

ELEKTRONISCH BEHEER VAN DE DOCUMENTEN

*Le monde est fait pour aboutir à un beau livre¹
(Stéphane Mallarmé)*


1. Inleiding



*Hervé Haut
is burgerlijk ingenieur en
Doctor in de Natuur-
wetenschappen. Na
verschillende jaren
onderzoek te hebben
verricht naar de
theoretische fysica aan
de UCL, bekleedde hij
diverse informatica-
functies in de privé-
sector. Hij begon bij de
SmalS-MvM in 1998 als
consultant bij de sectie
Onderzoek waar hij
vooral consultancy-
opdrachten vervult voor
projecten i.v.m.
documentair beheer.
Contact : 02/509.58.16*

Dit document wil een algemeen overzicht geven van het geheel van de tools die men kan moeten gebruiken wanneer men overgaat tot de invoering van het elektronisch beheer van documenten. We zullen zien hoe deze tools onderling geïntegreerd worden om de documenten van een onderneming zo goed mogelijk te beheren, ook in het dynamisch aspect van dit beheer dat, soms, invloed heeft op de werkprocedures.

1.1 Een beetje geschiedenis

 l meer dan tweeduizend jaar vormt het schrift het belangrijkste instrument om de menselijke kennis over te dragen. Sinds de steen van Rosetta en de manuscripten van de Dode Zee (die dateren van ongeveer twee eeuwen voor onze jaartelling), heeft de mens heel wat middelen ontwikkeld om zijn gedachten onsterfelijk te maken (papyrus, velijn, stoffen, papier, gravures op steen, op hout, daarna op metaal,...). De uitvinding van de boekdrukkunst voor Gutenberg maakte de geleidelijke ontwikkeling mogelijk van de druktechnieken en vandaag beschikken wij over een ontelbaar aantal documenten die getuigen van onze geschiedenis.

Reeds vroeg in onze geschiedenis ging men bewust de documenten die belangrijk geacht werden in de verschillende perioden verzamelen, organiseren en beheren. Zo ontwikkelden zich onze grote bibliotheken, van de legendarische Bibliotheek van Alexandrië tot de prestigieuze bibliotheken van vandaag. Zo ontwikkelden zich ook de scriptoria. Dit zijn ruimten die voorbehouden werden voor het schrijven van manuscripten in de kloosters en het waren de voornaamste plaatsen waar de literaire kennis werd doorgegeven tijdens het grootste deel van de Middeleeuwen.



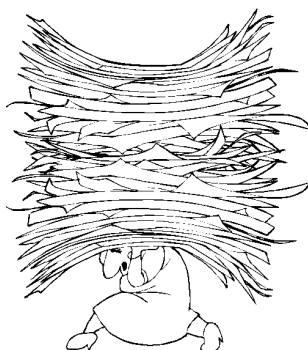
¹ De wereld is gemaakt om tot een mooi boek te komen

In onze bedrijven is het document vandaag nog steeds de belangrijkste opslagplaats van niet-gestructureerde informatie en het vertegenwoordigt de kracht van onze organisaties, hun communicatievermogen zowel binnen als buiten. En toch, ondanks hun strategisch belang, blijven de documenten één van de minst goed beheerde en minst geautomatiseerde resources in de onderneming die overigens steeds meer geïnformatiseerd wordt.

Tegenstrijdig genoeg is het de ontwikkeling van de techniek en de informatica die tot deze situatie geleid heeft.

Voor het verschijnen van het fotokopieerapparaat vond men in de meeste ondernemingen nog één of meer archivariissen die garant stonden voor het goed beheer van de papieren archieven en, in zekere zin, waakten over de kennis van de organisatie. Dankzij de ontwikkeling van het fotokopieerapparaat kon men de documenten gemakkelijk verveelvoudigen en afstappen van de verplichting tot een unieke en strikte archivering maar dit leidde wel tot veelvuldige en redundante archiveringen, wat vaak synoniem is van desorganisatie van de documentatie en van verspreiding van de kennis. De functie van archivaris is praktisch verdwenen, elke bediende besteedt een deel van zijn tijd aan deze functie, met zijn eigen methode, met een eerder persoonlijke dan organisatorische motivatie.

1.2 Het elektronisch documentenbeheer, archivaris van de moderne tijden ?



De ontwikkeling van de informatica heeft de situatie alleen maar verergerd. Zodra de computers verschenen, zorgde de eenvoud waarmee men listings kon aanmaken voor volumes papier die steeds moeilijker te beheren zijn. Een studie van 1996 in de Verenigde Staten stelde "dat het geheel van de Amerikaanse bedrijven ongeveer 90 miljard documenten per jaar produceert, dat deze documenten gemiddeld 11 keer gekopieerd werden wat leidt tot ongeveer 1000 miljard pagina's; mocht men dit papier opeenstapelen, dan zou men een hoogte van 300.000 kilometer bekomen!".

De SmalS-MvM recycleert elk jaar een hoeveelheid papier die overeenkomt met 30.000 pagina's per werknemer! We zullen zien dat men gemakkelijk meer economische oplossingen zou kunnen invoeren.

Al snel zag men in dat het noodzakelijk was de zogenaamde "gestructureerde" gegevens te organiseren en we zagen hoe alle database-tools om dit soort informatie te organiseren en bruikbaar te maken zich ontwikkelden.

Op het niveau van de documenten die aangemaakt worden in de onderneming (die men algemeen als "niet-gestructureerde" gegevens beschouwt), heeft men moeten wachten tot de jaren negentig om de geleidelijke ontwikkeling en invoering te zien van tools die toelaten deze documentaire massa elektronisch te beheren.

In het vervolg van dit artikel zullen wij deze tools voor Elektronisch documentenbeheer (EDM, Electronic Document Management) detailleren en de voornaamste modules van het documentair beheer overlopen. Het is onze bedoeling de bestaande tools te beschrijven, hun functionaliteit te preciseren en te tonen hoe zij onderling geïntegreerd worden. We zullen op een algemeen niveau blijven zonder in detail te treden over bijzondere tools en zonder een vergelijking te maken tussen tools van verschillende constructeurs.

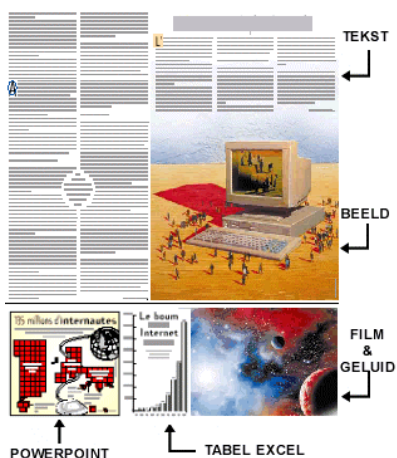


2. Het document in de onderneming

Alvorens te praten over het beheer van documenten, is het nuttig te definiëren wat men vandaag verstaat onder document, te wijzen op het belang ervan in de onderneming en de belangrijkste normen te vermelden die er momenteel zijn op het vlak van documenten.

2.1 Wat is een document?

Van Dale geeft als definitie van een document: "*oorkonde, ieder geschreven of gedrukt bewijsstuk, bescheid*". Het is waarschijnlijk de afbeelding van een papieren geschrift die ons eerst voor de geest komt wanneer we aan een document denken. Nochtans heeft de ontwikkeling van de digitale technieken de manieren om documenten weer te geven aanzienlijk uitgebreid door er ook bestanden in op te nemen die geproduceerd worden door bureauticootools (tekstverwerkers, rekenbladen,...), beelden, video's, geluid,... we zullen het dan hebben over samengestelde documenten (compound documents).



Binnen éénzelfde document zal men steeds vaker informatie vinden die afkomstig is van verschillende bronnen zoals, bijvoorbeeld, een tekst afkomstig uit een rapport, met een tabel die opgebouwd is op basis van een gegevensbank en illustraties die gescand werden in een commerciële brochure.

Deze samengestelde documenten beperken zich niet meer tot publicatie op papier maar worden vaak ook bewaard en verspreid onder elektronische vorm. In deze elektronische vorm zullen vaak navigatie-elementen (hyperlinks) toegevoegd worden in het document (of in een geheel van documenten) om het lezen ervan interactiever te maken.

2.2 Het belang van het document

Ook al is de weergave ervan geëvolueerd, het document, in de brede zin waarin wij het voortaan zullen bekijken, behoudt al zijn waarden als informatie, inlichting of getuigenis en het belang ervan binnen de onderneming groeide onafgebroken.

Men schat namelijk dat de informatie in de onderneming in het algemeen bestaat uit 20% gestructureerde gegevens (gegevensbanken,...) en 80% niet-gestructureerde gegevens, d.w.z. documenten. Deze massa niet-gestructureerde informatie raakt verspreid in de onderneming en wordt meestal "beheerd" door individuen zonder collectieve motivatie. Bovendien worden deze documenten meestal (één of meerdere keren!) "geklasseerd" en gebruikt onder papieren vorm hoewel de grote meerderheid ervan afkomstig is van de informatica.

Dit kan leiden tot belangrijke financiële verliezen als men het volgende weet²:

- ◇ elke bediende gebruikt slechts 65% van de in zijn kasten geklasseerde documenten;

² Business Week 1995.



- ◇ het klasseren en opzoeken van een document kost gemiddeld 20\$ en deze kosten kunnen oplopen tot 120\$ als het slecht geklasseerd is;
- ◇ de bediende besteedt tussen 10 en 40% van zijn tijd aan het opzoeken van of wachten op informatie;
- ◇ ook al worden de documenten bewaard onder elektronische vorm, dan nog zijn zij meestal alleen toegankelijk voor hun auteur.

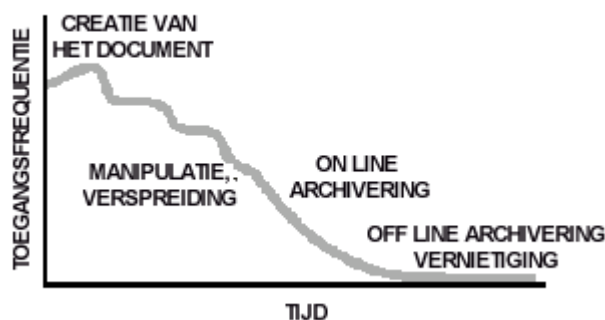
Hierbij komen ook de kosten om de veroorzaakte volumes aan papier op te slaan, het gevaar dat de onderneming kapitale informatie verliest, het verlies van het unieke karakter van de informatie door duplicering van documenten, het risico dat er mogelijk ongevalideerde documenten rondgaan of dat vertrouwelijke gegevens ongecontroleerd verspreid worden.

We kunnen deze kosten en risico's moeilijk ramen maar de bovenstaande beschouwingen tonen aan dat de meeste maatschappijen veel zouden kunnen winnen mochten zij alles ondernemen om hun niet-gestructureerde informatie te beheren zoals hun gestructureerde informatie. Dit is het doel van de tools voor het Elektronisch documentenbeheer.

We gaan de tools overlopen die de moderne technologie ons ter beschikking stelt en we volgen daarbij de levenscyclus van een document.

2.3 De levenscyclus van een document

Alle documenten hebben een levenscyclus die opgedeeld kan worden in vier fasen: de creatie van het document, een "manipulatie"-fase waarin het document wordt afgewerkt door opeenvolgende bijwerkingen (door de auteur of door een geheel van personen), een fase waarin het gearchiveerd wordt en gemakkelijk toegankelijk blijft zonder echter nog belangrijke wijzigingen te ondergaan en ten slotte een laatste stadium waarin het ofwel "off line" gearchiveerd wordt, ofwel vernietigd wordt.



Deze verschillende fasen komen overeen met verschillende tools voor documentair beheer. Meestal zijn deze tools onderling geïntegreerd rond een "repository", d.w.z. rond een gegevensbank die het geheel van de documentatie beheert.



3. De creatie van de documenten

We gaan niet uitweiden over de bureauticootools waarmee we documenten kunnen creëren. We gebruiken allemaal een tekstverwerker waarmee we tabellen van ons rekenblad kunnen opnemen, alsook beelden of meer algemeen om het even welk object afkomstig van een andere toepassing. Er bestaan uiteraard tools om veel complexere documenten te maken, gebruikt door specialisten in de uitgeverwereld maar we treden hierover niet in detail.

Naast deze software voor de "directe" creatie van elektronische documenten zullen wij dieper ingaan op twee andere tools waarmee men een documentaire gegevensbank kan voeden op basis van papieren documenten en listings die afkomstig zijn van mainframes.

Maar eerst willen wij de nadruk leggen op het belang van de documentaire normen.

3.1 De documentaire normen

Binnen een onderneming legt men meestal een standaard op voor bureauticootools zoals de tools van Microsoft bij de SmalS-MvM. Hiermee lost men uiteraard het probleem op van de communicatie en verdeling van documenten binnen de onderneming.

Maar wat de garanties voor de duurzaamheid van de informatie betreft, is het verre van voldoende. We weten allemaal dat bepaalde documenten soms zeer lang bewaard moeten worden, zeker in de sector waarin wij werken. Welke garantie hebben wij dat een Word-document dat vandaag opgeslagen wordt, binnen een twintigtal jaren nog "leesbaar" zal zijn? Is het redelijk te denken dat Microsoft dan nog tools zal aanbieden die compatibel zijn met de tools van vandaag? Het antwoord is uiteraard neen en dat is de reden waarom men een duurzaam en draagbaar formaat zou moeten kiezen waarbij men zich afstemt op "universeel erkende" normen om de documenten te archiveren waarvan de levensduur meer dan enkele jaren bedraagt.

Het is niet altijd eenvoudig om zijn weg te vinden in de wereld van de standaarden: er bestaan talrijke (zowel nationale als internationale) normaliseringsorganisaties. Sommige van hen leggen standaarden op, andere nemen genoegen met de aanbeveling ervan. Naast deze organisaties ontwikkelen sommige informaticamaatschappijen ook hun eigen formaten die soms door de markt uitgekozen worden als de facto standaard. Verder in deze paragraaf zullen wij enkele van deze standaarden vermelden zonder exhaustief te willen zijn.

Voor de inhoud van het document komen op de markt drie standaarden naar voor:

- ◇ **SGML** (Standard Generalized Markup Language) is een ISO-standaard (International Standard Organisation, een organisatie van de Verenigde Naties die standaarden publiceert). Dit is vandaag de meest flexibele en de meest verfijnde standaard voor het opbouwen van complexe documenten die losstaan van elke verkoper en overgedragen kunnen worden op de meeste operating systems. Het belangrijkste kenmerk van SGML ligt in het feit dat de aldus geproduceerde documenten zichzelf beschrijven: zij bevatten een volledige beschrijving van het lichaam van het document zowel vanuit het standpunt van de vorm als van de structuur (bijvoorbeeld de titelniveaus, de referentietabel,...).



- ◇ **XML** (eXtensible Mark-up Language) is een zeer vereenvoudigd subgeheel van de SGML-taal. Deze taal wordt voorgesteld door het consortium W3C (World Wide Web Consortium). Eén van de doelstellingen ervan is het verrijken van de **HTML**-taal (HyperText Mark-up Language) die men gebruikt in Internetpagina's, rekening houdend met de structuur van de documenten maar minder zwaar en moeilijk dan SGML.
- ◇ **PDF** (Portable Document Format) is een eigen formaat van Adobe gebruikt voor de elektronische publicatie van documenten. Het biedt het voordeel dat een document voorgesteld wordt in een formaat dat losstaat van de software die het gemaakt heeft, van de hardware en van het gebruikte besturingssysteem. De eenvoudige integratie in een Internetpagina, de gratis terbeschikkingstelling van de leestool (de reader Acrobat) van dit documenttype en de rijkdom aan zoek- en navigatiemogelijkheden maken het vandaag tot een echte standaard³. Hoewel het een eigen formaat is, wordt het perfect gedocumenteerd en wij zien het als een formaat met toekomst voor de archivering en publicatie van documenten (als vorm en inhoud van de documenten afgewerkt zijn).

Wat de beelden en het audiovisuele in het algemeen betreft, dient men de volgende ISO-standaarden te noteren:

- ◇ **TIFF** (Tag Image File Format) dat een industrieel gebruikt formaat is voor de transfer of het opslaan van beelden. Dit formaat kan voorkomen onder niet-gecomprimeerde vorm of onder gecomprimeerde vorm zonder verlies van gegevens (volgens de compressienormen CCITT groep III en IV).
- ◇ **JPEG** (Joint Photographic Experts Group) is een andere grafische standaard die gecomprimeerd wordt met meer of minder verlies van gegevens naargelang het gekozen compressiepercentage.
- ◇ Wat ten slotte de muziek en de numerieke video betreft, behoren de diverse **MPEG**-formaten (Moving Picture Experts Group) tot de meest gebruikte.

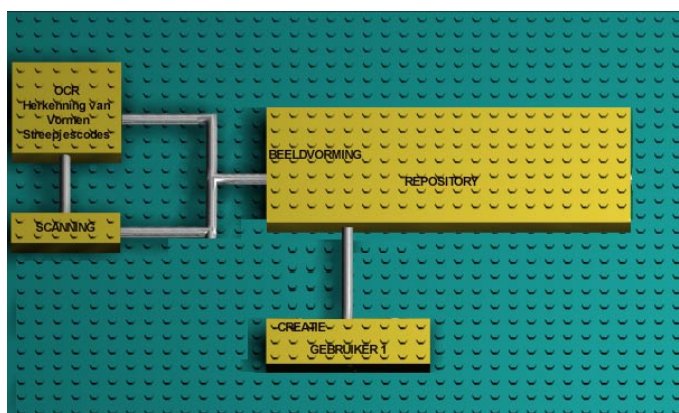
Naast deze standaarden verbonden met het formaat van het document is het nuttig om weten dat er andere standaarden bestaan met betrekking tot de onderlinge uitvoerbaarheid van de modules voor documentair beheer, maar we zullen deze standaarden hier niet detailleren.

3.2 De beeldvorming

In het centrum van een systeem voor documentair beheer vinden we een "repository": dit is de module die de gegevensbank bevat waarmee men het geheel van de documenten kan beheren; het is hij die instaat voor de integratie van de verschillende andere modules. Wij zullen hieraan een eerste module voor beeldverwerking (of captatie van documenten) toevoegen die gebruikt wordt om het documentair systeem te voeden vanaf papieren documenten. Dit is het eerste onderdeel in onze assemblage.

³ Bijvoorbeeld, de Gartner Group verspreidt onder zijn leden een cd-rom met de tekst van de conferenties van zijn jaarlijks symposium in het formaat PDF.





Het capteren van de documenten blijkt nuttig wanneer men interne documenten wil bewaren die alleen op papier bestaan of externe documenten die men op papier ontvangen heeft. Na het scannen zal men deze documenten opslaan in het systeem onder de vorm van een beeld. Deze vorm van opslaan heeft het nadeel dat hij veel meer schijf-

ruimte verbruikt dan een elektronisch document (men rekent meestal op 50 kbytes per zijde papier). Dit formaat zal echter verkozen worden voor de documenten waarvan het bestaan voor de onderneming belangrijker is dan de inhoud (een contract, een factuur, een notariële akte,...), daar deze inhoud vaak opgenomen is in andere gegevensbanken.

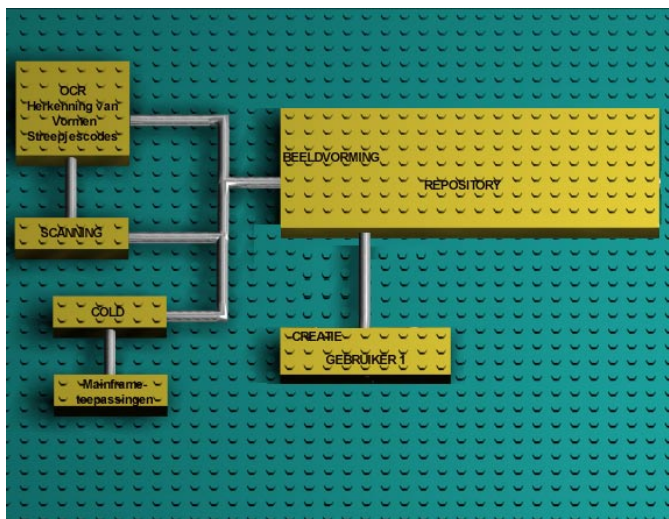
Wanneer het meer de inhoud van het document is die belangrijk is, dan zal men diverse technieken gebruiken om de inhoud van het document te identificeren:

- ◇ De OCR (Optical Character Recognition) of de herkenning van tekens laat toe de tekstuele inhoud van een document te extraheren door een elektronisch document aan te maken. Dit biedt twee voordelen : het vermindert aanzienlijk de grootte van het te stockeren document en het maakt dit bruikbaar voor standaardtools zoals tekstverwerkers. De tools die vandaag beschikbaar zijn op de markt, kunnen beschouwd worden als producten die goede resultaten opleveren voor zover men vertrekt van documenten die van goede kwaliteit zijn.
- ◇ De vormherkenning zal gebruikt worden wanneer men de inhoud van een formulier wil exploiteren. In dit geval kent men de plaatsen (de fysieke positie) van het formulier waar zich relevante informatie bevindt. De software voor vormherkenning laat toe deze informatie te extraheren, meestal om ze meteen op te slaan in de gegevensbanken waardoor het mogelijk wordt ze te exploiteren. Dit soort tool bestaat voor het herkennen van gedrukte formulieren, voor formulieren waarop men gedefinieerde hokjes moet aankruisen en ook voor formulieren waarvan de te herkennen zones manueel ingevuld zijn; in dit laatste geval liggen de foutenpercentages nog vrij hoog.
- ◇ Een andere wijze om de informatie van een document te coderen is er een streepjescode in opnemen bij het afdrukken ervan. De software voor het herkennen van streepjescodes geeft uitstekende resultaten en laat zeer vaak toe uit het document de "meta-informatie" te halen met het oog op de indexering ervan waarover wij het verder zullen hebben.



4.3 De COLD

Naast documenten die direct onder elektronische vorm geproduceerd zijn door bureautica-tools en documenten die via scanning in het systeem ingevoerd zijn, kan men met de COLD-tools (Computer Output to Laser Disk) in een documentair systeem listings (of geformatteerde rapporten) opslaan die geproduceerd zijn door mainframes (of andere computers).



Bij het afdrukken voeren de grote traditionele printers (zoals de Xerox-printer van het rekencentrum van de SmalS-MvM) een fusie uit tussen een bladachtergrond en gegevens afkomstig van de mainframe. De COLD-tools werken op een zeer gelijkaardige manier.

Als de door de mainframe gegenereerde rapporten in het documentair systeem gearhiveerd moeten worden, dan leidt men de

egevens naar het COLD-systeem en dit tegelijk met de informatie over de structuur van het document. Eerder dan de rapporten op papier af te drukken, stockeert het COLD-systeem deze rapporten op optische schijven (vanwaar zijn aanvankelijke benaming) of op om het even welk digitaal medium.

Dit soort tool biedt enorme voordelen en het is verbazend dat het niet vaker gebruikt wordt in de maatschappijen die elk jaar tonnen listings produceren:

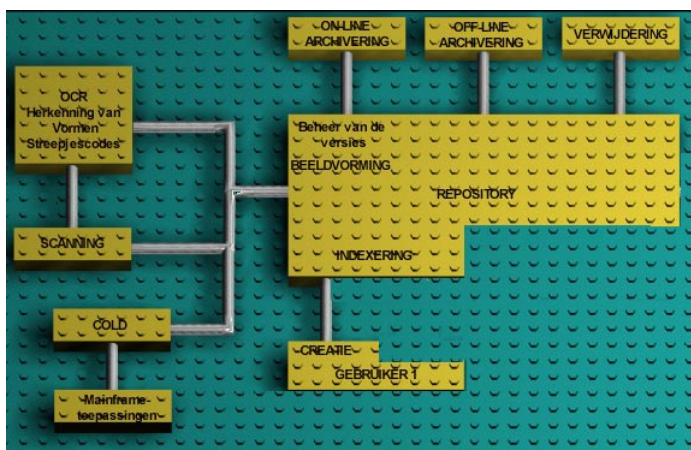
- ◇ Voor elk afdrukgeheel bevat het gecreëerde bestand de pagina-achtergrond (die zo slechts één keer opgeslagen is) en het geheel van de gegevens van het rapport; dit soort bestand is bijgevolg weinig omvangrijk.
- ◇ Door deze werkwijze kan men garanderen dat de listing op elk moment opnieuw aangemaakt kan worden zoals hij aanvankelijk was, ook al werd de pagina-achtergrond ondertussen gewijzigd.
- ◇ Aangezien de gegevens opgeslagen zijn onder digitale vorm, kunnen zij geëxploiteerd worden met tools voor gegevensanalyse gaande van een eenvoudig rekenblad tot geëvolueerde statistische tools. Verder laat deze stockeringswijze ook de automatische indexering toe van de pagina's van het rapport.
- ◇ Het is niet meer nodig een rapport volledig te verspreiden onder alle mensen die meestal slechts geïnteresseerd zijn in het deel waarmee zij moeten werken: dankzij de tools van het COLD-systeem kan elkeen die informatie extraheren die voor hem nuttig is zonder gehinderd te worden door gegevens waar hij niets mee te maken heeft.



- ◇ Het spreekt voor zich dat een COLD-systeem toelaat een groot aantal afdrucken te vermijden, waardoor heel wat minder papier verspild wordt.
- ◇ Tot slot tonen alle studies aan dat een COLD-systeem snel rendeert met een return on investment in de orde van 8 à 10 maanden!

4. De stockeringsoplossingen

De systemen voor documentair beheer zijn ontworpen om grote volumes aan documenten te beheren. Het is dus noodzakelijk gepaste middelen te voorzien voor het stockeren (of archiveren), alsook middelen om het document te verwijderen op het einde van zijn levenscyclus. Hieronder sommen wij de voornaamste kenmerken van de gebruikte media op.



Uiteraard is het de bedoeling efficiënt te kunnen werken met de gearcheiverde documenten. Het is dus belangrijk ze correct te organiseren om ze later eenvoudiger te kunnen opzoeken; elk document moet daarom nauwkeurig geïndexeerd worden alvorens men het archiveert in het systeem.

ot slot is het belangrijk in dit geheel van informatie de toegangsrechten correct te beheren om de nodige confidentialiteit te garanderen en voor iedereen slechts die informatie te openen waarop hij recht heeft.

Onze constructie wordt aangevuld met de modules voor indexering en stockering.

4.1 De stockeringsmedia

Traditioneel onderscheidt men drie stockeringsniveaus in functie van de gewenste toegangstijden.

De documenten waartoe men zeer vaak toegang moet hebben (voornamelijk gedurende de periode waarin zij "in wording" zijn) zullen bij voorkeur opgeslagen worden op magnetische schijven die een snelle toegangstijd bieden. De documenten die "on line" moeten blijven maar waarvoor men een iets langere recuperatietijd toelaat, zullen gearcheiverd worden op magnetische cassettes of optische schijven die gegroepeerd worden in jukeboxen (men spreekt van "near-line"-stockering). Tot slot zijn er de documenten die men gedurende een bepaalde tijd moet bewaren maar waartoe men slechts uitzonderlijk toegang heeft. Zij zullen gestockeerd kunnen worden op magnetische banden ("off-line"-archivering).

De toegangssnelheid van een medium is meestal omgekeerd evenredig met de capaciteit en de kostprijs ervan. Men zal dan ook altijd moeten trachten een juist evenwicht te vinden tussen de verwachte performantie en de aanvaardbare kosten. We moeten echter opmerken dat deze hiërarchie zich vandaag lijkt te wijzigen en dat het rendabel wordt om te werken met



oplossingen die uitsluitend gebaseerd zijn op magnetische schijven (en banden voor de off-line-archivering) ook al zijn de volumes aan documenten omvangrijk.

Andere verplichtingen dan performantie en kosten spelen een rol bij de keuze van het stockeringsmedium. Wanneer men bewijskracht wil invoeren voor de gearchiveerde informatie, d.w.z. garanderen dat de opgeslagen informatie trouw blijft aan het origineel en niet gewijzigd mag worden in de loop van de tijd, dan zal men er alle belang bij hebben stockeringsmedia te gebruiken die deze garantie bieden omwille van hun fysieke natuur: de CD-WORM's (Write Once Read Many) waarop men slechts één enkele keer kan schrijven.

In een nabije toekomst is het waarschijnlijk dat de DVD-technologie (digital versatile disk) de huidige optische schijven zal vervangen en belangrijkere capaciteiten zal aanbieden. Wanneer de norm van de DVD-R-schijven (write once) industrieel erkend zal zijn, zullen deze laatste de CD-WORM's vervangen.

Wanneer men het heeft over de bewaring van gegevens gedurende lange periodes, mag men zich echter niet laten misleiden en zich uitsluitend focussen op de gebruikte media: zelfs als een constructeur ons garandeert dat een optische schijf binnen vijftig jaar nog intact zal zijn, zullen we dan het leesapparaat hebben die haar kan lezen en zal het formaat dat gebruikt werd voor de bestanden nog ondersteund worden door de software van die tijd?

4.2 De indexering

Een goede indexering van het document wanneer het ingevoerd wordt in een systeem voor documentair beheer is een conditio sine qua non om te garanderen dat het later gemakkelijk toegankelijk zal zijn.

We zullen hier niet ingaan op de "full text"-indexeringstools waarbij een index opgebouwd wordt op basis van alle woorden van een document die niet behoren tot een woordenboek van uit te sluiten woorden (zoals "de", "het", "of"... die geen enkel belang hebben voor de opzoeking). Deze indexeringstools en de overeenkomstige opzoektools zouden aan bod kunnen komen in een afzonderlijke publicatie.

De indexering waarover wij hier praten betreft de "meta-gegevens" van een document die gestructureerd kunnen worden in drie delen, waarbij de eerste twee informatie leveren over het document zelf en het derde de inhoud beschrijft van het document:

- ◇ Een eerste geheel van indexen betreft de informatie m.b.t. het document als dusdanig: de auteur, de eigenaar, de creatiedatum, de toepassing die men moet gebruiken om het te visualiseren,...
- ◇ Een tweede reeks indexen beschrijft de levenscyclus van het document: de data van achtereenvolgende bijwerkingen, de auteurs van deze bijwerkingen, de versie nummers, de data van on-line-archivering, de voorziene verwijderingsdatum,...
- ◇ Het derde geheel van indexen betreft de beschrijving van de inhoud van het document. Om deze inhoud correct te beschrijven, moet men deze taak niet alleen toevertrouwen aan iemand die deze inhoud perfect kent maar men moet hem ook indexeren met gestandaardiseerde termen, met vooraf gedefinieerde sleutelwoorden. Men zal deze sleutelwoorden definiëren ofwel in eenvoudige lijsten, ofwel in complexere structuren van thesauri, van classificatie die men kan beschrijven zoals woordenboeken van genormaliseerde termen, die conceptueel georganiseerd zijn en onderling verbonden worden door semantische relaties.



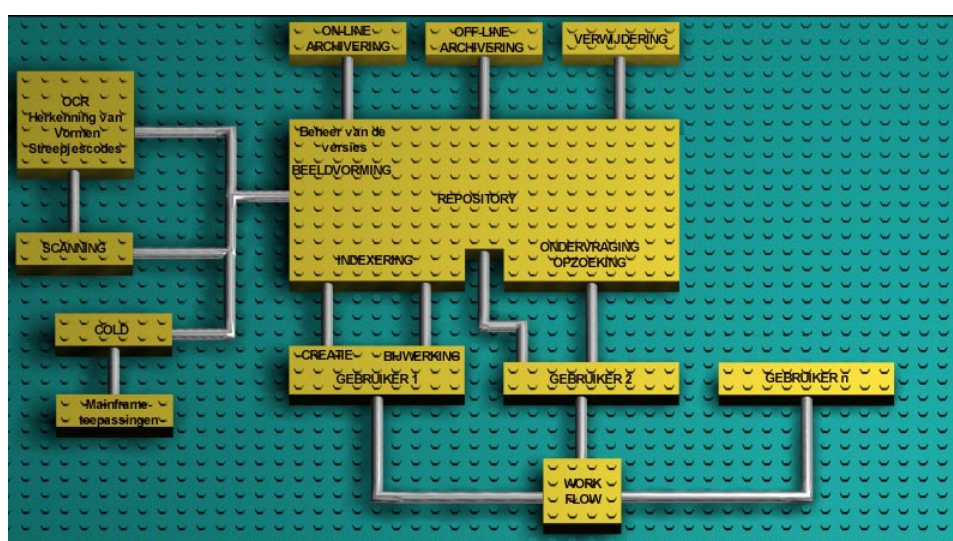
4.3 De toegangscontroles

Het is belangrijk de toegangen tot de informatie in een systeem voor documentair beheer te controleren. Elk document zal dus geassocieerd worden met de veiligheidsniveaus die toelaten de gebruikers te definiëren die het recht hebben om het document te visualiseren, te wijzigen, af te drukken, delen te knippen/plakken, te vernietigen. In meer geëvolueerde systemen zullen deze toegangsbeperkingen zelfs gedefinieerd kunnen worden voor delen van het document.

Naast deze toegangsbeveiligingen zal een systeem voor documentair beheer meestal audittools aanbieden die toelaten een spoor te behouden van alle acties die uitgevoerd werden in het systeem: wie deed wat, wanneer en hoe.

5. De manipulatie van de documenten

Om snel toegang te krijgen tot de gewenste informatie, is het noodzakelijk te beschikken over krachtige en voldoende uitgewerkte zoektools om deze informatie zo goed mogelijk af te bakenen. Bij het begin van zijn levenscyclus wordt een document vaak gewijzigd, hetzij door de auteur, hetzij door een geheel van mensen die samenwerken: tools voor het delen en laten doorstromen van de documenten zullen deze bewerkingen mogelijk maken. De modules voor het opzoeken, voor het beheren van de versies en voor workflow beantwoorden aan deze behoeften.



5.1 De opzoeking

De zoektool die ter beschikking van de gebruiker wordt gesteld, moet hem toelaten de documenten die hij nodig heeft gemakkelijk terug te vinden.

Een eenvoudige "full text"-opzoeking (bijvoorbeeld: alle documenten terugvinden die het woord "sociaal" bevatten) is vaak niet erg nuttig want het antwoord kan zoveel documenten bevatten dat dit resultaat niet bruikbaar zal zijn (dit is het geval met de talrijke zoektools op Internet die u duizenden documenten bezorgen als antwoord op dit soort vragen). Het is daarom dat er krachtigere zoektools ontwikkeld werden: wij zullen deze software niet in detail behandelen.



In meer klassieke zin zullen de opzoekingen in een documentaire gegevensbank gebeuren op de hierboven besproken "meta-gegevens". De mogelijkheden die deze zoektools bieden, gaan van de meest eenvoudige tot de meest complexe:

- ◇ Opzoeking op een sleutelwoord dat behoort tot de lijst van toegelaten sleutelwoorden;
- ◇ Booleaanse bewerkingen (en, of, en niet,...) op deze sleutelwoorden;
- ◇ Wanneer men een gehiërarchiseerde thesaurus gebruikt, zal men, bijvoorbeeld, alle documenten kunnen terugvinden m.b.t. een niveau van de hiërarchie door te zoeken op een woord van een lager niveau: een opzoeking op "arbeider" zal alle documenten opleveren die betrekking hebben op de "werknemers";
- ◇ Sommige zoektools zullen meertalige faciliteiten kunnen aanbieden: een opzoeking op "arbeid" zal ook de documenten vinden die geïndexeerd werden met "travail".

Deze zoektools worden meestal geassocieerd met tools die toelaten de documenten te manipuleren (rotatie van het beeld, zoom,...), ze af te drukken, er delen van te selecteren alsook ze te annoteren met een soort "post-its".

Benadrukken we ten slotte dat het ons, in een omgeving van documentair beheer, fundamenteel lijkt dat één en dezelfde zoektool de opzoekingen kan uitvoeren op alle documenten die in het systeem ingevoerd zijn ongeacht hun fysieke herkomst (bureautica, COLD of scanning).

5.2 De bijwerking en de verdeling van de documenten

Zoals wij hoger reeds stelden, worden de documenten van elektronische herkomst (tekstverwerkers, rekenbladen,...) meestal ingevoerd in het systeem voor documentair beheer alvorens zij volledig afgewerkt zijn.

De gebruikers die de gepaste toelatingsen hebben, moeten dus beschikken over tools die het mogelijk maken deze documenten bij te werken in alle veiligheid. Men moet immers vermijden dat meerdere gebruikers tegelijkertijd hetzelfde document wijzigen met het risico dat men bij het opslaan door de laatste gebruiker de wijzigingen van de andere gebruikers verliest. Een "check-in/check-out"-mechanisme laat toe dit soort problemen op te lossen: wanneer een gebruiker een document wil wijzigen, dan doet hij een "check-out" (hij vergrendelt het), hierdoor kunnen andere gebruikers het document nog wel lezen maar niet meer wijzigen; wanneer de gebruiker klaar is met zijn wijzigingen, dan doet hij een "check-in" (hij ontgrendelt het) zodat het gewijzigd kan worden door iemand anders.

Bij wijzigingen is het vaak nuttig nieuwe versies te creëren bij elke wijziging om de evolutie ervan terug te vinden, vooral als het document collectief wordt uitgewerkt. De tools voor versiebeheer laten toe deze problematiek te sturen. In de meeste gevallen worden de wijzigingen tussen twee opeenvolgende versies duidelijk geïdentificeerd, hetzij om de andere gebruikers hierover in te lichten, hetzij in afwachting dat de persoon die instaat voor de acceptatie van het document deze wijzigingen goedkeurt.

Bepaalde tools voor versiebeheer bieden tevens de mogelijkheid om "takken van versies" te beheren om daarna het document te onderscheiden in functie van zijn bestemming: bijvoorbeeld, de gebruiksaanwijzing van een automodel zal opgenomen worden in één enkel document terwijl delen van dit document in verschillende versies zullen bestaan in functie van de accessoires die toegevoegd worden aan het basismodel.



5.3 De workflow

Het ontwerpen van een document door een groep van gebruikers, zoals hierboven beschreven, komt overeen met een aanpak die men zou kunnen typeren als "pull" (trekken): wanneer een gebruiker een document wil bewerken, moet hij het zelf opzoeken in de documentaire database en het zich tijdelijk toe-eigenen om er zijn wijzigingen in aan te brengen. Deze aanpak biedt het voordeel dat de verschillende auteurs vrij zijn het initiatief te nemen, waardoor hun creativiteit bevorderd wordt.

Er zijn echter situaties waarin regels een goed gedefinieerde procedure opleggen om een document uit te werken: men kan zich inbeelden dat bepaalde delen opgesteld moeten worden door een aangeduide persoon, dat de redactie van de verschillende delen moet gebeuren in een bepaalde volgorde, dat er goedkeuringen van de hiërarchie nodig zijn in bepaalde stadia in de uitwerking van het document. In dit soort situatie zal men een aanpak aannemen van het type "push" (duwen): wanneer een actie uitgevoerd zal moeten worden door een bepaalde persoon, dan zal het document hem toegestuurd worden met de gepaste instructies.

Met de workflowtools kan men dit type procedures automatiseren. Voor een bepaald document (of eerder voor een bepaald documenttype) definieert men de te vervullen taken, de tussenkommende personen en hun verantwoordelijkheden, de te respecteren sequenties en, zodra het proces opgestart is, volgt de realisatie van het document automatisch haar verloop.

Naargelang de graad van hun complexiteit zullen de workflowtools meer of minder uitgewerkte routeringen mogelijk maken:

- ◇ lineaire routeringen waarbij het document bij elke etappe verstuurd wordt naar één enkele duidelijk gepreciseerde persoon;
- ◇ routeringen met vertakking waarbij het document, na een bepaalde etappe, verstuurd wordt naar meerdere personen tegelijk;
- ◇ routeringen waarbij de persoon naar wie het document gestuurd wordt, gekozen is in functie van zijn werkbelasting;
- ◇ conditionele routeringen waarbij de volgende tussenkommende persoon gekozen wordt in functie van een criterium dat samenhangt met de toestand van het document;
- ◇ routeringen met een alternatief om bijvoorbeeld rekening te houden met de afwezigheid van een medewerker (om het document niet te blokkeren in een etappe van het circuit);
- ◇ ...

De invoering van een workflow in de onderneming is ongetwijfeld het meest delicate deel in de ontplooiing van een oplossing voor documentair beheer. Men zal dus vermijden een workflow in te voeren daar waar de procedures niet goed vastgelegd zijn.

De voordelen die men er van kan verwachten zijn echter talrijk:

- ◇ rationalisering van de procedures van de onderneming;
- ◇ vermindering van hun kosten dankzij de automatisering ervan;
- ◇ tijdwinst door de elektronische transmissie van de delen nodig voor hun uitvoering;
- ◇ verbetering van de controle en van de opvolging van deze procedures.

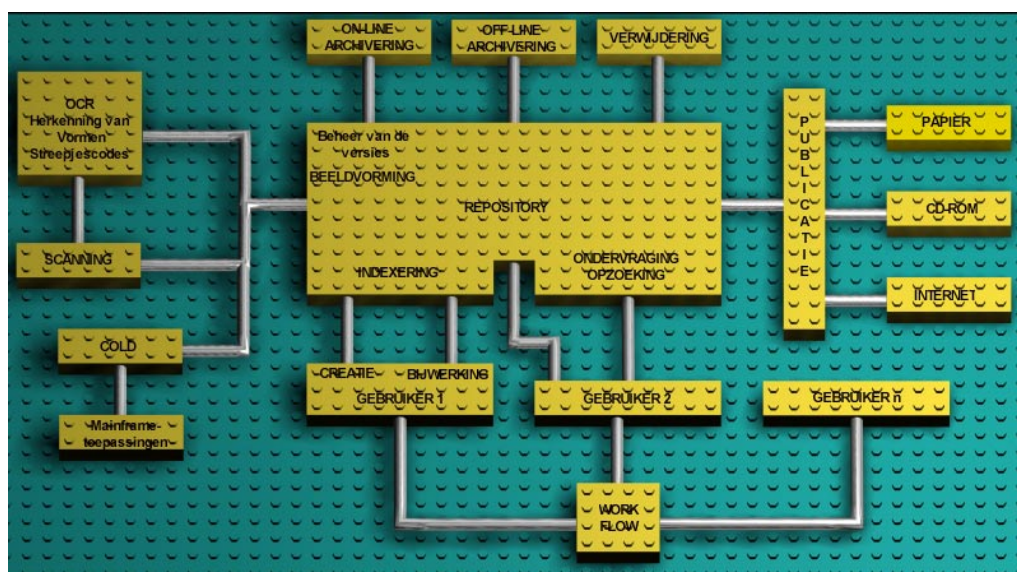
De analyse voorafgaand aan de invoering van een workflow toont vaak gebreken aan in een bestaande organisatie en de verleiding is groot om van deze gelegenheid gebruik te maken om deze organisatie te re-engineeren. Wij denken dat men het gewicht van een dergelijke beslissing ernstig moet evalueren en dat men deze richting slechts mag inslaan als er op het



terrein een wil bestaat en als de betrokken leiding zich werkelijk engageert in het project. Psychologisch is het soms rendabeler en realistischer om een eerste fase te doorlopen waarin de workflow de bestaande procedures overneemt, ook al blijven zij gedeeltelijk in gebreke.

6. De verspreiding van de documenten

We komen aan het einde van de opbouw van ons model voor documentair beheer en voegen het laatste onderdeel toe: de verspreiding of de publicatie van de documenten.



Zoals bovenstaand schema aangeeft, zijn er drie hoofdwegen om documenten te verspreiden: op papier, op cd-rom en op Internet of Intranet.

In vele situaties zal men ertoe aangezet worden dezelfde documenten onder verschillende vormen te publiceren. Het is dan ook wenselijk dat men over deze documenten beschikt in een elektronisch formaat zodat men ze eenvoudig kan verspreiden op deze verschillende manieren zonder aan kwaliteit in te boeten op het vlak van de presentatie.

In de praktijk bestaan er vandaag talrijke conversietools die toelaten een tekst die gemaakt werd in een standaardtekstverwerker te "vertalen" in een HTML-pagina om hem te verspreiden op Internet. Deze vertalingen zijn echter niet altijd perfect wat betreft de vorm, de kwaliteit van het aangemaakte document en men mag hopen dat de toekomstige ontwikkelingen van XML de realisatie van kwalitatief betere vertalingen zullen toelaten.

Steeds vaker wordt ook het formaat PDF gebruikt om te publiceren op Internet: men stelt deze evolutie vast in bedrijfsites waar de commerciële brochures, de activiteitenverslagen,... steeds vaker gepubliceerd worden onder het formaat PDF; er zijn ook veel uitgeverij die beginnen met het aanbieden van abonnementen via Internet door dit formaat te gebruiken voor hun publicaties. Naast de hogergenoemde troeven, biedt dit formaat ook het voordeel dat het afgedrukte document getrouw wordt weergegeven, dat het een hoog compressiepercentage heeft, (en dus redelijke bestandsgroottes heeft voor de verspreiding op Internet), en dat het, sinds de laatste versie, ook in staat is tot "streaming". Dit wil zeggen dat een gebruiker niet moet wachten tot wanneer een bestand volledig opgeladen is alvorens hij de eerste pagina's ervan kan lezen. Ten slotte laat de tool toe deze bestanden te lezen ofwel onmiddellijk



("reader") ofwel binnen een Internetnavigators ("plug-in") en is hij helemaal gratis, wat ongetwijfeld een belangrijke reden is voor het succes ervan.

De publicatie op cd-rom kan heel eenvoudig uitgevoerd worden. Ofwel kopieert men op de cd-rom de HTML-pagina's die reeds voor Internet gerealiseerd werden, ofwel graveert men op de cd-rom PDF-bestanden (met eventueel de geassocieerde reader).

Men neemt echter zelden vrede met deze elementaire oplossingen wanneer men een cd-rom aanmaakt: deze laatste biedt namelijk veel rijkere mogelijkheden zoals de invoeging van geluids- en videoclips en de toegangssnelheid is geen rem op het gebruik van deze technieken. Om dergelijke cd-roms aan te maken, gebruikt men edittools ("authoring") die hiervoor ontwikkeld werden.

7. Conclusie

We hebben de verschillende tools overlopen die, onderling geassembleerd, toelaten het geheel van de documenten van een onderneming te beheren.

Veel ondernemingen beginnen zich bewust te worden van het belang van hun niet-gestructureerde informatie die bestaat onder de vorm van documenten in de brede zin zoals wij ze hierboven gedefinieerd hebben. De groei van de markt van de tools voor het elektronisch beheer van documenten toont deze tendens zeer duidelijk: in 1998 werd de wereldmarkt geraamd op 14 miljard dollar en de studies voorzien een groei van om en bij de 20% per jaar in de vier volgende jaren, om te komen tot 33 miljard dollar in 2002.

We vinden deze evolutie terug in alle sectoren van de industrie: de banken en de verzekeringsmaatschappijen waar het EDM vooral gebruikt wordt voor de archivering en het verbeteren van de dienstverlening aan de klant; de fabrieksindustrie zowel voor het beheren van de fabricageprocedures als voor het aanmaken van handleidingen en gebruiksaanwijzingen; de farmaceutische industrie waar het bijna onmogelijk wordt om op papier de documentatie te beheren die voorgelegd moet worden aan de autoriteiten om toelating te krijgen om een medicijn op de markt te brengen; de openbare sector die niet alleen geconfronteerd wordt met grote massa's formulieren die beheerd moeten worden maar ook met wettelijke verplichtingen voor het bewaren van de documenten.

Zelfs al worden de huidige tools steeds gebruiksvriendelijker, toch mag men hieruit niet te snel besluiten dat een elektronisch beheer van documenten gemakkelijk in te voeren is: het is belangrijk deze oplossingen geleidelijk te ontplooien zodat de gebruikers geleidelijk kunnen wennen aan deze nieuwe werkwijzen en hen te begeleiden bij de verandering. Men dient onder meer te vermijden dat de bestaande procedures te veel gewijzigd worden bij de invoering van een workflow en men moet er vrede mee nemen deze procedures in een eerste fase over te nemen en ze pas daarna geleidelijk te optimaliseren.

De SmalS-MvM heeft, voor de RSZ, reeds een project voor documentair beheer (Juridoc) uitgewerkt om de exploitatie van de juridische documentatie te verbeteren. Binnenkort zal zij een oplossing voor documentair beheer uitwerken om, onder elektronische vorm, het geheel van de documentatie van de RSZ m.b.t. de werkgevers te beheren.

Wij denken dat het interessant is te onderzoeken of het opportuun zou zijn oplossingen van het type COLD in te voeren om de productie van papier door de mainframes te doen dalen en de gebruikers van deze listings moderne en gebruiksvriendelijke tools aan te bieden om de informatie te exploiteren.



De opkomst van deze nieuwe tools voor het beheer van documenten heeft nog een enorm groeipotentieel en we mogen denken dat we in een min of meer nabije toekomst echt zullen kunnen praten over het "minimal paper office", een kantoor zonder overbodig papier.

Dankzij deze nieuwe technologieën zal men in de ondernemingen een belangrijk deel kunnen terugvinden van de kennis die er door de jaren heen beetje bij beetje verspreid raakte. Maar, wat ons betreft, wij hopen dat het boek zijn reden van bestaan zal behouden en dat volgende generaties nog altijd zullen kunnen genieten van het lezen van boek dat vers van de pers is gerold, met de gevoelens die opgewekt worden door het ruiken en voelen van de versgedrukte pagina's.

