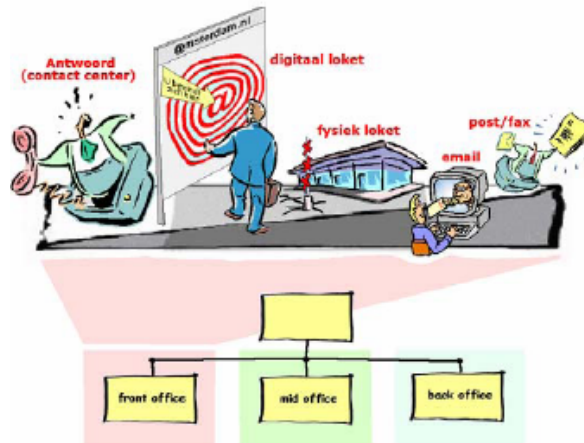
voortgangs-
bewaking

Elk dienstverleningsproces begint met het te woord staan van de klant, de aanvraag wordt op de een of andere manier geregistreerd en het verzoek wordt in behandeling genomen. Vaak zal er sprake zijn van een beoordelings- of beslissingsmoment in de procedure en als alles goed gaat wordt het gevraagde op

tijd en naar tevredenheid van de klant geleverd. Dat laatste, de bewaking van de voortgang van het proces, is cruciaal voor de eerder geformuleerde ambitie met betrekking tot de klanttevredenheid. Hier start de discussie over de zogenoemde mid office, die hieronder nader wordt toegelicht.

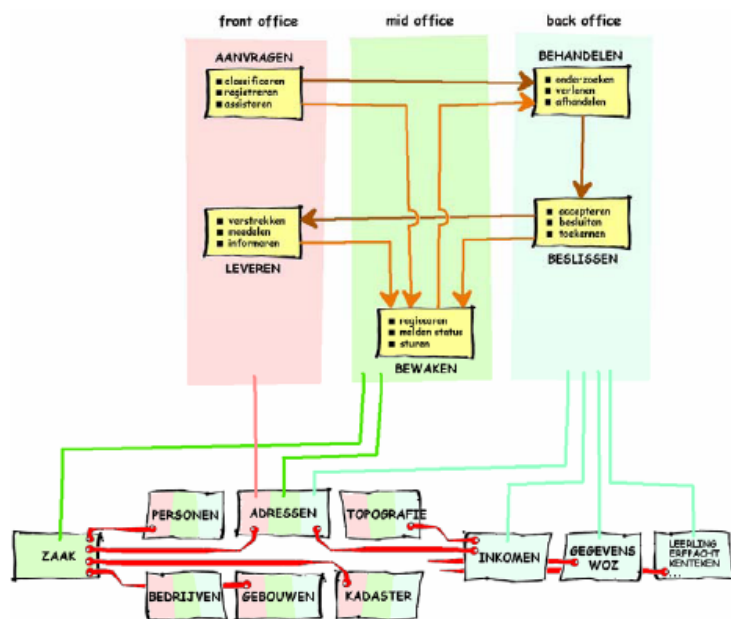
Channel management

Op naar de organisatielaag. De interactie met de klant kan via meerdere kanalen geschieden. In de eerste plaats is er het fysieke loket, de balie, en de traditionele communicatie via de gewone post. Daar zijn andere, elektronische kanalen naast gekomen: de website Amsterdam.nl, de correspondentie via email, het digitale loket en het telefonische contact center (Antwoord/14020). De vraag is aan de orde welke producten en diensten wij bij voorkeur via welk kanaal aan de man willen brengen. We moeten aan channel management gaan doen.



Informatie

En dan vanuit de proceslaag naar beneden: welke informatie is benodigd bij de uitvoering van de processen? De gegevens waarover het in de informatielaag gaat, kunnen in drie categorieën worden verdeeld.



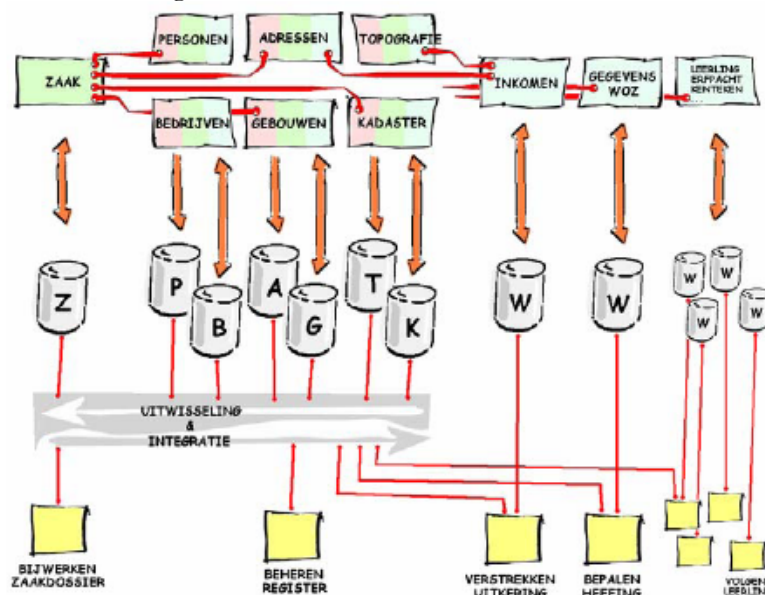
Basisregistraties; In de eerste plaats gaat het om de zes basisregistraties, het hart van de gegevenshuishouding. Deze zijn in vrijwel elke fase van het primaire proces benodigd, dus zowel in de front en de mid office als in de back office.

Bedrijfsgegevens; Ten tweede gaat het om gegevens die specifiek zijn voor de aard van het product of de dienst: inkomens- en vermogenspositie (DWI), WOZ-gegevens van de eigenaar van het onroerend goed (DBGA), in-, door- en uitstroom van leerlingen (DMO), erfpacht op het perceel (OGA), kenteken van de auto (Stadstoezicht), enzovoort. Deze gegevens worden vanwege de hoedanigheid met name in de back office gebruikt.

Zaakdossier; De derde categorie betreft de zaakgegevens, benodigd voor een adequate bewaking van de voortgang van het proces, en daarmee specifiek voor de mid office. Steeds sterker dient zich de behoefte aan de vereiste gegevens in een apart zaakdossier te registreren. Het zogenaamde datamodel van deze zaakgegevens is een van de meer technische diagrammen die in het handboek zijn opgenomen. Zo'n model laat zien hoe nauwkeurig de onderlinge relaties van de gegevensveldjes zijn gedefinieerd en hoe de verbindingen met de andere categorieën lopen.

Applicaties

De gegevens zijn hiermee op papier bepaald maar moeten nog wel in techniek worden omgezet. Op de applicatielaag spreken we van databases. Verder is er functionaliteit benodigd om deze databases te beheren en te bewerken.



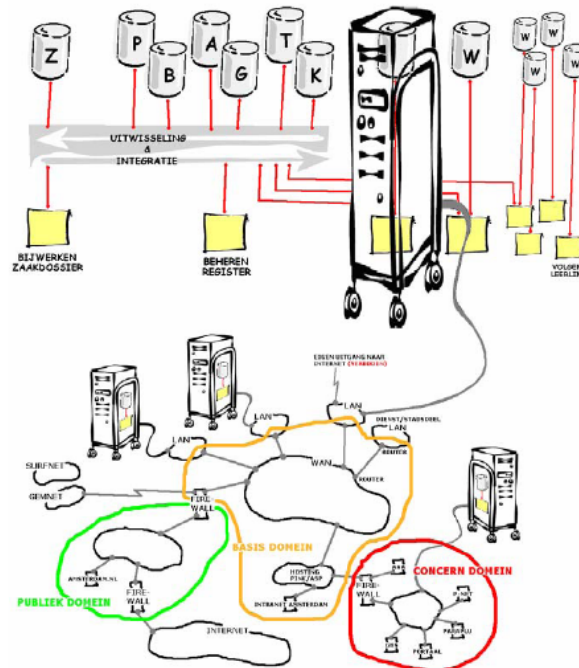
SGA en webservice

Van oudsher is sprake van systemen waarin functionaliteit en database exclusief voor elkaar zijn bestemd (rechts in de afbeelding), maar onderlinge uitwisseling en meervoudig gebruik van functionaliteit (links in de afbeelding) wordt steeds belangrijker, de zogenaamde Service Gerichte Architectuur (SGA). De applicatie hoeft zelfs niet eens meer in huis te draaien maar kan via internet worden aangeboden (webservices).

infrastructuur

Database en functionaliteit samen noemen we een applicatie of een systeem. De systemen draaien gewoonlijk in een lokaal netwerk (LAN), dat weer is gekoppeld aan een grotere ring (WAN). Deze infrastructuur heet in Amsterdam E-NET en is ingedeeld in verschillende domeinen met een toenemende mate van beveiliging. Extra voorzieningen (Amsterdam.nl en Portaal Amsterdam) zorgen voor de presentatie op het web. Om de uitwisseling van gegevens en

services mogelijk te maken wordt aan een gemeenschappelijke Service Bus Amsterdam gewerkt.



Heen en
weer

Het verhaal dat bij de burger begint, is zo uiteindelijk vertaald in techniek. Behalve van boven naar beneden kan de redenering ook andersom lopen. Op die manier ontstaat er consistentie in de beelden en een toetsing aan de praktijk. De architectuur is een levende, dynamische systematiek om grip te krijgen op de werkelijkheid.²⁰¹

²⁰¹ Ibidem, pag. management samenvatting - 6-9

5.5. Vergelijking karakteristieken ECM & handboek architectuur

ECM is een strategie en niet een specifieke technologie.

De noodzaak van een architectuur van de Amsterdamse organisatie en informatievoorziening wordt door de auteurs als volgt gedefinieerd. ‘...omdat we zonder een goede verbinding tussen de vereisten van de organisatie en de primaire processen enerzijds en de vormgeving van de informatievoorziening anderzijds (business alignment) niet kunnen leveren wat van ons verlangd wordt.’²⁰² Het handboek architectuur kan beschouwd worden als de strategie voor het realiseren van deze verbanden tussen de primaire processen en de informatievoorziening. ‘De strategie voor de inrichting van de informatiearchitectuur kent vier prioriteiten:

1. Implementeren van wettelijke afspraken rond basisregistraties
2. Benutten van potentie van kernadministraties
3. Het toewerken naar één standaard voor een zaakdossier en één zakenmagazijn.
4. Standaardisering²⁰³

Daarnaast moet het handboek worden beschouwd als een hulp- en communicatiemiddel gericht op gebruik door professionals die zich bezighouden met de organisatie en informatievoorziening. ‘Het geeft via concrete grondslagen, modellen, standaarden en beheerprocedures een beschrijving van de inrichting van informatievoorziening en de ‘bouwkundige’ kaders waarbinnen gewerkt moet worden (inrichten).’²⁰⁴ Het handboek is dan ook in de eerste plaats bedoeld als ‘toetskader en richtlijn voor nieuwe ontwikkelingen, projecten en vervanging van bestaande systemen. In het handboek zijn standaarden en richtlijnen opgenomen om de samenwerking vorm te kunnen geven.’²⁰⁵

Een van de uitgangspunten van het handboek is standaardisering (prioriteit nr. 4 in het lijstje hierboven). Dat het handboek architectuur zich niet richt op een specifieke technologie of systeem komt op dit punt duidelijk tot uitdrukking. ‘Standaardisering wordt meestal geassocieerd met de keuze voor één pakket voor een bepaalde functie die gemeentebreed wordt toegepast. Dit is niet juist. Zonder aanvullende standaardisering op proces en inhoud kan een dergelijk besluit zelfs contraproductief werken, ook in het geval dat een dergelijk besluit concernbreed wordt toegepast. Dat maakt standaardisering behalve een onderwerp voor informatici ook een organisatorisch onderwerp. Het organisatorisch vermogen van de Gemeente Amsterdam om tot standaardisering van processen en inhoud te komen is meer bepalend dan de vraag of architecten één of meerdere pakketten tot standaard kunnen benoemen.’²⁰⁶

Conclusie Zowel ECM als het handboek architectuur zijn een strategie. In beide gevallen wordt niet uitgegaan van een specifieke technologie of specifieke systemen.

²⁰² Ibidem, pag. 2-1

²⁰³ Ibidem, pag. 6-1

²⁰⁴ Ibidem, pag. 2-2

²⁰⁵ Ibidem, pag. 2-4

²⁰⁶ Ibidem, pag. 2-8

ECM heeft betrekking op het beheer van ongestructureerde informatie.

Op het moment dat het handboek architectuur spreekt van informatievoorziening gaat het in de eerste plaats over de basisregistraties en kernadministraties. ‘In de informatiearchitectuur geldt de grondslag van “éénmalige opslag, meervoudig gebruik” voor basisregistraties en kernadministraties. Die moeten overal beschikbaar zijn binnen processen en systemen (applicaties) waar zij relevant zijn. Dit kan in de vorm van raadpleegbare informatie, maar ook als onderdeel van dataverzamelingen die in specifieke deelprocessen waardevol zijn.’²⁰⁷ Zoals ik hierboven al heb aangegeven gaat het bij de basisregistraties om gestructureerde informatie met betrekking tot persoonsgegevens, adressen, gebouwen, bedrijven, percelen en topografie. Naast de basisregistraties is er een aantal andere administraties binnen het concern Amsterdam die intensief worden gebruikt, de zogenaamde kernadministraties. Voorbeelden hiervan die worden genoemd zijn ‘Onderwijs en jeugd’, ‘Bestuursinformatie’ en het zogenaamde ‘Zakenmagazijn’. Interessant in dit opzicht is dat er in het handboek architectuur van uit wordt gegaan dat de kernadministratie ‘Zakenmagazijn’ uit zowel gestructureerde als ongestructureerde informatie bestaat. ‘In het zakenmagazijn worden op centraal niveau alle relevante gegevens over een zaak opgeslagen en beschikbaar gesteld. Het gaat hierbij onder meer om informatie over de status van procedures, degene die een verzoek heeft ingediend, het organisatieonderdeel dat het verzoek behandelt en het moment van binnenkomst van de aanvraag. Een zakenmagazijn kan bestaan uit gestructureerde (records) en ongestructureerde (documenten) gegevens. In het laatste geval spreekt men ook wel van zakendossier.’²⁰⁸

Conclusie In het handboek architectuur wordt uitgegaan van het beheer van alle informatie. Er wordt daarbij geen expliciet onderscheid gemaakt tussen gestructureerde en ongestructureerde informatie. Op dit punt is er dus sprake van een verschil tussen ECM en het handboek architectuur.

ECM heeft als uitgangspunt ‘enkelvoudige opslag, meervoudig gebruik van informatie’

Zoals duidelijk is geworden, geldt ten aanzien van de basisregistraties en de kernadministraties de grondslag ‘éénmalige opslag, meervoudig gebruik’²⁰⁹ Het vormt een van de belangrijkste grondslagen van het handboek en er wordt dan ook maar liefst veertien keer aan gerefereerd. Doel van het principe is dat ‘voor zowel de burger als voor de uitvoerende organisaties geldt dat direct inzichtelijk moet zijn wat de voortgang van de behandeling is en welke andere vragen en/of klachten aan deze persoon gekoppeld zijn. Per vraag of klacht noemen we deze informatie een zaakdossier. Meerdere zaakdossiers van een burger tezamen vormen een klantendossier. Voor de informatievoorziening betekent dit dat er gebruik gemaakt moet worden van één voor het gehele concern toegankelijk zakenmagazijn waarin de informatie van de verschillende zaakdossiers te vinden is. In zo’n zaakdossier moet de vraag of klacht eenduidig vastgelegd worden, de basisinformatie direct gekoppeld zijn, moet de status en voortgang van de zaak zichtbaar gemaakt kunnen worden en een digitaal dossier gekoppeld.’²¹⁰ Om de enkelvoudige opslag van informatie te garanderen is het in de grondslagen van de

²⁰⁷ Ibidem, pag. 7-2

²⁰⁸ Ibidem, pag. bijlage - 60

²⁰⁹ Ibidem, pag. management samenvatting - 13, 1-2, 3-2, 6-1, 6-2, 6-2, 6-4, 6-6, 6-6, 6-11, 6-11, 6-21, 7-2, bijlage 22

²¹⁰ Ibidem, pag. 6-2

architectuur opgenomen

‘De basisregistraties en kernadministraties vormen een verplichte bron van de Amsterdamse gegevenshuishouding

Reden:

Bij het eenmalig vastleggen is het verplicht om de bestaande gemeentelijke basisregistraties en kernadministraties als bron te gebruiken.²¹¹

‘Gegevens worden éénmalig opgeslagen en meervoudig gebruikt.

Reden:

Gegevens (records, tekst, foto’s, enz.) zijn van gemeenschappelijk nut en worden dan ook (voor zover mogelijk en toegestaan) gedeeld door uiteenlopende interne en externe functies. Voor de kwaliteit en inzichtelijkheid is het van groot belang dat gegevens maar op één plek kunnen worden gewijzigd (vastgesteld), zij kunnen wel op meerdere plekken worden gebruikt.²¹²

Conclusie Zowel ECM als het handboek architectuur gaan uit van het principe ‘enkelvoudige opslag, meervoudig gebruik van informatie’.

ECM gaat uit van het koppelen van de verschillende applicaties

De huidige situatie wordt in het handboek als volgt beschreven. ‘Amsterdam kent 30 diensten, 7 bedrijven en 14 stadsdelen. Elk van deze organisatieonderdelen kent een relatief zelfstandige management-verantwoordelijkheid. De directeuren van concernonderdelen en stadsdeelsecretarissen zijn verantwoordelijk voor de eigen lokale netwerken en informatievoorziening. De stadsdelen, diensten en bedrijven hebben veel vrijheid bij het inrichten hiervan. Op het niveau van applicaties zien we dat zogenaamde eilandautomatisering is ontstaan. Historisch gezien is dit verklaarbaar en er zijn ook voordelen aan. Er kan namelijk maatwerk worden geleverd aan integrale managers en snel gereageerd worden op ad hoc vragen. De eilandautomatisering heeft echter ook een keerzijde. Schaalvoordelen worden gemist en het vormt een zware belemmering voor het realiseren van de doelstellingen van de Andere Overheid. Vanuit het oogpunt van architectuur is daarom een koerswijziging nodig richting gemeenschappelijkheid en standaardisering. Deze koerswijziging biedt ook kansen voor de kosten en de kwaliteit van het beheer. De eilandautomatisering maakt dat het beheer op dit moment ook versnipperd is ingericht met grote kwaliteitsverschillen. Via professioneel gemeenschappelijk beheer is hier dus nog winst te boeken.²¹³

De drie belangrijkste problemen in het huidige Amsterdamse applicatielandschap zijn:

- ‘Verticale inrichting. Applicaties zijn op dit moment per organisatie (dienst, bedrijf, stadsdeel) ingericht en daarbinnen per probleemveld of behoefte. Deze vorm van

²¹¹ Ibidem, pag. bijlage - 22

²¹² Ibidem, pag. bijlage - 22

²¹³ Ibidem, pag. 7-27

inrichting wordt ook wel ‘verticale silo’s’ of ‘eilandautomatisering’ genoemd.

- Monolithische inrichting. Informatiesystemen zijn ingericht als monolithische blokken oftewel ze vormen één homogeen geheel. Daarmee zijn ze gesloten voor de omgeving en bijna onmogelijk ‘op te hakken’ in verschillende deeloplossingen.
- Versnipperde inrichting. De praktijk is dat er per probleemveld of behoefte naar een nieuwe oplossing wordt gezocht. Het aantal oplossingen is hiermee op dit moment zeer gevarieerd en we missen schaalvoordelen (door overlap in type oplossingen).²¹⁴

Dat de huidige situatie verre van ideaal is en er dus een noodzaak tot verandering is, wordt krachtig uitgedragen. ‘We kunnen onze business-doelstellingen alleen halen als we met één loket naar buiten treden, processen delen, gegevens delen en middelen delen. En om alle organisaties op elkaar aan te sluiten zijn er voorzieningen “in het midden” nodig om de processen te regisseren en de gezamenlijke gegevens te ontsluiten.’²¹⁵ Dit roept direct de vraag op hoe deze ‘voorzieningen in het midden’ er dan uit zien. Maar daar wordt een duidelijk antwoord op gegeven. ‘Over de oplossingsrichting bestaat consensus. Amsterdam moet toewerken naar een zogenaamde service gerichte architectuur [...] waarbinnen processen zijn opgedeeld in componenten en de techniek (applicaties en infrastructuur) meer en meer op een standaard manier wordt ingevuld. De technische bouwblokken behoren daarbij niet langer exclusief tot een bepaalde applicatie, maar kunnen van verschillende kanten worden aangeropen. De belangrijkste gevolgen van deze strategie voor de inrichting van de applicatiearchitectuur zijn:

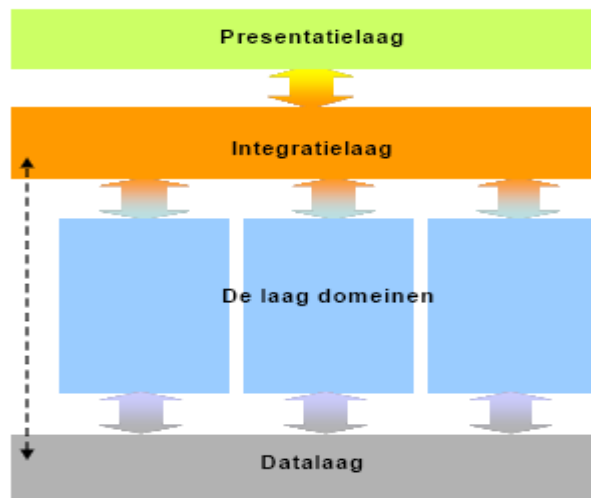
1. Integratie van presentatiefuncties;
2. Openheid bij de inrichting van applicaties en gemeenschappelijke integratievoorzieningen;
3. Gemeenschappelijke ontwikkeling en beheer van applicaties;
4. Standaardisering van functies;
5. Modulaire opbouw van applicaties.²¹⁶

²¹⁴ Ibidem, pag. 7-2

²¹⁵ Ibidem, pag. 4-2

²¹⁶ Ibidem, pag. 7-2, 7-3

Typologie applicatielandschap



Figuur 48: typologie applicatielandschap (globaal)

‘Bij de inrichting van het Amsterdamse applicatielandschap kijken we naar de bedrijfsmatige (functionele) aspecten van de applicaties. We onderscheiden op functioneel niveau vier verschillende lagen

1. de presentatielaag
2. de integratielaag
3. de laag domeinen
4. de data laag

In de *presentatielaag* zijn alle functies ondergebracht die te maken hebben met de interactie tussen gemeente en de ‘buitenwereld’, waaronder samenwerkingspartners, burgers en

ondernemers. Ook de interne medewerkers wordt als (doel)groep daarin meegenomen. Dit wordt ook wel met de front office geïdentificeerd. De *integratielaag* bevat functies die het mogelijk maken om gegevens te kunnen uitwisselen tussen de lagen presentatie, domeinen en data. Dit wordt ook wel met de mid office geïdentificeerd. Een applicatiedomein is makkelijk te vereenzelvigen met (op organisatieniveau) een dienst of stadsdeel of (op infrastructuurniveau) een netwerk van een deelnemer in de E-net infrastructuur. Binnen het Amsterdamse applicatielandschap wordt met *de laag domeinen* echter bedoeld op functies binnen de inhoudelijke domeinen (beleidsterreinen) waar de gemeente zijn primaire processen heeft geconcentreerd. In de referentiearchitecturen van onder andere NORA en EGEM worden deze domeinen meestal aangeduid als back office. In de *data laag* worden de basisregistraties en kernadministraties beschikbaar gesteld aan de overige applicatielagen. Dit kan rechtstreeks naar de laag domeinen of via de integratielaag.²¹⁷

Richtlijn integratie:

Met betrekking tot de integratie staat er een aantal duidelijke uitspraken in het handboek architectuur:

- ‘Integratie vindt plaats binnen één platform. Integratie van applicaties in verschillende domeinen vindt alleen plaats via het platform en niet rechtstreeks.’²¹⁸
- ‘Van individuele applicaties mogen de gegevens- en de applicatielaag binnen de applicatie niet door gebruikers rechtstreeks worden benaderd. Een gebruiker krijgt alleen toegang tot de applicatie via de presentatielaag. De presentatielaag roept vervolgens de applicatielaag aan en die roept weer de gegevenslaag aan. Alleen beheerders mogen met extra authenticatievoorzieningen rechtstreeks toegang verkrijgen tot applicatielaag en gegevenslaag.’²¹⁹
- ‘Als regel worden de gegevenslagen van meerdere systemen nooit direct aan elkaar gekoppeld. De applicatielaag is verbonden met een integratievoorziening die de datacommunicatie tussen de systemen orkestreert.’²²⁰

²¹⁷ Ibidem, pag. 7-5, 7-6

²¹⁸ Ibidem, pag. 7-22

²¹⁹ Ibidem, pag. 7-29

- Integratiefuncties ondersteunen op verschillende niveaus zoals processen (bijv. tijdsbewaking, monitoring), informatie (verschillen in semantiek en syntax), applicaties

(koppelen van zeer uiteenlopende systemen op een beveiligde en stabiele manier) en infrastructuur (locatie onafhankelijk werken). Integratie is dus geen zuiver technische oplossing binnen de applicatiearchitectuur, maar strekt zich uit over alle architectuurlagen.²²¹

Conclusie Zowel ECM als het handboek architectuur gaat ervan uit dat applicaties open dienen te worden ingericht voor hun omgeving. Bij beide is het uitgangspunt dat door middel van een gemeenschappelijke voorziening de applicaties worden geïntegreerd.

ECM gaat uit van generieke applicaties

Op het niveau van applicaties wordt er in het handboek architectuur een scheiding gemaakt tussen processpecifieke en procesgenerieke functies:

Processpecifieke functies

‘Dit zijn functies die ontstaan ter ondersteuning van specifieke bedrijfsprocessen die uniek zijn binnen de gemeente. Het gaat vrijwel altijd om functies waarbij de inhoud van het proces in de functie is verweven, zoals het registreren van erfpachtbelasting of de vaststelling van het recht op bijstandsuitkering.’²²²

Procesgenerieke functies

‘Dit zijn functies die in meerdere, zo niet alle bedrijfsprocessen voorkomen. Een goed voorbeeld is het beheer van werkstromen. In de meeste processen speelt een combinatie van werkverdeling en tijdsbewaking een rol. Deze kunnen daarom generiek ondersteund worden met zogenaamde workflow managementfuncties. Hetzelfde geldt bijvoorbeeld voor functies voor documentbeheer.’²²³

In het applicatielandschap manifesteert de scheiding tussen processpecifieke- en procesgenerieke functies zich in de domeinlaag. Dit is de laag van de inhoudelijke domeinen (beleidsterreinen) waar de primaire processen zijn geconcentreerd. ‘In Model 7.9 is met een arcering aangegeven welke functies binnen welke lagen van het Amsterdamse applicatielandschap gemeenschappelijk gebruikt kunnen worden. Het betreft dus de functies die de mogelijkheid in zich hebben om gemeenschappelijk ontwikkeld en beheerd te worden. Wat opvalt is dat alleen de processpecifieke functies binnen de laag domeinen buiten het gemeenschappelijk domein blijven.’²²⁴

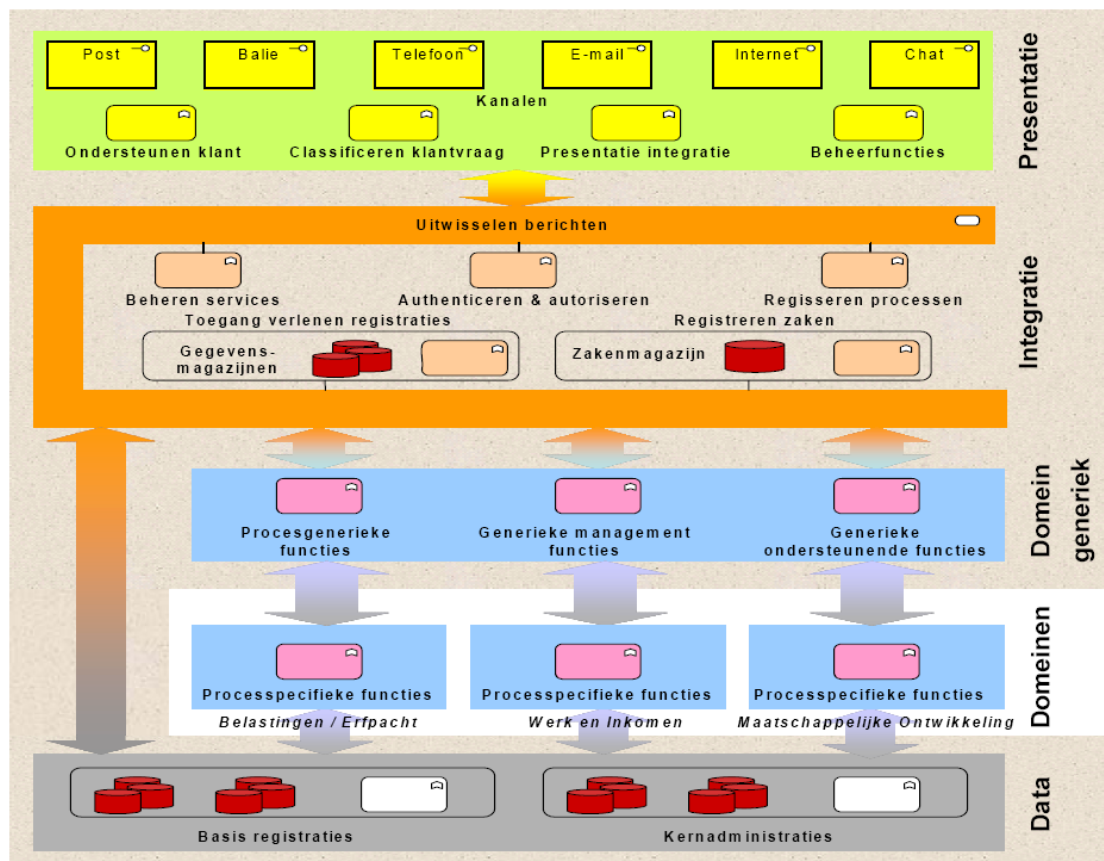
²²⁰ Ibidem, pag. 7-29

²²¹ Ibidem, pag. 7-9

²²² Ibidem, pag. 7-14

²²³ Ibidem, pag. 7-14

²²⁴ Ibidem, pag. 7-19



Figuur 49: model 7.9 het toekomstige applicatielandschap: wat kan gemeenschappelijk?

Conclusie In het handboek architectuur wordt uitgegaan van zowel specifieke als generieke applicatie. Op dit punt is er dus sprake van een verschil tussen ECM en het handboek architectuur

ECM gaat uit van standaardiseren van metadata

Over nut en noodzaak van standaardisering van metadata wordt o.a. het volgende gesteld. 'Om informatie te kunnen uitwisselen zijn afspraken nodig over de vorm en taal waarin dit wordt gedaan. Standaardisering dus. Een goed voorbeeld is standaardisering op het terrein van digitaal informatiebeheer. Als stadsdelen en diensten op een eigen manier informatie registreren en archiveren, wordt het bijna onmogelijk om informatie terug te vinden en te ontsluiten voor geïnteresseerden. Een Andere Overheid vereist standaardisering op allerlei terreinen. De keuzes in de informaticarchitectuur zijn van invloed op hoger- en lageregelegen architectuurlagen. Met name het principe van 'éénmalige opslag, meervoudig gebruik' is bepalend voor de vormgeving van alle hoofdprocessen. In deze processen wordt immers intensief gebruik gemaakt van basisregistraties, kernadministraties en het zaakdossier. Ook zal in deze processen het digitaal informatiebeheer verankerd moeten worden. Ook op de lager gelegen architectuurlagen heeft de informaticarchitectuur impact. De applicatie- en infrastructuurarchitectuur zullen onbelemmerde gegevensuitwisseling mogelijk moeten maken binnen de randvoorwaarden die de Gemeentelijke InformatieBeveiligingsNorm (GIBN) stelt.²²⁵

Conclusie Zowel ECM als het handboek architectuur gaan uit van het principe dat standaardisering van metadata cruciaal is om informatie te kunnen uitwisselen.

ECM heeft als doel het openbreken van informatiesilo's

‘Applicaties zijn nu in hoofdzaak “organisatiegebonden”. Elke gemeentelijke organisatie in Amsterdam is vrijwel autonoom in de keuze van applicaties, hoe die zijn ingericht en wie ze mogen gebruiken. En over het algemeen bevinden ze zich ook in de infrastructuur van de betreffende organisatie. Feitelijk vormt de applicatiehuishouding van een dienst nu een gesloten domein, met eigen regels en standaarden. Binnen zo’n domein zijn applicaties redelijk op elkaar afgestemd en afstemming met applicaties van derden gebeurt voornamelijk op ad hoc basis. Het voordeel van deze autonomie is dat elke organisatie applicaties kan gebruiken die het best aansluiten op de interne bedrijfsvoering en die “passen” in de “eigen” infrastructuur. Maar de autonomie leidt ook tot versnippering. Verschillende diensten gebruiken voor dezelfde functionaliteit verschillende applicaties. Soms zijn er zelfs binnen één organisatie meerdere oplossingen voor dezelfde functie. Ook in geval van het gebruik van dezelfde applicaties verschilt de inrichting nogal eens per dienst. Deze versplintering brengt extra kosten met zich mee en bemoeilijkt de uitwisseling van gegevens. Het realiseren van een Andere Overheid vraagt om samenwerking tussen concernonderdelen. Meer standaardisering in de laag domeinen is daarvoor een noodzakelijke voorwaarde.’²²⁶

‘Tot voor kort had elke dienst en elk stadsdeel zijn eigen loket aan de voorkant voor het eigen werkpakket aan de achterkant. Maar nu willen we graag dat de gezamenlijke front office van 30 diensten plus 14 stadsdelen in principe alle ruim 500 gemeentelijke producten levert vanuit misschien wel even zovele back offices (M keer N). Dan moet er dus een voorziening komen om die aansluiting optimaal te laten verlopen: het uitzetten van de vraag, het bewaken van de voortgang en het terugleveren van het resultaat. Dat kan organisatorisch van aard zijn of technisch, maar zal in ieder geval het aantal relaties moeten reduceren (M plus N). Die voorziening is mid office gaan heten omdat het de koppeling en de afstemming moet verzorgen tussen de front office en de back office.’²²⁷

‘De huidige “eilandautomatisering” in Amsterdam maakt dat de elektronische dienstverlening aan Amsterdamse burgers en bedrijven versnipperd is. Dit ondanks het feit dat er de afgelopen jaren al veel is verbeterd en het gebruik van gemeenschappelijke voorzieningen toeneemt. Toch heeft elk organisatieonderdeel vaak nog een eigen loket als schil rond de eigen applicaties aangelegd. Dit is niet goed voor de kwaliteit van de dienstverlening. Het leidt bijvoorbeeld tot:

- verschillen in toegankelijkheid en bereikbaarheid van onderdelen van het concern Amsterdam;
- overlap in de externe communicatie met Amsterdammers;
- aanbod- in plaats van vraaggestuurde dienstverlening: de Amsterdammer kan nog

²²⁵ Ibidem, pag. 6-2, 6-3

²²⁶ Ibidem, pag. 7-14, 7-15

²²⁷ Ibidem, pag. 4-8

onvoldoende zelf “aan het stuur zitten” en niet overal naartoe “rijden” omdat er nog geen verbindingswegen zijn.

Om dit te doorbreken is een éénduidig gezicht en een éénduidig geluid aan burgers, bedrijven en instellingen nodig. De buitenwereld moet met de gemeente kunnen communiceren, onafhankelijk van de interne procesinrichting van het concern en dat betekent:

- Eén loket gedachte, met meerdere kanalen naar de klant, maar met één gezicht vanuit het perspectief van dienstverlening;
- Eenmalige informatieaanlevering, met de mogelijkheid van regie over eigen persoonsgegevens;
- Eenmalige registratie: basisregistraties en kernadministraties
- Eenmalige aanmelding en authenticatie (het zogenaamde single sign on) bij communicatie met de overheid.²²⁸

‘De kwaliteit en beschikbaarheid van informatie is cruciaal voor verbetering van dienstverlening en stroomlijning van processen. Het “no wrong door” principe vereist beschikbaarheid van dezelfde informatie, ongeacht de locatie. Kanaalafhankelijkheid betekent dat dezelfde informatie (en antwoorden) ongeacht het communicatiekanaal beschikbaar moeten zijn. “Eenmalige opslag, meervoudig gebruik” vereist heldere afspraken over gegevens (lokaal en landelijk) en heldere afspraken over relaties tussen gegevens en beheer van gegevens.’²²⁹

Conclusie Zowel ECM als het handboek architectuur heeft tot doel bestaande informatiesilo’s af te breken.

ECM heeft als doel gedurende de gehele levenscyclus grip te hebben op ongestructureerde informatie

Information Lifecycle Management en de mogelijkheden die dit biedt voor het beheer van de informatiehuishouding worden niet als zodanig in het handboek architectuur besproken. Juist omdat het onderwerp in het geheel niet wordt besproken, is het moeilijk er een conclusie aan te verbinden. Want waarom maakt ILM geen onderdeel uit van het handboek? Wordt het als onbelangrijk ervaren? Is men het vergeten? Is het er bewust uitgelaten omdat men bijvoorbeeld van mening is dat het een onderdeel van de informatiehuishouding is waar de diensten, bedrijven en stadsdelen van Amsterdam zelf voor moeten zorgen en zelf beleid voor moeten maken? Een antwoord op de vraag *waarom* ILM geen onderdeel uitmaakt van het handboek architectuur kan ik op basis van de literatuurstudie niet geven en mijn conclusie zal zich dan ook beperken tot de constatering *dat* het onderwerp niet wordt besproken.

Conclusie Het doel gedurende de gehele levenscyclus grip te hebben op ongestructureerde informatie wordt niet expliciet genoemd in het handboek architectuur.

ECM gaat niet slechts betrekking tot het ‘managen’ van ongestructureerde informatie

²²⁸ Ibidem, pag. 7-7

²²⁹ Ibidem, pag. 6-1

maar heeft tot doel het bedrijfsproces optimaal te ondersteunen

‘De architectuur richt zich op meer dan alleen informatievoorziening en ICT. Het is een zogenaamde “enterprise architectuur”. De architectuur gaat niet alleen over informatievoorziening en techniek, maar richt zich ook op de koppeling met het zogenaamde businessdomein (bedrijfsprocessen en de organisatie). Dit is noodzakelijk omdat de realisatie van de overheid vraagt om samenhang en samenwerking op alle niveaus: de business kan niet zonder ICT en vice versa. De uitwerking van het businessdomein is op een hoger abstractieniveau dan de uitwerking van de informatievoorziening’.²³⁰

Conclusie Zowel ECM als het handboek architectuur hebben niet slechts betrekking op het ‘managen’ van ongestructureerde informatie maar hebben tot doel het bedrijfsproces optimaal te ondersteunen.

²³⁰ Ibidem, pag. 2-5

5.6. Reflectie

5.6.1. Uitkomst vergelijking

Laten we de onderzoeksvraag er nog eens bij halen, daar is het immers allemaal om begonnen. De vraag was:

‘Zijn Enterprise Content Management (ECM) en het Handboek Architectuur twee verschillende termen voor hetzelfde principe of zijn het twee verschillende benaderingen voor het beheren van dezelfde digitale informatie?’

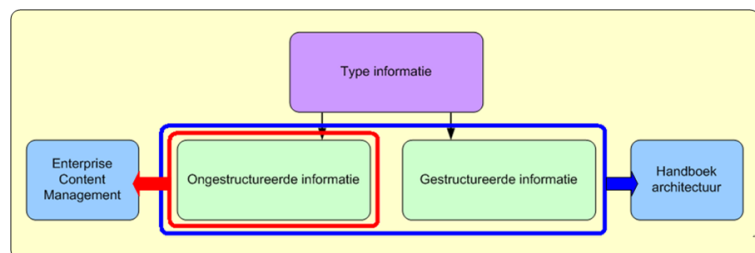
Als we de conclusies van de bovenstaande vergelijking op een rijtje zetten dan kan geconstateerd worden dat de twee op veel punten overeenkomsten vertonen:

- beide kunnen als een strategie worden beschouwd;
- beide gaan uit van het principe van ‘enkelvoudige opslag, meervoudig gebruik’;
- bij beide wordt standaardisering van metadata als cruciaal beschouwd om informatie-uitwisseling mogelijk te maken;
- bij beide staat het koppelen van applicaties centraal;
- bij beide is het afbreken van informatiesilo’s het doel van het koppelen van applicaties;
- beide zien informatiebeheer niet als een doel op zich, maar als een manier om het bedrijfsproces te optimaliseren.

Tot zover de overeenkomsten; uit de vergelijking blijkt ook dat er op drie punten verschillen zijn tussen de twee benaderingen:

1) Gestructureerde en ongestructureerde informatie.

ECM heeft betrekking op ongestructureerde informatie, het handboek architectuur daarentegen richt zich op de gehele informatiehuishouding (gestructureerd en ongestructureerd) van het concern Amsterdam. Het verschilt op dit punt dan ook fundamenteel van de ECM-strategie.

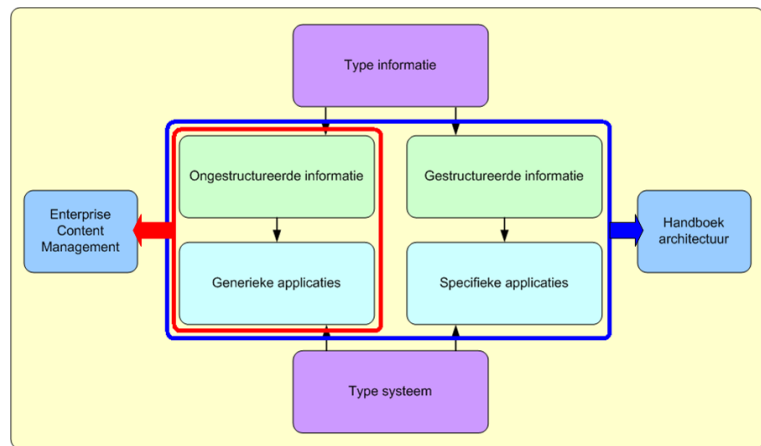


Figuur 50: verhouding ECM - handboek architectuur: informatie

2) Generieke en specifieke applicaties

Dit onderscheid in informatietype vertaalt zich (zoals al in hoofdstuk 2 is beschreven) ook naar applicatietypes die gebruikt worden voor het informatiebeheer. Enigszins gechargeerd kan de verdeling als volgt worden gemaakt: gestructureerde informatie wordt beheerd in specifieke

applicaties en ongestructureerde informatie in generieke applicaties. Daar waar ECM zich richt op ongestructureerde informatie door middel van generieke applicaties, richt het handboek architectuur zich op de gehele informatiehuishouding en dus op zowel generieke als specifieke applicaties. Dus ook op dit punt verschillen ECM en het handboek architectuur fundamenteel van elkaar.

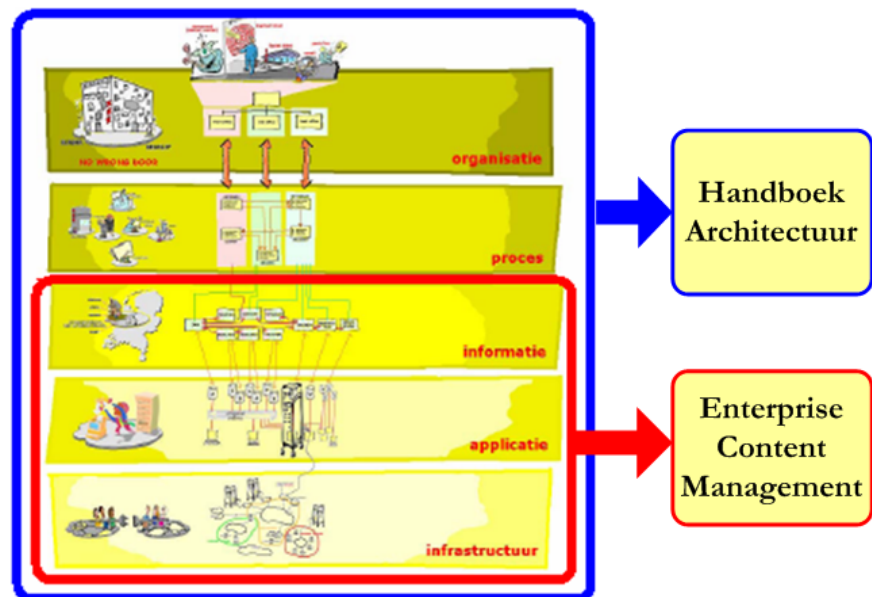


Figuur 51: verhouding ECM - handboek architectuur: applicatie

3) Scope van strategie

Het derde punt waarop ECM en het handboek architectuur fundamenteel van elkaar afwijken is de scope. Ik heb al aangegeven dat het handboek architectuur uitgaat van vijf lagen (organisatie, proces, informatie, applicatie, infrastructuur). ECM heeft betrekking op opname, gebruik, beheer, ontsluiting, opslag en distributie van ongestructureerde informatie. Hiermee beperkt ECM zich tot de onderste drie lagen uit het model: *informatie*, *applicatie* en *infrastructuur*.

Dit wil natuurlijk niet zeggen dat ECM geen invloed heeft op de twee bovenste lagen (*organisatie* en *proces*). Wat ik hiermee bedoel zal ik uitleggen aan de hand van een voorbeeld. De implementatie van workflow gaat vaak gepaard met het optimaliseren en hermodelleren van de processen (laag 2). De invoering van ECM heeft hiermee effect op de proceslaag. Maar, en dat is het verschil, het optimaliseren en hermodelleren van de processen is



Figuur 52: verhouding ECM - handboek architectuur: scope

een (mogelijk) gevolg van de implementatie van ECM. Het maakt echter geen onderdeel uit van de ECM-strategie *an sich*. In dit opzicht kan dan ook gesteld worden dat de scope van het handboek architectuur breder is dan die van de ECM-strategie.

5.6.2. Kantekening bij vergelijking

Het doel van deze scriptie is antwoord te krijgen op de vraag of ECM en het handboek architectuur verschillende termen zijn voor hetzelfde principe. De scriptie heeft niet tot doel het handboek architectuur inhoudelijk te toetsen op bijvoorbeeld kwaliteit of haalbaarheid. Toch wil ik een inhoudelijke kanttekening plaatsen bij het handboek architectuur. Dit is nodig omdat er anders wellicht een vertekend beeld ontstaat. Zetten we de bovenstaande verschillen tussen ECM en het handboek architectuur in een tabel dan ziet de conclusie er als volgt uit:

	ECM	Handboek Architectuur
Gestructureerde informatie	✗	✓
Ongestructureerde informatie	✓	✓
Generieke applicaties	✓	✓
Specifieke applicaties	✗	✓
Scope		
• Organisatie	✗	✓
• Proces	✗	✓
• Informatie	✓	✓
• Applicatie	✓	✓
• Infrastructuur	✓	✓

Deze tabel schetst naar mijn idee echter een vertekend beeld. Ik zal uitleggen waarom. De tabel geeft aan dat de scope van het handboek de gehele informatiehuishouding (het totaal van gestructureerde en ongestructureerde informatie) omvat. De tabel wekt daarmee de suggestie dat het handboek een uitgewerkte strategie voor de gehele informatiehuishouding behelst. De kanttekening die hierbij geplaatst moet worden is dat de focus van het handboek architectuur zich met name richt op de gestructureerde informatie (basisregistraties).²³¹ Dit wordt overigens ook in het handboek onderkend.

'Deze eerste versie van de architectuur is ontstaan vanuit de businessdoelen in het programma Basisregistraties en ICT infrastructuur. In onze visie moet de architectuur zich verder ontwikkelen door nadrukkelijk mee te liften op belangrijke verandertrajecten binnen de gemeentelijke organisatie.'²³²

Kijken we vanuit het ECM-perspectief, dan zien we dat het beheer van ongestructureerde informatie in het handboek architectuur onderbelicht blijft. Dit wordt met name duidelijk als we naar het zakenmagazijn kijken. Het zakenmagazijn is een essentieel onderdeel van de strategie en gaat een van de kernadministraties van Amsterdam worden.²³³

Zoals aangegeven moeten in het zakenmagazijn alle zaakdossiers van de Gemeente Amsterdam komen. Een zaakdossier bevat alle gestructureerde en ongestructureerde informatie die tot een unieke zaak behoren. Daarbij wordt aangegeven dat 'met name voor het hoofdproces dienstverlening [...] een standaard zaakdossier en zakenmagazijn van cruciaal belang [is]'.²³⁴ Vooral het woord *standaard* is in deze zin van belang, want standaardisering is voor de

²³¹ Ibidem, pag. 2-4

²³² Ibidem, pag. 2-5

²³³ Ibidem, pag. 7-11, 7-13

²³⁴ Ibidem, pag. 6-2

uitwisselbaarheid van de informatie van cruciaal belang. De opvatting wordt in het handboek architectuur onderstreept, want ook daar wordt benadrukt dat om informatie te kunnen uitwisselen het noodzakelijk is afspraken te maken over de vorm en taal waarin dit wordt gedaan. Worden er geen afspraken gemaakt en registreren en archiveren de stadsdelen en diensten de informatie ieder op hun eigen manier, dan is het gevolg, zo wordt ook in het handboek gesteld, dat het bijna onmogelijk wordt informatie te ontsluiten en terug te vinden.²³⁵

Ondanks dat standaardisering als cruciaal wordt beschouwd, zijn er in het handboek met betrekking tot ongestructureerde informatie opvallend weinig standaarden opgenomen.

‘Een goed voorbeeld van de noodzaak tot standaardisering van digitale informatie is het koppelen van een digitaal dossier aan een zaakdossier. In een volgende versie van dit handboek moet daarom een objectenmodel opgenomen worden met daarin de relatie tussen de ongestructureerde data (gedigitaliseerde documenten) en het zaakdossier. Ook is het wenselijk dan een aantal standaarden vast te stellen voor de metadata. [...] Voor metadata moeten afspraken gemaakt worden over de volgende aspecten: elementen, waarden, classificaties en syntax.’²³⁶

Zoals ik ook in hoofdstuk 4 heb betoogd, vormt de standaardisering van metadata een absolute voorwaarde voor de uitwisseling (op inhoudelijk gebied) van ongestructureerde informatie. De koppeling van de generieke systemen en het daarmee afbreken van de informatiesilo's biedt pas echt meerwaarde als de daarin opgeslagen informatie inhoudelijk aan elkaar gerelateerd kan worden. Het ontbreken van afspraken met betrekking tot de standaardisering van het zaakdossier en de metadata maakt duidelijk dat het beheer van ongestructureerde informatie onderbelicht blijft en nog onvoldoende is uitgewerkt. Laat ik het anders stellen: zonder standaarden geen zaakdossier, geen zakenmagazijn en dus geen concernbreed beheer van ongestructureerde informatie. Eerlijk is eerlijk, het handboek architectuur is onvolledig, maar dat wordt ook in het stuk zelf duidelijk aangegeven.

‘Deze eerste versie van de architectuur is niet volledig. Het bevat alleen die elementen die cruciaal zijn voor de realisatie van de Andere Overheid (en meer specifiek het programma BRI) en zaken die “op de plank lagen” en zo makkelijk in de architectuur konden worden “gehangen”.’²³⁷

Samenvattend kan gesteld worden dat ondanks dat ongestructureerde informatie en generieke applicaties (het domein van ECM) onderdeel zijn van het handboek architectuur, juist deze onderdelen onvoldoende zijn uitgewerkt. Hierdoor omvat het handboek architectuur wel het domein van ECM, maar wordt het slechts op zeer abstract niveau uitgewerkt en is er eigenlijk geen sprake van een volwaardige strategie.

In het ontbreken van standaarden met betrekking tot het beheer van ongestructureerde informatie schuilt het gevaar dat juist op dat vlak de verschillende onderdelen van de Gemeente Amsterdam hun eigen beleid gaan ontwikkelen, wat in de toekomst tot problemen kan leiden. Dat ongestructureerde informatie zo onderbelicht is gebleven, is des te opvallender aangezien het overgrote deel (80%) van de informatie ongestructureerd is (zie hoofdstuk 2) en in generieke applicaties wordt beheerd. Het is juist deze ongestructureerde informatie die qua omvang exponentieel groeit en alleen al om die reden extra aandacht verdient.

²³⁵ Ibidem, pag. 6-2

²³⁶ Ibidem, pag. 6-15

²³⁷ Ibidem, pag. 2-6

De kritische kanttekening dat veel informatie nog uitgewerkt moet worden doet echter niets af aan het feit dat het beheer van ongestructureerde informatie, standaardisering van metadata en generieke applicaties wel degelijk onderdeel uitmaken van het handboek architectuur. Ik heb alleen willen aangeven dat die onderdelen in deze versie van het handboek onderbelicht zijn gebleven.

5.7. Conclusie

Het moge duidelijk zijn dat de vraag of ECM en het handboek architectuur twee verschillende termen zijn voor hetzelfde principe negatief moet worden beantwoord. Het zijn evenmin twee verschillende benaderingen voor het beheer van dezelfde digitale informatie.

De conclusie is dat het handboek architectuur breder georiënteerd is dan ECM.

Daar waar ECM betrekking heeft op ongestructureerde informatie, gaat het handboek over de hele informatiehuishouding. Daar waar ECM bestaat uit generieke applicaties, is het handboek van toepassing op alle applicaties. En daar waar ECM zich beperkt tot de informatie en de techniek, heeft het handboek daarnaast ook betrekking op de organisatie en het proces.

Gesteld zou kunnen worden dat de ECM-strategie deel uitmaakt van het breder georiënteerde handboek architectuur.

5.8. Nawoord: positionering ECM

Het voorgaande stelt ons in staat ECM te positioneren binnen de informatiehuishouding. Zoals duidelijk is geworden, hebben organisaties in de regel een veelheid van applicaties voor allerlei uiteenlopende doeleinden in gebruik. Grofweg kunnen echter de volgende type systemen worden onderscheiden:

- Enterprise Resource Planning (ERP) t.b.v. inzicht in middelen;
- Customer Relationship Management (CRM)²³⁸ t.b.v. klantcontact en inzicht in contacten;
- Business Intelligence (BI)²³⁹ t.b.v. inzicht in bedrijfsprocessen;
- Enterprise Content Management (ECM) t.b.v. inzicht in content/informatie;
- Applicaties voor organisatiespecifieke doeleinden (hier aangeduid als *Andere applicaties*).²⁴⁰

Als deze indeling wordt aangehouden, dan kan de positie van ECM binnen de organisatie (naast de andere applicaties) als volgt worden weergegeven.

²³⁸ Customer Relationship Management oftewel CRM (ook wel bekend als relatiemarketing of verkoopbeheersysteem) is een werkwijze alsmede een techniek waarbij het optimaliseren van alle contacten met de klant centraal staat. Een CRM-systeem is een software-pakket voor het beheren van klantgegevens en interacties met deze klanten.

²³⁹ Business Intelligence is gericht op het verzamelen en analyseren van informatie over klanten, beslissingsprocessen, concurrentie, markttoestand en algemene economische, technologische en culturele trends, teneinde beslissingsondersteunende informatie (intelligence) te verkrijgen.

²⁴⁰ *Kernmodel II, ervaring met en visie op informatiemanagement bij de rijkssoeverheid* – Interlab, Den Haag (2006) pag. 31
http://www.rijks-interlab.nl/Images/KernmodelII%20definitief%20met%20copyright_tcm134-94192.pdf



Figuur 53: positionering ECM in informatie huishouding

ECM richt zich op het beheer van ongestructureerde informatie en het openbreken van de informatiesilo's waarin de informatie gevangen zit. Eenzelfde ontwikkeling vindt plaats op het gebied van gestructureerde informatie, daarbij gaat het om strategieën en technologieën als ERP, CRM en business intelligence. Echter, doordat gestructureerde en ongestructureerde informatie als losse entiteiten worden beschouwd met ieder hun eigen strategie en technologie, bestaan er nog altijd twee informatiesilo's. Dit betekent dat er ook na een succesvolle implementatie van de bovengenoemde strategieën en technologieën nog steeds beperkingen zijn bij het beschikbaar stellen van bedrijfsinformatie.

In dat kader introduceer ik een nieuw begrip: Enterprise Information Management. Het informatie- en technologie-onderzoeksbedrijf Gartner ziet een evolutionaire ontwikkeling, waarbij een steeds verdere integratie plaatsvindt van gestructureerde- en ongestructureerde informatie. Gartner noemt deze ontwikkeling Enterprise Information Management.²⁴¹

'Enterprise information management (EIM) is a technology strategy for organizing, designing, cataloguing and safeguarding structured and unstructured content to maximize its value, usefulness, accessibility and security. Gartner defines EIM as an organizational commitment to define, secure and improve the accuracy and integrity of information assets and to solve semantic inconsistencies across all boundaries, thus supporting the technical, operational and business objectives within the company's enterprise architecture strategy.'²⁴²

Zo op het eerste gezicht lijken EIM en het handboek architectuur twee verschillende termen voor hetzelfde principe. Maar of dit ook echt zo is, of dat het misschien toch twee verschillende benaderingen zijn voor het beheer van dezelfde digitale informatie, is een vraag die door iemand anders beantwoord mag worden.

²⁴¹ Shegda, Karen M.; e.a. - *Magic Quadrant for Enterprise Content Management* - Gartner (2007); pag 3

²⁴² Newman, David - *Business Drivers and Issues in Enterprise Information Management* (Gartner RAS Core Research Note G00129712) – Gartner (2005)
<http://www.avanade.com/uploaded/pdf/avanadearticle4124441.pdf>

Lijst van geraadpleegde literatuur

Analysis, Selection, and Implementation Guidelines Associated with Electronic Document management Systems (EDMS) - An AIIM Recommended Practice Report prepared by the Association for Information and Image Management International (2006); pag. 7,14-15
<http://www.aiim.org/documents/standards/arp1-2006.pdf>.

Arnold, A.G.; Glashouwer, B. - *De moderne informatiehuishouding van de digitale overheid, het archief op het bureau* - Stichting Het Expertise Centrum, Amsterdam (2005) http://www.hec.nl/index.php?page_id=58

Benneker, Frank - *Quickscan naar mogelijkheden voor automatische metadatageneratie* - Digitale Universiteit (2006) http://www.darenet.nl/nl/page/repository.item/show?saharaIdentifier=ou_11:oai:dspace.ou.nl:1820/803

Boiko, Bob - *Content Management Bible (2nd edition)* - Indianapolis (2005)

Bussel, G.J.; Horsman P.J.; Waalwijk H. - *Softwarespecificaties voor Record Management Applicaties voor de Nederlandse Overheid 2004 (ReMANO 2004)* - Archiefschool, Amsterdam (2004)
http://www.archiefschool.nl/docs/ReMANO_2004.pdf

Bussel, G.J. - *'Fixed content', context en SOX* - VBDS (2005) http://www.vbds.nl/index.php?option=com_content&task=view&id=88&Itemid=63

Cate ten, Timo; Groeneweg, Ronald - *Workflow en documentflow, Over mensen, organisaties, processen en documenten* - Digital Display, Deventer (2007)
<http://www.ddisplay.nl/beheer/UserFiles/File/Seminar/Seminar%202006/workflow%20en%20documentflow.pdf>

Cate ten, Timo; Groeneweg, Ronald - *De Zaak X..., zaakgewijs werken in theorie en digitale praktijk* - Digital display (2007)
<http://www.ddisplay.nl/beheer/UserFiles/File/De%20Zaak%20X.pdf>

Chen, Ying; Ong, Shauchi - *Holistic information management solution* - IBM research report (2005)

Claassen, Laurent - *Compliance is een 'modewoord'* - Ernst & Young
http://www.ey.nl/download/publicatie/Compliance_is_een_modewoord.pdf

Colard, Robin; Zutphen, Frank van; Székely, Tom - *Personele en organisatorische gevolgen digitalisering backoffice* - ECORYS-NEI Arbeid & Sociaal Beleid (2002)
http://www.rmconventie.nl/ozo/p_en_o-gevolgen_digi_backoffice.pdf

Damen, J.; Leeuwen, S. van - *Document beheer in de praktijk: Trends, voordelen en voorwaarden bij een document management software implementatie* - Management Kennisbank Vakspecialisten (2006)
<http://rubeo.eu/files/Document%20beheer%20in%20de%20praktijk.pdf>

Damme, Maarten van - *Business Case implementatietraject DMS stadsdeel Amsterdam-Centrum* - stadsdeel Amsterdam Centrum (2005)

Damme, Maarten van - *Functioneel Ontwerp Vervanging GRS Stadsdeel Amsterdam-Centrum Vierde concept* - stadsdeel Amsterdam Centrum (2006)

Davenport, Thomas H.; Cohen, Don - *Solving the information management puzzle: A life cycle approach* - (2005)

http://forbes.bitpipe.com/detail/RES/1147100054_420.html

Doekes, M. - *RMA of DMS: cruciaal nuanceverschil* - Digital Display

http://digidiv.amsterdam.nl/handleidingen_en_instrumenten/documentaire_informatiesystemen/kenmerken_en_cases/rmad.doc

Enterprise Content Management, de route naar brede acceptatie - Microsoft whitepaper (2006)

<http://www.microsoft.com/netherlands/adviseurs/productiviteit/ecmoplossingen.aspx>

Gable, Julie - *Enterprise application integration* - Information Management Journal (March/April 2002)

Gantz, John F.- *The expanding digital universe, a forecast of worldwide information Growth through 2010* - IDC (2007)

http://www.emc.com/about/destination/digital_universe/pdf/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf

Giesbers, Saskia - *Records Management Terminologie, Resultaten van een onderzoek in opdracht van de Records Management Conventie* - RMC bureau (2004); pag. 19-23, 42-44

http://www.rmconventie.nl/uploads/RecordsManagementTerminologiev4juli2004_1.pdf

Griffiths, J. R.; Brophy, P. - *Students searching behaviour and the web* - in Library Trends 53(4) 2005: 539-554

Handboek architectuur, de samenhang in organisatie en informatievoorziening van de Gemeente Amsterdam - adviesgroep architectuur versie 1.0 april 2007

Hartman Erik, M. - *Content management: een groeimarkt* - In: VIP vakblad voor documentmanagement, nummer 3 jaarboek (2004)

http://content.hartman-communicatie.nl/assets/binaries/Documenten/cm_een_groeimarkt_x_hartman.pdf

Heijkoop, Harry - *Basicursus XML (herziene editie)* - Den Haag (2005)

Hofman, Hans - *De digitale archivaris: een nieuwe wereld. De invloed van de informatietechnologie op het archiefvak* - In: Horsman, Peter J.; Ketelaar, Frederick C.J.; Thomassen, Theo H.P.M. (red.) - Naar een nieuw paradigma in de archivaliek. Jaarboek 1999 Stichting Archiefpublicaties ('s-Gravenhage 1999)

Hofman, Hans - *Een uitdijend beelal? Context van archiefbescheiden* - In: Horsman, Peter J.; Ketelaar, Frederick C.J.; Thomassen, Theo H.P.M. (red.), Context, interpretatiekaders in de archivaliek, Jaarboek 2000 Stichting Archiefpublicaties ('s-Gravenhage 2000)

Hollingsworth, David - *The Workflow Reference Model* - Workflow Management Coalition (1995); <http://www.wfmc.org/standards/docs/tc003v11.pdf>

Horsman, Peter J. - *Archiveren. Een inleiding* - Stichting Archiefpublicaties, 's-Gravenhage (2004)
How Much Information? 2003 - onderzoeksrapport Universiteit van Berkeley (2003)
http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf

In paper we trust, why paper endured the digital Age: an investigation into the business, psychological and technological aspects of paper use - Lexmark international (2002)

Informatie: grondstof met toekomstwaarde - advies opgesteld door Raad van Cultuur en de Raad voor het Openbaar Bestuur (2008)
http://www.rob-rfv.nl/default.aspx?skin=Rob&inc=detail&nieuws_id=851&type=actueel

Information lifecycle management maturity model - Sun Microsystems white Paper (2005);
http://www.sun.com/storagetek/white-papers/ILM_Maturity_Model.pdf

Information Lifecycle Management: Nieuw buzzword of doordachte strategie? - In: VIP Vakblad voor documentmanagement, nr 5 juni (2004); pag 31-34
<http://www.reflux-txt.nl/ILM.pdf>

Jenkins, Tom - *Enterprise Content Management Technology, what you need to know* - Ontario (2006)

Jenkins, Tom; Forquer, Bill; Jelinski Peter - *Enterprise Content Management Solutions, what you need to know* - Ontario (2006)

Jenkins, Tom; Kölher, Walter; Shackleton, John - *Enterprise Content Management Methods, what you need to know* - Ontario (2006)

Kaderdocument, Afbakening functionaliteitsgebieden regionale midoffice architectuur Drechtsteden Digitaal, Versie 1.2 - (2006)
http://www.drechtstedendigitaal.nl/Dordrecht/up/ZiayqlnHiB_Kaderdocument_regionale_midoffice_concept_v1.2.pdf

Kampffmeyer, Ulrich - *ECM enterprise content management* - (2006)
http://www.project-consult.net/Files/ECM_White%20Paper_kff_2006.pdf

Kernmodel II, ervaring met en visie op informatiemanagement bij de rijksoverheid – Interlab, Den Haag (2006)
http://www.rijks-interlab.nl/Images/KernmodelII%20definitief%20met%20copyright_tcm134-94192.pdf

Klein, Johan - *Enterprise Content Management, trends voor 2005* - Lost Boys (2004)

Koenen, Kees; Baak, Paul - *Het gebeugen als actieve kracht, De archieffunctie binnen de digitale overheid* - (1999)
<http://www.rmconventie.nl/uploads/GeheugenAlsActiveKracht.pdf>

Ladley, John - *Beyond Rows and Columns, Unstructured Information, Part 3* - DM Review Online (2003)
http://www.dmreview.com/article_sub.cfm?articleId=6284

- Leer, Paul de - *ECM-market in beweging* - Verdonck, Klooster en Associates (2004)
http://www.jokietje.be/school/mmPublishing/Extra_Informatie/Extra_Informatie/040728_VI_P_ECM_PL.pdf
- Lesk, Michael - *Understanding digital libraries* - (second edition) San Francisco (2005)
- Logisch metadatamodel 'Functionaliteiten kernmodel Interlab' Versie 1.0* - Werkgroep Metadata Interlab (2004)
http://www.rijks-interlab.nl/Images/Kernmodel-LogischMetadatamodel-DI_tcm134-82198.pdf
- Mackenzie Owen, John - *Kennis in veelvoud* - Rede uitgesproken bij de aanvaarding van het ambt van hoogleraar in de Documentaire Informatiewetenschap aan de Universiteit van Amsterdam op woensdag 24 juni 1998
<http://igitur-archive.library.uu.nl/DARLIN/2005-0130-130700/ORATIE02.pdf>
- Mackenzie Owen, John - *The scientific article in the age of digitization* - (2005)
<http://dare.uva.nl/document/17843>
- McKemmish, Sue - *Yesterday, today and tomorrow: a continuum of responsibility* - (pag. 195-210) In: Horsman, Peter J.; Ketelaar, Frederick C.J.; Thomassen, Theo H.P.M. (red.) - *Naar een nieuw paradigma in de archivistiek*. Jaarboek 1999 Stichting Archiefpublicaties ('s-Gravenhage 1999)
- Meester, G. J. - *Collaboration Tools* - (2002)
<http://www.betterbe.com/docs/Colabtools.pdf>
- Mensink, Dennis - *Je zit op een goudmijn maar je weet het niet* - Facto magazine nr. 4 (2005)
- Möller, Jan; Bogaarts, Jacques; Nijborg, Hans - *Minimum functionele eisen voor record managementapplicatiesoftware* - Ministerie van Verkeer en Waterstaat (1998)
- Mouris, Harry - *Interdepartementaal ordeningsmodel, versie 1.0* - Interlab (2005)
http://www.rijks-interlab.nl/Images/Rapport%20Interlab%20ordeningsmodel%201-0_tcm134-82203.pdf
- Munkvold, Bjørn; Erik, Päivärinta; Tero, Hodne; Anne, Kristine; Stangeland, Elin - *Contemporary Issues of Enterprise Content Management: The Case of Statoil* - In: *Scandinavian Journal of Information Systems*, 2006, 18(2):69-100
- NEN-ISO 15489-1(nl), Informatie en documentatie – Archiefbeheer. Deel 1: Functionele eisen voor informatie en archiefmanagement* - Nederlands Normalisatie-instituut (2001)
- NEN-ISO 23081-1:2006 information and documentation – record managementprocesses – Metadata for records – part 1: Principles* - Nederlands Normalisatie-instituut (2006)
- Newman, David - *Business Drivers and Issues in Enterprise Information Management (Gartner RAS Core Research Note G00129712)* – Gartner (2005)
<http://www.avanade.com/uploaded/pdf/avanadearticle4124441.pdf>

NORA 2.0, *Nederlandse Overheid Referentie Architectuur, Samenhang en samenwerking binnen de elektronische overheid* - Kenniscentrum (2007)

http://www.elektronischeoverheid.nl/data/files/architectuur/NORAv2_0.pdf

Notitie Zoekmogelijkheden, Ordeningsmodel en Documentbeheer, ten behoeve van Werkgroep DMS - Stadsdeel Amsterdam-Centrum, (2005)

http://digidiv.amsterdam.nl/handleidingen_en_instrumenten/documentaire_informatiesystemen/kenmerken_en_cases/zoeken_ordenen_en_beheren.doc

Orde op zaken, zaken op orde, rendement met informatiemanagement (versie 8) - Koenen Baak en Partners & VHIC (2003)

http://digidiv.amsterdam.nl/div_nu_en_straks/elektronische_overheid/ozo.doc

Olij, P.M. - *Deelproject Proeftuin, Functioneel ontwerp DMS* - LogicaCMG (2007)

Poppe, Jeroen - *Metadata: ISO 23081 en andere standaarden (2)* – In: OD 2006 nr. 10

Prins, J.E.J.; Matthijssen, L.J. - *De Digitale Overheid en de wet, de juridische kaders voor gebruik van digitale documenten bij overheden* - Den Haag (2000)

Regeling duurzaamheid archiefbescheiden - Uit: Staatscourant 18 september 2001, nr. 180 / pag. 10

http://www.nationaalarchief.nl/images/3_2562.pdf

Reiner, D. - *I Have to Show Them What?! E-Mail and the process of electronic discovery* - In: Information storage and security journal, 06-2005

Reiner, David; Press, Gil; Lenaghan, Mike; Barta, David; Urmston Rich - *Information lifecycle management: the EMC perspective* - in: Data Engineering, 2004. Proceedings of the 20th international conference on data engineering (ICDE'04); pag. 804- 807

Rockley, Ann; Kostur, Pamela; Manning, Steve - *Managing enterprise content, a unified content strategy* - Berkeley (2003)

Rooij, R. de - *Maak meer mogelijk met metadata, het DNA van een website* – Naar voren (2005)

<http://www.naarvoren.nl/artikel/metadata/>

Shegda Karen M.; Bell, Toby; Chin, Kenneth; Gilbert, Mark R. - *Magic Quadrant for Enterprise Content Management (Gartner RAS Core Research Note G00150426)* - Gartner (2007)

<http://mediaproducts.gartner.com/gc/reprints/ibm/external/volume2/article16/pdf/article16.pdf>

Short, James E. - *Information Lifecycle Management: An Analysis of End User Perspectives* - The Information Storage Industry Center Graduate School of International Relations and Pacific Studies University of California (2006)

http://www.storagenetworking.org/pdfs/ISIC_ILM_WP_06-01_1.pdf

Starre, Jan van der - *Omvallende bitjes: ons digitaal geheugen bedreigt* - Informatie professional 2002 (6)

Strategic archiving, using information lifecycle management to archive data more efficiently and comply with new regulations - Sun Microsystems (2005)

http://www.sun.com/storagetek/white-papers/index.xml?displayPage=9&sort=modify_date&order=false&product=null

Storage optimization, Information lifecycle management - Sun Microsystems (2005)

http://www.sun.com/storagetek/white-papers/Storage_Optimization_for_ILM.pdf

Syllabus VHIC faculty - *Leergang aankomend records manager* - VHIC

Testbed Digitale Bewaring - *Functionele specificaties bewaarsysteem (versie 1.0)* - Den Haag (2005)

http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Functionele_specs_bewaarsysteem_v1.pdf

The ECM Suites Report 2008, Comprehensive Product Evaluations (sample edition version 2) - CMS Watch (2008)

Themistocleous, Marinos; Irani, Zahir; Kuljis, Jasna; Love, Peter E.D. - *Extending the Information System Lifecycle through Enterprise Application Integration: A Case Study Experience* - Proceedings of the 37th Hawaii International Conference on System Sciences (2004)

Waeyenbergh, S. - *Standaarden voor dynamisch archiefbeheer* - In *Bibliotheek- en archiefgids* 80 (2004) 2; pag. 10-17, alhier pag. 11

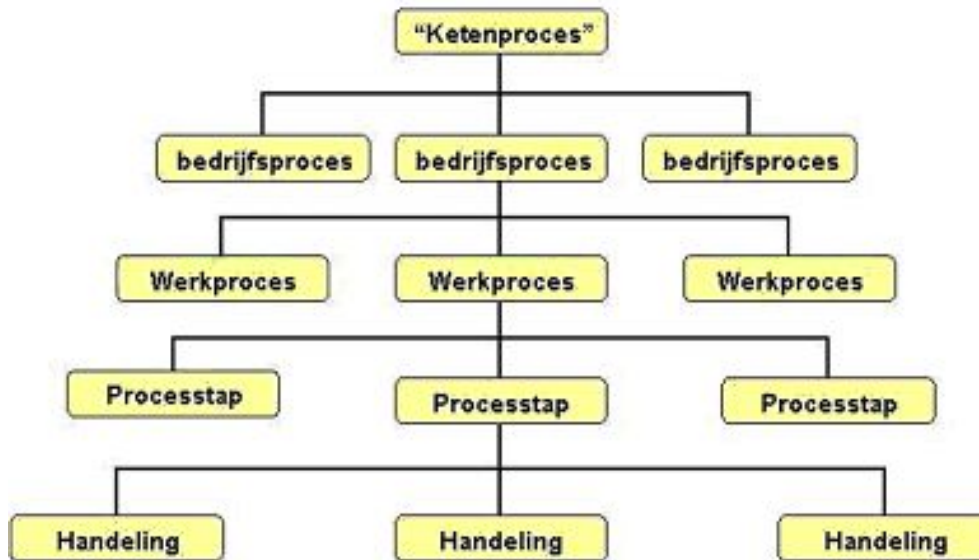
<http://www.vvbad.be/files/690-standaarden.PDF>

Ying, Chen; Shauchi, Ong - *Holistic information management solutions* - IBM (2005)

Zaken in zicht, GFO-ZAKEN - VNG Uitgeverij, Den Haag, (2004)

<http://egem-iteams.nl/system/files/gfozakeninzicht.pdf>

Bijlage 1: Procesbeschrijving



Ketenproces	Een ketenproces is een geordende reeks services die door verschillende organisaties aan elkaar worden geleverd met als doel om via één organisatie een (combinatie van) dienst(en) te leveren aan een burger of een bedrijf.
Bedrijfsproces	Een bedrijfsproces is een geordende reeks werkprocessen die binnen één organisatie wordt uitgevoerd met als doel om een (combinatie van) dienst(en) te leveren aan een burger, bedrijf of andere organisatie.
Werkproces	Een geordende reeks van processtappen die binnen één organisatorische eenheid binnen een organisatie wordt uitgevoerd met als doel een specifieke bijdrage (prestatie) te leveren aan een dienst die uiteindelijk zal worden geleverd aan een burger, een bedrijf of een andere organisatie.
Processtap	Een geordende reeks handelingen die ononderbroken wordt uitgevoerd door één mens of machine binnen één bedrijfsfunctie.
Handeling	Kleinst mogelijke eenheid van werk, uitgevoerd door één persoon of machine op één plek op één moment (eenheid van tijd, plaats en handeling). ²⁴³

²⁴³ NORA 2.0, *Nederlandse Overheid Referentie Architectuur, Samenhang en samenwerking binnen de elektronische overheid* - Kenniscentrum (2007) pag. 96
http://www.elektronischeoverheid.nl/data/files/architectuur/NORAv2_0.pdf

Bijlage 2: Britney Spears

488941 britney spears	29 britent spears	9 brinttany spears	5 brney spears	3 britiy spears	2 brirreny spears
40134 brittany spears	29 brittnany spears	9 britanay spears	5 broitney spears	3 britmeny spears	2 brixrany spears
36315 brittney spears	29 britttany spears	9 britinany spears	5 brotny spears	3 britneeyy spears	2 bixrttany spears
24342 britany spears	29 btiney spears	9 britn spears	5 bruteny spears	3 britnehy spears	2 bixrttney spears
7331 britny spears	26 birttney spears	9 britnew spears	5 btiinye spears	3 britnely spears	2 britain spears
6633 briteny spears	26 breitney spears	9 britneyn spears	5 brittney spears	3 britnesy spears	2 britane spears
2696 brittney spears	26 brinity spears	9 britney spears	5 gritney spears	3 britnetty spears	2 britaneny spears
1807 briney spears	26 britenay spears	9 brtiny spears	5 spritney spears	3 britnex spears	2 britania spears
1635 brittney spears	26 britneyt spears	9 brittney spears	4 bittny spears	3 britneyxxx spears	2 britann spears
1479 brintey spears	26 brittan spears	9 brtny spears	4 bnritney spears	3 britnity spears	2 britanna spears
1479 britanny spears	26 brittne spears	9 brytny spears	4 brandy spears	3 britntey spears	2 britannie spears
1338 brittany spears	26 brittany spears	9 rbitney spears	4 rbitney spears	3 britntey spears	2 britannt spears
1211 britnet spears	24 beitney spears	8 birtiny spears	4 breatingy spears	3 britttxny spears	2 britannu spears
1096 brittney spears	24 birtney spears	8 bithney spears	4 breetney spears	3 brittneyy spears	2 britanyl spears
991 britaney spears	24 brightney spears	8 brattany spears	4 bretiney spears	3 brittney spears	2 britanyt spears
991 britnay spears	24 brintiny spears	8 breitny spears	4 brfitney spears	3 britntey spears	2 britaney spears
811 brithney spears	24 britanty spears	8 breteny spears	4 rbrittany spears	3 brityny spears	2 britenany spears
811 brtiney spears	24 britenny spears	8 brightny spears	4 brieteny spears	3 brityney spears	2 brixenet spears
664 birtney spears	24 britini spears	8 birtay spears	4 brietny spears	3 britny spears	2 briteny spears
664 brittney spears	24 britny spears	8 brittty spears	4 briitny spears	3 broteny spears	2 britenyys spears
664 britney spears	24 brittni spears	8 briotney spears	4 britttany spears	3 britny spears	2 britianey spears
601 bitney spears	24 brittnie spears	8 britanys spears	4 brinie spears	3 brtliany spears	2 britan spears
601 brinty spears	21 brittney spears	8 britley spears	4 brinteny spears	3 brtinay spears	2 britinary spears
544 brittaney spears	21 britany spears	8 birtney spears	4 brintne spears	3 britneyy spears	2 britny spears
544 brittnay spears	21 biteny spears	8 birtney spears	4 britaby spears	3 brtitany spears	2 britnany spears
364 britey spears	21 bratney spears	8 brinty spears	4 britay spears	3 brtitny spears	2 britnat spears
364 brittany spears	21 britani spears	8 brittner spears	4 britainey spears	3 brtnet spears	2 britnbey spears
329 brtney spears	21 britanie spears	8 brottany spears	4 britinie spears	3 brytiny spears	2 britndy spears
269 bretney spears	21 briteany spears	7 haritney spears	4 britinny spears	3 btney spears	2 britneh spears
269 britneys spears	21 brittay spears	7 birtney spears	4 britmney spears	3 drittney spears	2 britnemy spears
244 britne spears	21 brittinay spears	7 biteney spears	4 britnear spears	3 pretney spears	2 britney6 spears
244 brytney spears	21 brtany spears	7 bitiny spears	4 britnel spears	3 rbritney spears	2 britneye spears
220 breatney spears	21 britiany spears	7 breateny spears	4 britneuy spears	2 bbrittany spears	2 britneyh spears
220 brittany spears	19 birney spears	7 brianty spears	4 briutney spears	2 hbrittney spears	2 britneym spears
199 brittney spears	19 birtney spears	7 brintye spears	4 britmney spears	2 hbitney spears	2 britneyyy spears
163 britny spears	19 britnaey spears	7 britianny spears	4 brittaby spears	2 hbrittny spears	2 britneyt spears
147 breatny spears	19 britnee spears	7 britly spears	4 brittery spears	2 hbrittany spears	2 britnjoy spears
147 brittney spears	19 britony spears	7 britney spears	4 brittthey spears	2 beitany spears	2 brittne spears
147 britty spears	19 brittany spears	7 britneyu spears	4 brittney spears	2 beitny spears	2 britnu spears
147 brotney spears	19 brittney spears	7 britnley spears	4 brittntat spears	2 bertney spears	2 britoney spears
147 brutney spears	17 birtny spears	7 britnny spears	4 brittney spears	2 bertny spears	2 brittany spears
133 brittney spears	17 brieny spears	7 brittitan spears	4 brittnye spears	2 betney spears	2 britreny spears
133 brinyey spears	17 brintty spears	7 briyny spears	4 brittteny spears	2 betny spears	2 britry spears
121 bittany spears	17 brithy spears	7 brrittany spears	4 briutney spears	2 bhriney spears	2 briteany spears
121 bridney spears	17 brittanie spears	7 brttiney spears	4 briyeny spears	2 biney spears	2 brittanay spears
121 britainy spears	15 brinney spears	7 briteny spears	4 brnity spears	2 bintey spears	2 brittang spears
121 britmey spears	15 briten spears	7 brrittany spears	4 brtteny spears	2 birtny spears	2 britttans spears
109 brietney spears	15 brixerney spears	6 beritny spears	4 brtttany spears	2 biritany spears	2 britttanyh spears
109 brithny spears	15 britheny spears	6 bhritley spears	4 bryney spears	2 birittany spears	2 britttanyyn spears
109 britni spears	15 britneny spears	6 birthney spears	4 brythney spears	2 birittny spears	2 britttany's spears
109 brittant spears	15 britttany spears	6 breathney spears	4 brytne spears	2 birnty spears	2 britttanyt spears
98 bittney spears	15 brittmey spears	6 breaty spears	4 brytni spears	2 birtey spears	2 britttanyy spears
98 brithey spears	15 brytnei spears	6 bretany spears	4 brytnie spears	2 birtheny spears	2 britttany spears
98 brittiany spears	15 btrney spears	6 briatany spears	4 bvritney spears	2 birttany spears	2 britttenie spears
98 bitney spears	15 rittney spears	6 brint spears	4 dritney spears	2 birtny spears	2 britttenty spears
89 brietny spears	14 brinet spears	6 britteney spears	4 priteny spears	2 birtnet spears	2 britttinney spears
89 brinetny spears	14 britneyy spears	6 britian spears	3 beittany spears	2 bitnet spears	2 brittley spears
89 brintny spears	14 britten spears	6 britinty spears	3 bichney spears	2 bitttany spears	2 britton spears
89 britnie spears	12 beritney spears	6 britittany spears	3 biritny spears	2 bnrittany spears	2 britttney spears
89 brittney spears	12 britiny spears	6 britnsy spears	3 birnety spears	2 bntney spears	2 britttneyt spears
89 brittnet spears	12 briatny spears	6 britrey spears	3 birny spears	2 bpitney spears	2 britttney's spears
89 brity spears	12 brieney spears	6 britsny spears	3 birteney spears	2 brandi spears	2 britttneyu spears
89 ritney spears	12 brinany spears	6 brittine spears	3 birthny spears	2 brattney spears	2 britttneye spears
80 bretny spears	12 britaany spears	6 brittny spears	3 birtiney spears	2 breanty spears	2 britttany spears
80 britnany spears	12 britan spears	6 brittssy spears	3 birttteny spears	2 breany spears	2 britttany spears
73 brittney spears	12 britine spears	6 brittany spears	3 birttny spears	2 breittany spears	2 britttney spears
73 brittainy spears	12 britnea spears	6 britttany spears	3 bitanny spears	2 brently spears	2 brittyne spears

Figuur 54: zoektermen 'Britney Spears' bij Google

Bijlage 3: Records Continuüm model

Het records continuüm model gaat uit van vier dimensies:

1. Creëren

De eerste dimensie omvat de stap in het proces, de personen die een stap in een proces uitvoeren en de neerslag (representatie) van de stap in het proces. De vastlegging van proceshandelingen gebeurt in het document en in de gegevens die over het document en de handelingen die ermee zijn gebeurd hebben plaatsgevonden: de bijbehorende context en metadata. De representatie omvat dus zowel het informatieproduct als de bijbehorende metadata. De representatie is in de eerste dimensie nog niet onder het archiefregime gebracht.

2. Verwerven/opnemen

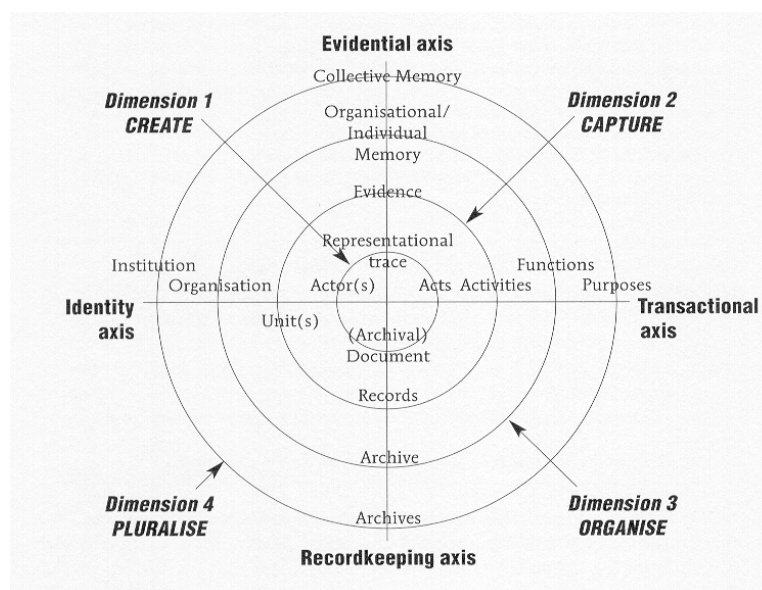
De tweede dimensie omvat zowel de persoonlijke archiefsystemen als de archiefsystemen op bedrijfsniveau. Deze systemen leggen op een zodanige manier de informatie en de context vast dat op basis daarvan ondersteuning kan worden geboden bij het afleggen van verantwoording over het handelen. Deze documenten worden tot archief verklaard, de documenten (en de bijbehorende metadata) dienen eenduidig en onveranderbaar te worden opgeslagen. Om dit te waarborgen worden de documenten en metagegevens onder het archiefregime gebracht.

3. Organiseren

De derde dimensie omvat de organisatie (vormgeving) van de archiveringsprocessen. Deze dimensie betreft de wijze waarop een organisatie haar archiefregime ziet en het archief realiseert als geheugen van haar organisatiedoelen.

4. Verbreden

De vierde dimensie ontstijgt de archiefvormer, ongeacht of het een individu of een organisatie betreft. Het gaat hier over de wijze waarop archieven in een groter geheel worden ingebracht om gezamenlijk een collectief (nationaal) sociaal, historisch en cultureel geheugen te vormen.²⁴⁴



Figuur 55: records continuüm model

²⁴⁴ Arnold e.a. - *De moderne informatiehuishouding* - pag. 29

Dimensie één en twee zijn de zogenaamde ‘implementatie dimensies’. De betrokken personen in deze dimensie zijn de uitvoerders van de werkprocessen. Doel van deze dimensies is om de neerslag van de uitvoering van de werkprocessen zodanig te verkrijgen en vast te leggen dat het geheel als bewijs kan dienen. De derde en de vierde dimensie hebben betrekking op de inrichting, standaardisering, beheersing, regulering, inspectie en controle. Bij deze dimensies zijn voornamelijk de bestuurders en archiefmanagers betrokken. De bestuurders als dragers van de verantwoordelijkheid voor het archiefregime en de archiefmanagers als verantwoordelijken voor het vormen, onderhouden en verschaffen van toegang tot het institutionele geheugen. De vierde dimensie betreft de vorming van een collectief geheugen op een wijze die de grenzen van organisaties overstijgt.²⁴⁵

Vertaald naar de praktijk zou het model als volgt kunnen worden geïnterpreteerd. Het zal in eerste instantie een individu zijn die de uitvoering van het bedrijfsproces documenteert om daarover verantwoording te kunnen afleggen (dimensie 1). Vervolgens zullen deze documenten worden opgenomen in het archiveringssysteem van de afdeling (dimensie 2) dat weer onderdeel uitmaakt van het archiefsysteem van de gehele organisatie (dimensie 3). De organisatie maakt op haar beurt weer onderdeel uit van een groter verband en kan om die reden beslissen het archiefbestand te laten opnemen in een historische archiefcollectie (dimensie 4). In tegenstelling tot het levenscyclus van analoge archiefdocumenten hoeft dit echter niet te betekenen dat een bestand fysiek wordt overgedragen aan een archiefbewaarplaats of dat het bestand niet meer gebruikt/geraadpleegd wordt door de organisatie zelf.

“The dimensions of the continuum are not time-based. Records are both current and historical from the moment of their creation. By definition they are ‘frozen’ in time, fixed in a documentary form and linked to their context of creation. They are thus time and space bound, perpetually connected to events in the past. Yet they are also disembedded, carried forward through time and space, and re-presented in the contexts of their use.”²⁴⁶

²⁴⁵ McKemmish, Sue - *Yesterday, today and tomorrow: a continuum of responsibility* (pag. 195-210) in: Horsman Peter J., Ketelaar Frederick C.J. en Thomassen Theo H.P.M. (red.) - *Naar een nieuw paradigma in de archivistiek. Jaarboek 1999 Stichting Archiefpublicaties* (’s-Gravenhage 1999), alhier 202-203

Arnold e.a. - *De moderne informatiehuishouding* - pag. 29

²⁴⁶ McKemmish, Sue – *Yesterday, today and tomorrow: a continuum of responsibility*; pag. 203