

# *DAVID*

---

Archiveren van websites

Filip Boudrez  
Sofie Van den Eynde

Versie 1.0

Wettelijk depot D/2001/9.213/8

Antwerpen-Leuven, mei 2002

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

E-mailadres: [david@stad.antwerpen.be](mailto:david@stad.antwerpen.be)

# INHOUDSTAFEL

|                                                                   |           |
|-------------------------------------------------------------------|-----------|
| <b>INHOUDSTAFEL .....</b>                                         | <b>3</b>  |
| <b>I. INLEIDING .....</b>                                         | <b>6</b>  |
| <b>II. INTERNET, WWW &amp; WEBSITES .....</b>                     | <b>8</b>  |
| <b>III. BELANG VAN WEBSITES ARCHIVERING .....</b>                 | <b>13</b> |
| <b>IV. BESTAANDE INITIATIEVEN .....</b>                           | <b>17</b> |
| <b>V. ARCHIVEREN VAN WEBSITES .....</b>                           | <b>20</b> |
| A. KWALITEITSEISEN VOOR GEARCHIVEERDE WEBSITES .....              | 21        |
| B. SELECTIE: WAT ARCHIVEREN? .....                                | 22        |
| B.1 Het acquisitiebeleid: welke websites archiveren ? .....       | 22        |
| B.2 Wat archiveren van websites? .....                            | 24        |
| B.2.1 Websites met een vaste inhoud .....                         | 25        |
| B.2.2 Websites met een dynamische inhoud .....                    | 27        |
| B.2.3 Besluit .....                                               | 31        |
| C. HOE WEBSITES VASTLEGGEN VOOR ARCHIVERING ? .....               | 31        |
| C.1 Websites met een vaste inhoud .....                           | 31        |
| C.2 Websites met een dynamische inhoud .....                      | 33        |
| C.2.1 Snapshots .....                                             | 33        |
| C.2.2 Logbestanden .....                                          | 37        |
| C.2.3 Databanken .....                                            | 38        |
| D. FREQUENTIE .....                                               | 38        |
| D.1. Bepalen van de frequentie .....                              | 38        |
| D.2. Hoe versies archiveren? .....                                | 39        |
| D.2.1 De wijzigingen aan websites met een vaste inhoud .....      | 39        |
| D.2.2 De wijzigingen aan websites met een dynamische inhoud ..... | 40        |
| E. DIGITALE DUURZAAMHEID .....                                    | 41        |
| F. BEHEER VAN DE GEARCHIVEERDE WEBSITES .....                     | 45        |
| F.1 Beschrijven .....                                             | 45        |
| F.2 Veilige bewaring .....                                        | 50        |
| F.3 Opslag op informatiedragers .....                             | 51        |
| H. TER BESCHIKKING STELLEN .....                                  | 51        |
| <b>VI. UITBOUWEN VAN HET ARCHIVERINGSSYSTEEM .....</b>            | <b>52</b> |
| A. ROL VAN DE ARCHIVARIS .....                                    | 52        |
| B. EFFICIËNT ONTWERP EN BEHEER .....                              | 54        |
| C. ARCHIVERING .....                                              | 57        |

|                                                                                                                           |           |
|---------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>VII. HET AUTEURSRECHT: EEN OBSTAKEL VOOR DE ARCHIVERING VAN WEBSITES .....</b>                                         | <b>60</b> |
| A. INLEIDING .....                                                                                                        | 60        |
| B. AUTEURSRECHT IN EEN NOTENDOP .....                                                                                     | 61        |
| B.1. Beschermde werken .....                                                                                              | 61        |
| a) Werken van letterkunde of kunst .....                                                                                  | 61        |
| b) En websites? .....                                                                                                     | 62        |
| c) Hyperlinks .....                                                                                                       | 63        |
| d) Uitzondering voor officiële akten van de overheid .....                                                                | 63        |
| B.2. Wie is drager van de auteursrechten? .....                                                                           | 64        |
| B.3. Wat houden de auteursrechten in voor de auteursrechthebbende? .....                                                  | 65        |
| a) Vermogensrechten .....                                                                                                 | 65        |
| b) Morele rechten .....                                                                                                   | 67        |
| B.4. Hoe verkrijgt men auteursrechtelijke bescherming? .....                                                              | 68        |
| C. ARCHIVERING VAN WEBSITES: PROBLEEMSTELLING .....                                                                       | 68        |
| C.1. Reproductiehandelingen .....                                                                                         | 69        |
| C.2. Het aanbrengen van wijzigingen .....                                                                                 | 72        |
| C.3. Digitale duurzaamheid .....                                                                                          | 73        |
| C.4. Ter beschikking stelling .....                                                                                       | 74        |
| C.5. Bijzondere regimes .....                                                                                             | 74        |
| a) Computerprogramma's .....                                                                                              | 74        |
| b) Databanken .....                                                                                                       | 76        |
| C.6. Besluit .....                                                                                                        | 78        |
| D. OPLOSSING: UITZONDERING VOOR PRESERVERINGSDOELEINDEN .....                                                             | 79        |
| <b>VIII. ARCHIVEREN VAN PERSOONSGEGEVENS .....</b>                                                                        | <b>82</b> |
| A. PERSOONSGEGEVENS EN INTERNET .....                                                                                     | 82        |
| A.1. Organisatie en beheer van het Internet .....                                                                         | 83        |
| A.2. IP-adressen .....                                                                                                    | 85        |
| A.3. Logbestanden .....                                                                                                   | 87        |
| B. VERWERKING VAN PERSOONSGEGEVENS VOOR HISTORISCHE DOELEINDEN: IN<br>OVEREENSTEMMING MET HET OORSPRONKELIJKE DOEL? ..... | 88        |
| <b>IX. DE VLAAMSE OVERHEDEN OP HET INTERNET .....</b>                                                                     | <b>90</b> |
| A. DE SITUATIE ANNO 2002 .....                                                                                            | 90        |
| B. CASE: HET ELEKTRONISCH LOKET BEVOLKING VAN CEVI .....                                                                  | 91        |
| <b>X. AANSPRAKELIJKHEID VAN DE OVERHEID VOOR HAAR EIGEN WEBSTEK ..</b>                                                    | <b>92</b> |
| A. KAN DE BURGER RECHTEN ONTLENEN AAN DE INHOUD VAN OVERHEIDSWEBITES? .....                                               | 92        |
| B. ZIN OF ONZIN VAN DISCLAIMERS .....                                                                                     | 94        |
| <b>XI. PORTAALSITES: DE TOEKOMST .....</b>                                                                                | <b>95</b> |
| A. VAN INFORMATIE NAAR INTERACTIE EN INTEGRATIE .....                                                                     | 95        |
| B. ÉÉN VIRTUELE OVERHEID .....                                                                                            | 96        |
| C. HET WETTELIJKE KADER .....                                                                                             | 96        |
| C.1. Front office: de elektronische identiteitskaart .....                                                                | 96        |

|                                                  |            |
|--------------------------------------------------|------------|
| C.2. Back office: uniek identificatienummer..... | 98         |
| <b>XII. BESLUIT &amp; AANBEVELING.....</b>       | <b>99</b>  |
| <b>XIII. BIBLIOGRAFIE .....</b>                  | <b>101</b> |

# I. INLEIDING

Websites zijn al lang geen onbekende meer voor archivariissen. Steeds meer archiefdiensten -en instellingen maken van deze internettechnologie gebruik om informatie en diensten on line ter beschikking te stellen van hun klanten. In een aantal gevallen is de interactieve dienstverlening via het net sterk uitgebouwd en kan de vorser zelfs al digitale archiefdocumenten on line raadplegen. Op die manier wordt de digitale leeszaal een stukje tot bij de mensen thuis gebracht. Daarnaast is er nog een twee manier waarop archiefdiensten en -instellingen met deze nieuwe technologie worden geconfronteerd, nl. het archiveren van websites. Dit rapport spitst zich toe op de vraag hoe websites het best worden gearchiveerd.

Algemeen is men het er over eens dat via internet een schat van informatie beschikbaar is. Internet heeft zijn succes aan diverse factoren te danken. Eén van de voornaamste is de snelheid waarmee informatie wereldwijd beschikbaar is en wordt aangepast. Precies de vluchtigheid van de informatie heeft er toe geleid dat er werd gezocht naar een manier om die informatie vast te leggen. De eerste Zweedse elektronische nieuwsbrief is bijvoorbeeld verloren gegaan<sup>1</sup>. De officiële website van de Olympische Spelen van 2000 staat niet meer on line. Al vroeg zag men de noodzaak in om de informatie op het internet te archiveren. Zo stelt men vast dat aan websites een module ‘archief’ wordt toegevoegd waar de vorige versies nog steeds on line beschikbaar zijn of dat verouderde webpagina’s gewoon actief blijven en van een banner ‘gearchiveerd’ worden voorzien<sup>2</sup>. Dit zijn echter losse initiatieven van webdesigners of content managers die om één of andere reden oudere versies van webpagina’s verder raadpleegbaar willen houden.

Dit rapport belicht de mogelijkheden voor het archiveren van websites en de aandachtspunten waarmee men rekening moet houden. In het eerste hoofdstuk wordt het Internet, het World Wide Web, de architectuur en de evolutie van websites toegelicht. Wie voldoende vertrouwd is met de client-serverinteracties en de architectuur van websites kan dit overslaan. Na het belang van het archiveren van websites komt vervolgens de digitale archiveringsstrategie aan bod. Aangezien niet alleen archiefdiensten en -instellingen maar ook bibliotheken en documentatiecentra op dit terrein actief zijn, wordt dit zo ruim en zo algemeen mogelijk benaderd. Bij het uitstippelen van een archiveringsbeleid voor websites komen opnieuw meerdere vragen aan bod. De belangrijkste zijn hier: Wat archiveren? Hoe verwerven? Met welke frequentie? Hoe websites beheren? Hoe het website archief ter beschikking stellen? Bij elke vraag zijn meerdere scenario’s mogelijk. Iedere instelling heeft zijn doelstelling en volgt de optie die er het best bij aansluit. Dit resulteert uiteindelijk in een ander archiveringsbeleid. Het deel met betrekking tot het archiveren wordt afgesloten met een praktijkgericht archiveringsvoorbeeld. Zoals steeds wordt hierbij uitgegaan van een minimale IT-infrastructuur.

Bij het uittekenen van een archiveringsstrategie voor websites zal de organisatie ook rekening moeten houden met de juridische implicaties van het auteursrecht. Voor heel wat

---

<sup>1</sup> K. PERSSON, *The Kulturarw3 Project - The Swedish Royal Web Archiw*<sup>3</sup>, Lezing gehouden in Svetlogorsk, aug. 2000.

<sup>2</sup> Bijv. <http://europa.eu.int/ISPO/dlm/documents/guidelines.html> met de boodschap: “This website has been archived. Please visit the new site ...”.

reproductiehandelingen zal hij de toestemming moeten vragen van de auteur van de (inhoud van de) website. Na een korte inleiding in het Belgische auteursrecht, brengen we in kaart met welke regels de archivaris op dit vlak rekening moet houden. Ook de privacyreglementering schept voor archiefinstellingen en -diensten heel wat problemen. Voor deze juridische problemen trachten we in dit rapport oplossingen te formuleren.

Tenslotte gaan we na hoe het gesteld is met de aanwezigheid van de Vlaamse overheden op het Internet en bekijken we de plannen van de federale overheid met betrekking tot de ontwikkeling van elektronische identiteitskaart voor iedere burger. De introductie van deze kaart in de loop van 2003 zal de ontwikkeling van e-government ongetwijfeld ten goede komen. Het probleem van het archiveren van overheidswebsites zal zich vanaf dan nog meer opdringen. Met dit rapport willen we hierop anticiperen.

Sofie Van den Eynde schreef het juridisch gedeelte van dit rapport. Filip Boudrez nam de digitale archivering voor zijn rekening.

Antwerpen-Leuven, mei 2002.

## II. INTERNET, WWW & WEBSITES

Het Internet kan misschien het gemakkelijkst worden omschreven als een wereldomspannend computernetwerk<sup>3</sup>. De ontwikkeling van het Internet gaat terug tot de jaren zestig van de vorige eeuw. Tijdens de warme fase van de Koude Oorlog zocht het Amerikaanse Ministerie van Defensie naar een manier om computers met elkaar te verbinden zodat informatiebeheer en raketbesturing vanop verschillende plaatsen mogelijk werd. Het eerste netwerk kreeg de naam Arpanet mee. Het uitbouwen van de voorloper van het huidige Internet ging samen met het vastleggen van een protocol voor het adresseren en versturen van informatie. TCP/IP (Transmission Control Protocol / Internet Protocol) werd het standaardprotocol en bevat eigenlijk tientallen protocollen waarvan TCP en IP de meest gebruikte zijn. Elke computer die in het netwerk werd opgenomen, kreeg een uniek IP-adres waar later een domeinnaam werd aan gekoppeld (DNS: Domain Name System)<sup>4</sup>. Het Internet is ondertussen sterk uitgebouwd en kent verschillende toepassingen: gopher<sup>5</sup>, FTP-archieven, e-mail, usenet, nieuwsgroepen, Telnet en het World Wide Web (WWW).

De ontwikkeling van het WWW is een mijlpaal in de geschiedenis van het Internet (CERN: 1989-'92). Het WWW is inmiddels geëvolueerd van een technologie voor tekstitwisseling tot een interactieve en dynamische client-servertoepassing waarbij documenten met hypertext, multimediatoepassingen en databanken aan elkaar zijn verbonden. Het accent is verschoven van pure tekst naar grafische vormgeving en muisbesturing. Hypertext is tekst die hyperlinks naar andere documenten bevat. Voor het ophalen van WWW-documenten werd een nieuw protocol ontwikkeld: HTTP (Hypertext Transfer Protocol). HTTP ondersteunt de communicatie van hypertextbestanden tussen server en client. Voor de uitwisseling van bestanden tussen client en server wordt FTP (File Transfer Protocol) gebruikt.

Voor het samenstellen van WWW-documenten werd eveneens een standaard ontwikkeld: HyperText Mark-up Language (HTML). HTML is een vastgelegde taal die het mogelijk maakt om aan de hand van mark-upopdrachten de onderdelen van een webpagina en hun functie af te bakenen. De opdrachten staan tussen < en > en worden HTML-tags genoemd.

---

<sup>3</sup> <http://www.isoc.org/internet/history/brief.html>; <http://www.davesite.com/webstation/net-history.shtml>; J. HONEYCUTT (e.a.), *Het complete handboek Internet*, Schoonhoven, 1997; C.J.M. MOSCHOVITIS, *History of the Internet: a chronology, 1843 to the Present*, Santa-Barbara (California), 1999; I. ENGHOLM, *Digital design history and the registration of web development*,

<sup>4</sup> Bij het opgeven van een URL (Universal Resource Locator) in de adresbalk van een browser zet DNS de domeinnaam om in een IP-adres.

<sup>5</sup> gopher: een internettoepassing voor de uitwisseling van informatie die populair was begin de jaren 1990. Gopher werkt niet met MIME-types zoals HTTP maar met Gopher object types. Geleidelijk aan moest gopher de plaats ruimen voor HTTP.

Voorbeeld 1: HyperText Markup LanguageBRONCODE

```
<p><i>Deze zin wordt cursief weergegeven</i></p>
```

```

```

```
<a href="http://www.antwerpen.be/david">Hyperlink  
naar DAVID website</a>
```

```
<% ASP, PHP of JSP-code %>
```

WEBBROWSER

*Deze zin wordt cursief weergegeven*



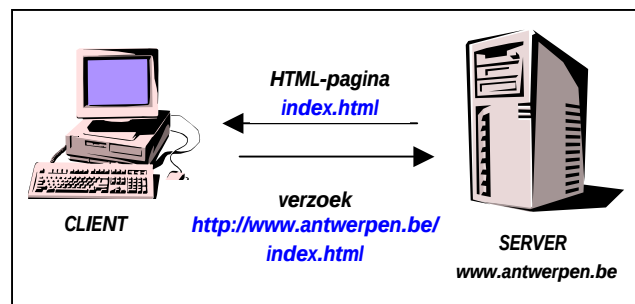
[Hyperlink naar DAVID website](http://www.antwerpen.be/david)

HTML-output

De clientcomputer is de computer waarop een website wordt bekeken. Het programma op de client dat de WWW-documenten genereert is de webbrowser. De webbrowser stelt het HTTP-verzoek samen en stuurt dit door naar de server. De recentste versies zijn grafische browsers die ook andere internettoepassingen zoals FTP en gopher ondersteunen en die ook automatisch andere applicaties (bijv. MS Office, Acrobat Reader, MP3-speler, RealPlayer, Shockwave, Flash Player, enz.) opstarten om bestanden in een ander formaat dan HTML te openen. Sommige van deze applicaties zijn plug-ins in de webbrowser. De computer die de website host en via het web ter beschikking stelt, wordt de server genoemd. Op deze computer draaien webserverprogramma's (Apache, Internet Information Server, Netscape FastTrack / Enterprise Server, enz.), modules voor scripts en uitvoerbare programma's.

Een website bestaat uit een geheel van afzonderlijke computerbestanden die in een bepaalde mappenstructuur is opgeslagen. De band tussen twee computerbestanden van een website wordt vastgelegd door middel van hyperlinks. Via deze links krijgt men toegang tot het online materiaal. Webpagina's en de bijhorende bestanden zoals afbeeldingen en downloads kunnen op twee manieren aan elkaar worden gelinkt: absoluut en relatief. De absolute links verwijzen naar en vertrekken van de root van de website. De rootaanduiding start doorgaans van de URL of het IP-adres: `http://www.antwerpen.be/david/nl/index.htm`. Bij een relatieve link vertrekt de pathaanduiding van de positie van waar gelinkt wordt: `../nl/index.htm`. In beide soorten pathaanduidingen wordt verwezen naar de mappenstructuur, de mapnamen en de bestandsnamen. Als één van deze onderdelen wijzigt, moet ook de pathaanduiding worden aangepast, anders werkt de link niet meer.

De oudste WWW-toepassingen waren heel statisch. De communicatie tussen server en client was éénrichtingsverkeer. Deze websites zijn niets meer dan een aantal HTML-bestanden en afbeeldingen die in een bepaalde bestandenstructuur op de webserver werden geplaatst met links als koppeling tussen de pagina's en afbeeldingen. De HTML-pagina's en eventueel bijhorende stylesheets bevatten de inhoud en de opmaakgegevens van de webpagina die naar de client wordt verstuurd. De interactie tussen server en client beperkt zich tot het versturen van een HTTP-verzoek en het terugsturen van de gevraagde webpagina. Met de

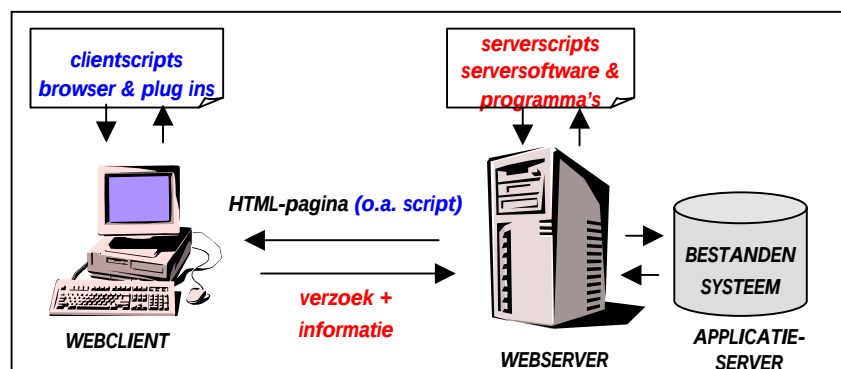


Afbeelding 1: De client-server interactie bij statische websites

instructie `http://www.antwerpen.be/index.html` wordt de server met als domeinnaam 'www.antwerpen.be' om de HTML-pagina `index.html` verzocht. Een deel van de domeinnaam is een alias die verwijst naar de rootmap van de website op de webserver. Op de webserver is een HTTP-daemon actief die wacht op de verzoeken van de webbrowsers, ze uitvoert en HTML-pagina's en bijhorende bestanden naar de client terugstuurt. De server voegt een HTTP-header aan de bestanden toe. Op de client genereert de browser vervolgens het WWW-document. De inhoud van deze webpagina's ligt vast en is voor elke bezoeker dezelfde. Tot op de dag van vandaag bestaan een groot aantal websites op het WWW uit webpagina's met een vaste inhoud. De interactie tussen client en server beperkt zich hoofdzakelijk tot het uitwisselen van verzoeken en HTML-pagina's. Deze websites worden in dit rapport 'statische websites' genoemd.

Het verstrekken van informatie op basis van statische documenten heeft zijn beperkingen en er werd bijgevolg al snel naar een mogelijkheid gezocht om een grotere interactie tussen server en client mogelijk te maken. Dit werd in eerste instantie mogelijk door het gebruik van CGI (Common Gateway Interface)-serverscripts. CGI-scripts zijn echter niet zo veilig en kunnen enkel op de server worden uitgevoerd. Een aantal producentgebonden scripttalen bieden hier oplossingen voor en zijn inmiddels wijdverspreid: ASP (Active Server Pages, Microsoft), PHP (Php Hypertext Processor, Unix-Linux), JAVA Servlets, ColdFusion (Macromedia) en JSP (JavaServer Pages, Sun). Serverscripts zijn ingebed in HTML-pagina's of worden in aparte bestanden opgenomen. Voor het uitvoeren van de serverscripts is de passende webserversoftware vereist. In veel gevallen werkt het serverscript samen een computerprogramma dat op de webserver draait (bijv. `form.exe`, `query.exe`, `rightsite.exe`). Bij het uitvoeren van een serverscript voert een webserver volgende handelingen uit: aanroepen script(programma), inlezen van de gevraagde bestanden, uitvoeren van het script, gegevens doorsturen of opvragen, uitvoeren van de bewerking en tenslotte het doorsturen van het resultaat als HTML naar de webbrowser.

Afbeelding 2: De client-server interactie bij websites met formulieren, waarlangs databanken worden bevraagd of die "on the fly" worden gegenereerd. Het gekoppelde bestandensysteem bestaat meestal uit een databank of documentbeheersysteem die op een applicatieserver lopen



Veel gebruikte toepassingen zijn het verwerken van verstuurd formuliergegevens, het bevroegen van databanken, het beschikbaarstellen van documenten via een documentbeheerssysteem. De website wordt hierbij als interface gebruikt en is niet langer een op zichzelf functionerende entiteit. De website maakt deel uit van een informatiesysteem dat opgebouwd is uit: webserversoftware, webserverconfiguratie, scriptbestanden, modules voor het uitvoeren van scripts, uitvoerbare programma's en databanken of documentsbeheerssystemen. De gekoppelde databanken staan meestal op afzonderlijke applicatieservers. Deze databanken worden het 'deep web' of het 'back-office systeem' genoemd. De gekoppelde websites zijn 'database-driven'. De inhoud van de webpagina's die de client ontvangt, is bij dergelijke toepassingen afhankelijk van de zoekopdracht of van de informatie die op dat ogenblik in de databanken aanwezig is. De webpagina's worden "on the fly" op de webserver samengesteld en vervolgens naar de webbrowser doorgestuurd. Een andere mogelijkheid is

dat het ‘back-office systeem’ op vastgelegde tijdstippen een HTML-pagina publiceert die op de website raadpleegbaar is. De inhoud van deze websites is dynamisch en is niet voor elke bezoeker dezelfde.

Om de server te ontlasten wordt het uitvoeren van de scripts zoveel mogelijk naar de clientzijde verschoven. Binnen een dergelijke toepassing bezorgt de server aan de client de nodige HTML-pagina met de bijhorende clientscripts, ingebed in een HTML-pagina of in een afzonderlijk bestand. De scripts worden vervolgens door de client uitgevoerd. Absolute voorwaarde is dat de clientcomputer over de nodige software beschikt. In de meeste gevallen zal het evenwel volstaan dat de vereiste browser(versie) is geïnstalleerd. Scripts die aan de clientzijde worden uitgevoerd zijn ondermeer JAVA, JavaScript en VBScript<sup>6</sup>. Voorbeelden van toepassingen met clientscripts zijn het controleren van gegevens in een formulier voor verzending, automatische doorlinking naar een andere webpagina, menubalken, roll-over afbeeldingen, knoppen met animatie, automatisch weergeven van datum en tijd, het toevoegen van URL aan favorieten, enz. De ontvangen HTML pagina bevat in al deze gevallen client-scripting.

Om websites meer functionaliteit te geven, worden applets en ActiveX-toepassingen toegevoegd. Applets zijn kleine Java-tools die niet functioneren bovenop een besturingssysteem, maar die door de Java Virtuele Machines binnen webbrowsers worden uitgevoerd. In tegenstelling tot ActiveX hebben applets geen toegang tot de harde schijf van de clientcomputer nodig. ActiveX-toepassingen breiden de webbrowserfunctionaliteit uit en worden op de harde schijf geïnstalleerd zodat ze na het bezoeken van een bepaalde website nog bruikbaar zijn.

De laatste tijd zijn ook websites in Flash sterk in opmars. Flash is een Macromediatoepassing waarbij animatie, tekst, afbeeldingen, interactie en geluid met elkaar worden gecombineerd. Websites in Flash bestaan uit één of meerdere ‘filmpjes’ (\*.fla-bestanden) die voor verspreiding via het web als \*.swf-bestanden worden gepubliceerd. Voor het bekijken van Flash-websites heeft men de passende plug-in nodig. De Flash-website wordt nog in veel gevallen in de vorm van een statische HTML-versie aangeboden.

Met de jongste generatie WWW-toepassingen is het mogelijk om gepersonaliseerde informatie ter beschikking te stellen. Deze webpagina’s kunnen dan ook niet louter meer als publicatie of gemeenschappelijke interface worden beschouwd. Wat op het scherm verschijnt kan afhankelijk zijn van de gebruikersrechten, het gebruikersprofiel (o.a. wie? waar vandaan? wanneer?), het tijdstip, vorige geraadpleegde webpagina’s, de toegepaste zoekopdracht en de software waarover de gebruiker beschikt<sup>7</sup>. Deze websites hebben dus geen vaste vorm en kunnen misschien wel het best met een computerprogramma worden vergeleken. De output van deze websites is wel een statische HTML-

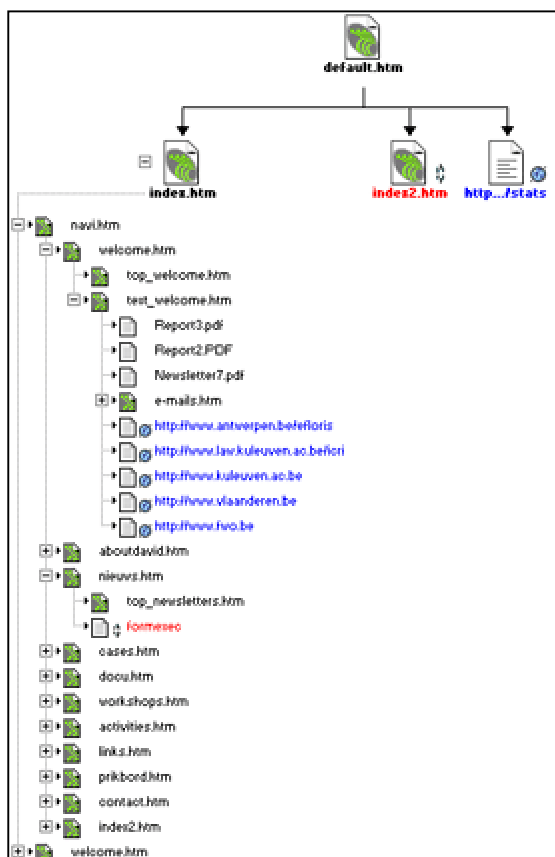
---

<sup>6</sup> Java is een programmeertaal van Sun Microsystems. Een gecompileerd JAVA-programma dat aan een website is toegevoegd wordt een *applet* genoemd. JAVA draait op een virtuele machine (VM). Applets worden samen met de webpage naar de client gestuurd. Een applet wordt niet op de server uitgevoerd. JavaScript is een andere manier om multimediabestanden naar webbrowsers te sturen. JavaScript is het resultaat van een samenwerking tussen Netscape en Sun Microsystems. De JavaScriptcode van het programma wordt rechtstreeks in een HTML-page ingesloten. JavaScriptcodes worden in tegenstelling tot JAVA-applets niet gecompileerd. VBScript is een combinatie van Visual Basic en Microsoft OLE-scripting. Met VBScript kunnen zowel aan de clientzijde als aan de serverzijde scripts worden uitgevoerd.

<sup>7</sup> Een voorbeeld van dit laatste is de homepagina van de website <http://www.antwerpen.be>. Bij de aanmelding wordt automatisch een browsercheck uitgevoerd. De bezoekers met een IE-browser krijgen het linkerframe met de uitklapbare navigatie als Dynamic HTML toegestuurd. Aangezien deze functionaliteit enkel werkt in IE-browsers is er van ditzelfde frame een meer statische versie beschikbaar voor bijvoorbeeld Netscape-gebruikers. Deze statische versie is opgesteld conform de W3C-standaard.

pagina, maar die wordt pas samengesteld op het ogenblik dat de webserver een HTTP-verzoek ontvangt. De webserver neemt kennis van het gebruikersprofiel en de voorkeur van de bezoeker op basis van een cookie of de informatie die een webclient altijd mee verstuurt<sup>8</sup>. Voorbeelden hiervan zijn websites die gekoppeld zijn aan documentbeheerssystemen waarin gedefinieerd is wie welke documenten kan openen of websites waarvan de inhoud afhankelijk is van de voorkeuren van de gebruiker bij eerdere bezoeken.

Bij het archiveren van websites gaat het zowel om websites met een statische als een dynamische inhoud. De huidige websites variëren van statische digitale publicaties tot interactieve dynamische websites die gebruikt worden voor het verlenen van diensten of voor het verrichten van transacties. Dynamische websites kunnen eveneens een aantal statische webpagina's bevatten. De DAVID-website is bijvoorbeeld een overwegend statische website, maar een aantal pagina's zijn dynamisch doordat ze gelinkt zijn aan het documentbeheerssysteem van het stadsarchief Antwerpen (bijv. de webpagina's publicaties en nieuwsbrieven). De link tussen de website en het documentbeheerssysteem is een koppeling op basis van een Rightsitescript die door de server wordt uitgevoerd. Voor het automatisch in- en uitschrijven op de maillijst voor de nieuwsbrief wordt een ASP-script gebruikt.



Afbeelding 3: Sitemap van de DAVID website (versie 6).

De startpagina `index.htm` bestaat uit een leftframe (`navi.htm`) en een rightframe (`welcome.htm`). Via het navigatiepaneel worden andere HTML-pagina's bereikt die in het rightframe worden weergegeven. Die achterliggende HTML-pagina's (`welcome.htm` tem `index2.htm`) bestaan elk uit twee frames. Het bovenste frame (`top_*.htm`) bevat de titelbalk. Het onderste frame (`text_*.htm`) bevat de tekst die op de website wordt weergegeven.

De sitemap is een belangrijk instrument dat de structuur en de links in een website toont. Veel websites bevatten een sitemap zodat informatie snel kan teruggevonden worden. Er bestaan ook computerprogramma's die automatisch een sitemap kunnen genereren. Een sitemap is een handig instrument om de grenzen van een website vast te leggen.

<sup>8</sup> Een deel van deze informatie wordt in 'cookies' opgeslagen. Persoonsgegevens, interesses, wachtwoorden, voorkeur voor een bep. taal, enz. worden door de server in een cookie opgeslagen en naar de browser gestuurd. Bij een volgend bezoek aan dezelfde website zendt de browser de cookie in de HTTP-header naar de server. Hierdoor moet men bijvoorbeeld de taalkeuze niet meer herhalen. Een cookie is een tekstbestandje op de harde schijf van de webclient. Andere gegevens die door een browser naar een server wordt gestuurd, zijn onder andere: IP-adres, besturingssysteem, browserprogramma, schermresolutie, geïnstalleerde plug-ins.

De laatste tijd wordt ook volop gewerkt aan het toegankelijk maken en het beheren van webdocumenten. Ook op dit terrein wordt gebruiksvriendelijkheid, versiebeheer en integratie met het werkproces nagestreefd. *Web content management* is uitgegroeid tot een nieuw werkterrein binnen de IT en heeft tot doel de data van een website zo efficiënt mogelijk te beheren en op een dynamisch manier actueel te houden. Documentbeheerssystemen worden uitgebreid met modules voor het beheren van de inhoud van websites: metadata, versiebeheer en -controle, beheer van de website- en webpaginastructuur, web publishing, link management, enz. Eén van de nieuwe functionaliteiten in link management is het gebruik van virtuele links of persistent identifiers. Bij gewoon websitebeheer moeten de links handmatig worden gelegd of aangepast. Virtuele links verwijzen niet naar één bepaalde versie van een document, maar naar een object-ID waaraan de versies zijn gekoppeld. Bij het updaten van een document wordt dan zonder het aanpassen van de link de laatste versie op de website weergegeven. Persistent Identifiers zorgen ervoor dat de link blijft werken wanneer het doel wordt verplaatst.

Eén van de laatste ontwikkelingen is de verdere evolutie van HTML in de richting van een echte markuptaal. In de laatste (X)HTML-specificaties worden de (X)HTML-tags minder voor de opmaak van de pagina gebruikt. In de plaats daarvan worden stylesheets gebruikt.

### III. BELANG VAN WEBSITES ARCHIVERING

Het archiveren van een website kan vanwege diverse redenen belangrijk zijn. De archiefwaarde van websites hangt samen met de evolutie van statische tot interactieve websites.

De archivering van websites wordt ten eerste verantwoord vanuit de documentaire waarde die ze zelf hebben. Gearchiveerde websites zijn onmisbaar als bronnenmateriaal voor onderzoek naar de geschiedenis en de evolutie van dit medium. Het Internet en in het bijzonder het WWW heeft de manier waarop we informatie verspreiden en opzoeken grondig veranderd. Zonder gearchiveerde websites is het bijna onmogelijk om zich een beeld te vormen over hoe websites er uitzagen, voor wat ze werden gebruikt, welke informatie ze bevatten, hoe ze zich verhielden tegenover andere media, hoe een bepaalde instelling of onderneming zich via het web profileerde, welke mogelijkheden ze boden, hoe webdesign er omstreeks 1997 uitzag, hoe HTML en welke scripts werden gebruikt, enz.

Het ligt voor de hand dat websites ten tweede worden gearchiveerd vanwege de informatie die ze bevatten. Websites hebben in veel gevallen een grote informatieve waarde en zijn in de toekomst zeker bruikbaar als bronnen voor onderzoek van gelijk welke aard. Vanaf de eerste generatie werden websites overwegend voor de verspreiding van informatie gebruikt. Hun informatieve waarde is ondertussen nog sterk toegenomen. Bij het begin van de ontwikkeling van het WWW kon men immers nog vaststellen dat informatie ook via een website in digitale vorm ter beschikking werd gesteld. De informatie op deze websites (bijv. in HTML of PDF) was meestal nog in een andere vorm (bijv. in een tekstverwerkingsbestand) en op een andere plaats bij de archiefvormer aanwezig. Websites worden nu meer de exclusieve publicatieplaats. Websites verdringen alsmaar meer de traditionele kanalen

waarlangs informatie wordt verspreid. Deze evolutie heeft onder meer tot gevolg dat informatie bestemd voor een website meer onmiddellijk in een geschikt formaat voor het Internet wordt aangemaakt. Een ander gevolg van deze evolutie is dat de inhoud van papieren informatiedragers verandert omdat hun vroegere inhoud voortaan via een website wordt bekend gemaakt<sup>9</sup>. Inzake informatievoorziening worden websites en papieren publicaties dus meer complementair in plaats van elkaar te overlappen. Tot op de dag van vandaag heeft een groot aandeel van de websites de digitale verspreiding van informatie als voornaamste doel. Aansluitend bij de informatieve waarde is het ook belangrijk dat websites als on line bronnen op één of andere manier kunnen vastgelegd worden. Het “bevriezen” van websites is een praktische noodzaak want de vluchtigheid van het medium leidt tot snelle aanpassingen. De inhoud wijzigt snel en URL’s veranderen, waardoor het achteraf moeilijk kan zijn om bepaalde informatie terug te vinden.

Websites hebben naast een documentaire en informatieve waarde ten derde ook een culturele waarde. Ze behoren tot ons digitale erfgoed. Het zijn materiële getuigenissen van onze samenleving die zonder archiveringsbeleid voor de toekomst verloren gaan.

Websites worden tenslotte ook gearchiveerd omdat ze archiefbescheiden zijn, archiefbescheiden bevatten of omdat er door middel van een website archiefbescheiden worden gecreëerd. De huidige generatie websites hebben meer functionaliteiten dan de traditionele informatiekkanalen. Websites spelen een steeds grotere rol in het werkproces. Ze worden niet louter meer als publicatie gebruikt, maar worden volop ingeschakeld in het dienstverlenings- en bedrijfsvoeringsproces. De huidige tendens naar e-government en e-commerce illustreert dit. Transacties of handelingen die verantwoording vereisen, worden steeds meer via Internet- en intranetsites uitgevoerd. De (digitale) neerslagen van deze transacties zullen als bewijs of als documentatie gelden en moeten mee in het archiveringssysteem worden opgenomen. Websites geven ook aanleiding tot de creatie van archiefdocumenten via het web. Dit kunnen e-mails en databanken zijn. De websites bevatten dan doorgaans het formulier waarlangs de gegevens worden ingevoerd, als ook de richtlijnen en de procedure. In die zin behoren deze websites tot de context van deze archiefbescheiden.

Deze laatste motivering gaat zeker op voor de websites van overheidsdiensten. De overheid moet zich kunnen verantwoorden voor de informatie die ze via het WWW verspreiden, want haar website kan de handelingen en beslissingen van burgers en andere diensten sturen. Overheidswebsites zijn op hun beurt ook een informatiebron over de organisatie, taken, bevoegdheden, beleid, richtlijnen, enz. Vanwege de latere aansprakelijkheid of verantwoording en hun historische waarde moeten deze websites worden gearchiveerd. Wat als een bepaalde burger meent iets te hebben gelezen op de website en daarop aanspraak maakt? In Australië bijvoorbeeld is heel duidelijk vastgelegd dat de websites van de overheid archiefbescheiden zijn, want de overheid moet rekenschap kunnen afleggen over de informatie die ze via het Internet verspreidt. Elke wijziging die een website ondergaat wordt er gearchiveerd<sup>10</sup>. Als dekking tegen mogelijke aansprakelijkheidseisen is het aangewezen om zelf gearchiveerde versies van de eigen website bij te houden.

Websites van overheidsdiensten zijn bovendien in vele gevallen bestuursdocumenten. Het is niet alleen een middel om overheidsinformatie beschikbaar te maken, hoewel men meestal alleen dit aspect van overheidswebsites belicht. Het Vlaamse decreet inzake openbaarheid van bestuur van 18 mei 1999 definieert het begrip “bestuursdocument” als “de drager, *in welke vorm ook*, van informatie waarover

---

<sup>9</sup> G. VOERMAN (e.a.), *Het belang van het archiveren van websites*, in *Information Professional*, 2001, p. 17.

<sup>10</sup> *A policy for keeping records of web-based activity in the Commonwealth Government*, p. 11-12.

een administratieve overheid beschikt.”<sup>11</sup> Hoewel men zich over de kwestie van websites als bestuursdocument nooit formeel heeft uitgesproken in Vlaanderen, kan een overheidswebsite aangemerkt worden als een elektronische publicatie van die overheid met de bedoeling om de burger zo goed mogelijk over het beleid te informeren en om een snelle, gebruiksvriendelijke en transparante dienstverlening mogelijk te maken. De reden voor dit stilzwijgen over overheidsinformatie en ICT in de doctrine ligt voor hand. Om het statuut van overheidswebsites als bestuursdocumenten te benadrukken, zetten we deze evidentie hieronder toch even in de verf.

Openbaarheidswetgeving bekommert zich om de legitimiteit en de geloofwaardigheid van het bestuur als belangrijk onderdeel van een democratische rechtsstaat. Om deze reden heeft iedere burger het recht om toegang te krijgen tot bestuursdocumenten.<sup>12</sup> Het recht op toegang tot bestuursdocumenten wil paal en perk stellen aan de praktijk van geheimhouding en een gebrek aan transparantie binnen het bestuur, die kenmerkend was voor de vooroorlogse periode.<sup>13</sup> Het Vlaams decreet inzake openbaarheid voorziet een aantal maatregelen in het kader van de zgn. actieve en passieve openbaarheid van bestuur om ervoor te zorgen dat de burgers effectief toegang hebben tot bestuursdocumenten. Actieve openbaarheid houdt de verplichting in om “*de bevolking systematisch, tijdig en in begrijpelijke vorm voor te lichten over het beleid, de decreten, de besluiten en andere regelgeving, alsook over hun dienstverlening en over de informatie die bij hen beschikbaar is*”.<sup>14</sup> In de praktijk gebeurt dit tegenwoordig door middel van brochures en informatie die via verschillende websites<sup>15</sup> wordt megedeeld. De Vlaamse informatieambtenaar heeft als taak om de bevolking in te lichten over het beleid en de dienstverlening van de Vlaamse regering en het ministerie van de Vlaamse Gemeenschap en om de uitbouw, coördinatie en realisatie van het voorlichtingsbeleid van de Vlaamse regering te stimuleren en te coördineren.<sup>16</sup> Passieve openbaarheid betekent dat de burger de bestuursdocumenten mag raadplegen, uitleg mag vragen over de inhoud ervan en een afschrift ervan kan verkrijgen.

Welnu, een overheidswebsite is een overheidspublicatie (een digitale brochure als het ware, of een verzameling van digitale formulieren in het geval van proactieve dienstverlening) die per definitie toegankelijk is voor iedere burger die over een internetaansluiting beschikt. De kwestie van toegankelijkheid van dit bestuursdocument stond dus nooit ter discussie. Toch is het duidelijk dat ook de digitale publicatie aan de hand waarvan overheidsinformatie ter beschikking wordt gesteld, nl. de website, een bestuursdocument uitmaakt. Hoewel de openbaarheidsreglementering op zich weliswaar geen bepalingen bevat over het bewaren en vernietigen van bestuursdocumenten, leggen deze regels aan de administratieve overheid onrechtstreeks toch de plicht op om alle bestuursdocumenten ter beschikking te houden van de burger die erom verzoekt. Administratieve overheden moeten bijgevolg de verschillende versies van hun website archiveren.

---

<sup>11</sup> Voor meer uitleg over het begrip “bestuursdocument”, zie: BOUDREZ, F. en VAN DEN EYNDE, S., *Archiveren van e-mail*, Stadsarchief Antwerpen – I.C.R.I., Antwerpen - Leuven, oktober 2001, 34-35.

<sup>12</sup> Art. 32 van de Grondwet

<sup>13</sup> Hoewel in België de eerste stappen naar een echt openbaar bestuur pas werden gezet met de wet van 29 juli 1991 betreffende de uitdrukkelijke motivering van bestuurshandelingen (*B.S.* 12 september 1991).

<sup>14</sup> Art. 21 §1 van het decreet.

<sup>15</sup> Bijv. de portaalsite van de Vlaamse overheid: <http://www.vlaanderen.be>, of de webstek van de Vlaamse infolijn: <http://www2.vlaanderen.be/infolijn>

<sup>16</sup> Art. 22 §2 van het decreet.

Indien de burger gebruik wenst te maken van zijn recht op inzage, dan stelt de administratieve overheid in overleg met de aanvrager de plaats, de datum en het tijdstip van de inzage vast.<sup>17</sup> Voor de versie van de website die on line staat, heeft deze bepaling natuurlijk geen zin. De burger kan ten allen tijde de website bekijken. Ten aanzien van de gearcheverde versies, kan de overheid twee dingen doen. Zij kan aan haar website een subpagina “archief” toevoegen, waar de vorige versies van de website nog steeds on line raadpleegbaar zijn. De inzage kan dan op afstand “ter plaatse” gebeuren via het Internet. “Ter plaatse” wordt dan “van op het scherm”.<sup>18</sup> Zij kan de website ook lokaal archiveren in het archiveringssysteem. In dit laatste geval moet de burger zich overeenkomstig het decreet wenden tot de dienst waar de website zich bevindt.

Wat moet er in dit geval gebeuren indien de website gearcheveerd werd door een normaal vooruitziende en zorgvuldige overheid, maar zich niet bevindt bij de dienst waartoe het verzoek gericht is?<sup>19</sup> De Commissie voor de toegang tot bestuursdocumenten is van mening dat een administratieve overheid niet kan zeggen dat ze niet over de informatie beschikt indien het verzoek betrekking heeft op documenten waarvan zij zelf de auteur is.<sup>20</sup> Bovendien mogen we ervan uitgaan dat elke overheidsdienst een kopie bijhoudt van al haar bestuursdocumenten.<sup>21</sup> Ze kan zich dus niet ontdoen van haar verplichting door te stellen dat ze de vorige versies van de website niet meer in haar bezit heeft sinds de archivering.

Het decreet verleent ook het recht om een afschrift te verkrijgen van een bestuursdocument. Door de (huidige of gearcheverde) website te bekijken op zijn PC, genereert de webbrowser automatisch een “afschrift” van de website. De komst van het Internet doet de grenzen tussen actieve en passieve openbaarheid vervagen. Indien in het verzoek door de burger gevraagd wordt om de gearcheverde websites die niet on line raadpleegbaar zijn, elektronisch ter beschikking te stellen, bijv. via e-mail, dan moet de overheid dit verzoek inwilligen.<sup>22</sup> In de wet van 12 november betreffende de openbaarheid van bestuur in de provincies en de gemeenten (B.S. 19 december 1997) wordt er evenwel bepaald dat een mededeling in afschrift van een auteursrechtelijk beschermd werk niet is toegestaan dan met voorafgaande toestemming van de maker of van de persoon aan wie de rechten zijn overgegaan. Men kan zich de vraag stellen of het auteursrecht hier niet een ongewenste hinderpaal is bij de uitoefening van zijn recht op informatie. Bovendien heeft deze beperking op de openbaarheid nog weinig zin ten aanzien van digitale bestuursdocumenten op Internet, aangezien de webbrowser automatisch (en dus zonder de toestemming te vragen van de auteur) een afschrift genereert.

Alleen de administratieve overheden moeten met deze openbaarheidsreglementering rekening houden. Dit betekent dat noch de websites van de wetgevende overheden <sup>23</sup>, noch die van de

<sup>17</sup> Art. 11 §4 van het decept.

<sup>18</sup> DUMORTIER, J., JANSSEN, K. e.a., *Transparante overheidsinformatie als competitief voordeel voor Vlaanderen. Literatuurstudie*, K.U.Leuven, 85.

<sup>19</sup> Bijv. omdat de bewaring gecentraliseerd verloopt door en bij één dienst.

<sup>20</sup> Commissie voor de Toegang tot Bestuursdocumenten, advies 95/58.

<sup>21</sup> DUMORTIER, J., JANSSEN, K. e.a., o.c., 84.

<sup>22</sup> DUMORTIER, J., JANSSEN, K. e.a., o.c., 85.

<sup>23</sup> Over de openbaarheid van wetgeving stelt art. 190 van de Grondwet dat “geen wet, geen besluit of verordening van algemeen, provinciaal of gemeentelijk bestuur verbindend is dan na te zijn bekendgemaakt in de vorm bij de wet bepaald.” Voor de wetten, de koninklijke en ministeriële besluiten is dit algemeen rechtsbeginsel uitgewerkt in de wet van 31 mei 1961 (B.S. 21 juni 1961), die het Belgisch Staatsblad noemt als kanaal voor de bekendmaking. Of de overheid verplicht kan worden om overheidsinformatie zoals wetgeving, bekend te maken via het Internet, is een vraag die we hier niet behandelen. Deze problematiek wordt door ICRI en het departement Communicatiewetenschappen van de K.U.Leuven bestudeerd in het PBO-project: <http://www.law.kuleuven.ac.be/icri/projects/pbo.htm>.

rechterlijke overheden <sup>24</sup> specifiek omwille van de openbaarheid moeten bewaard worden. Deze websites moeten evenwel toch gearchiveerd worden omdat iedere overheid zich ten allen tijde moeten kunnen verantwoorden over de informatie die ze via het Internet verspreidt.<sup>25</sup>

Een website van een administratieve overheid kan niet alleen zelf een bestuursdocument uitmaken, een website kan ook bestuursdocumenten bevatten of creëren. Ook zij dienen zorgvuldig bewaard te worden, ook al gaat het hier op het eerste zicht om vluchtige transacties van digitale informatie die via de website worden uitgevoerd.

## IV. BESTAANDE INITIATIEVEN

De vluchtigheid van de informatie op het Internet heeft er toe geleid dat al vrij vroeg met het vastleggen van websites werd gestart. Aangezien de eerste generatie websites de status van digitale publicatie hadden, hoeft het geen verwondering te wekken dat de bibliotheekwereld met het verzamelen van (een deel) van het WWW begon.

De eerste echte projecten om websites op een systematische manier te archiveren schoten in 1996 uit de startblokken. Tijdens de zomer van 1996 begon het *Internet Archive* met het verzamelen van alle tekstuele informatie op het internet, zowel van nieuwsgroepen als van websites<sup>26</sup>. Rekening houdende met de grootschaligheid van het project kan men niet anders dan volledig geautomatiseerd werken. Het *Internet Archive* werkt samen met een commerciële partner die de data verzamelt en alle 6 maanden oude data naar het Internet Archive doorstuurt<sup>27</sup>. Sedert begin 2002 is het website-archief (de “Waybackmachine”) on line raadpleegbaar. Hetzelfde jaar startte de Nationale Bibliotheek van Zweden een gelijkaardig project op dat zich weliswaar beperkt tot de archivering van alle Zweedse websites (*Kulturarw<sup>3</sup> Project* <sup>28</sup>). De Zweedse websites worden verzameld met behulp van robots die de sites indexeren en in databanken stockeren. De doelstelling is niet alleen het bewaren van het web maar ook de bewaring van de originele “look and feel” en de surfervaring. In Australië begon de

<sup>24</sup> De overkoepelde website van de rechterlijke macht, is <http://www.juridat.be>. Juridat is gecreëerd door leden van de rechterlijke macht.

<sup>25</sup> Cf. infra: Aansprakelijkheid van de overheid voor haar eigen webstek.

<sup>26</sup> <http://www.archive.org>; <http://www.alexacom>

<sup>27</sup> De websites bewaard bij het *Internet Archive* zijn opgeslagen in ARC-bestanden. Alexa bezorgt ze in dit formaat aan het *Internet Archive*. ARC-bestanden zijn in essentie niets meer dan bestanden waarin de nodige metadata als headerinformatie aan de HTML-webpagina's is toegevoegd (*encapsulation*). De ARC-bestandsspecificatie is vrijgegeven voor Alexa (<http://www.alexacom/company/arcformat.html>). In het ARC-formaat worden zowel de metadata als de HTML-file opgenomen. De metadata omvatten: versie, URL en IP-adres, archiveringsdatum, MIME-type, aantal karakters in het HTML-bestand, servernaam, datum laatste wijziging. In één ARC-bestand (ongeveer 100 Mb groot) worden meerdere HTML-bestanden opgeslagen. Voor de ontsluiting van de ARC-bestanden wordt een externe databank bijgehouden.

<sup>28</sup> <http://www.kb.se/eng/kbstart.htm>; <http://kulturarw3.kb.se>; K. PERSSON, *The Kulturarw3 Project - The Swedish Royal Web Archiw<sup>3</sup>*, Lezing gehouden in Svetlogorsk, aug. 2000; A. ARVIDSON, *Harvesting the Swedish webspace*, Lezing gehouden in Darmstadt, 8 sept. 2001.

Nationale Bibliotheek in het kader van het *Pandoraproject*<sup>29</sup> met het archiveren van websites, nieuwsgroepen en mailinglijsten. Dit project richt zich op de belangrijkste Australische on line publicaties en maakt dus op voorhand een selectie van de websites. Het Amerikaanse *Minerva-project*<sup>30</sup> heeft een gelijkaardige aanpak. Minerva verzamelde de websites van de kandidaten voor de presidentsverkiezingen van 2000. Beide projecten volgen een selectieve benadering en leggen met behulp van een off line browser een website vast. Op basis van de opgedane ervaring wil men overstappen naar de vastlegging van hun respectieve webruimtes. In Nederland archiveert het *Occasio-project*<sup>31</sup> internieuwsgroepen en het Documentatiecentrum Nederlandse Politieke Partijen<sup>32</sup> de websites van politieke partijen. Binnenkort start de Bibliothèque nationale de France met een project om de Franse webruimte vast te leggen<sup>33</sup>. Ook in Oostenrijk loopt er momenteel een pilootproject<sup>34</sup>.

Parallel met de functieverschuiving is het vastleggen van websites geleidelijk aan ook een archiefaangelegenheid geworden. Websites zijn immers geëvolueerd van statische digitale publicaties tot dynamische en interactieve instrumenten waarmee op maat gemaakte diensten en informatie wordt verstrekt. Immers, de mogelijkheden van websites breiden onder invloed van de IT-evolutie alsmaar verder uit. In de landen waarin e-government en digitale dienstverlening zoals het e-loket al volop operationeel is, gaat men nog een stap verder. De archiefdiensten werken er een archiveringsbeleid uit voor de archiefdocumenten die door middel van een website worden gevormd of voor de transacties die langs een website gebeuren. Begin 2000 startte het NARA (National Archives and Records Administration) de archivering van de websites van alle federale diensten. Elke dienst werd gevraagd om op het einde van de Clinton-ambtstermijn een snapshot van de website bij het NARA neer te leggen<sup>35</sup>. De Australische Nationale archiefdienst publiceerde begin 2001 een beleidsnota en richtlijnen voor het archiveren van websites binnen de overheid. In Australië wordt ook onderzocht welke metadata worden bijgehouden<sup>36</sup>. Het Public Record Office archiveerde de website van Downstreet 10 naar aanleiding van de verkiezingen van juni 2001<sup>37</sup>.

<sup>29</sup> <http://pandora.nla.gov.au>; W. CATHRO, C. WEBB en J. WHITING, *Archiving the web: the pandora archive at the National Library of Australia*, Lezing gehouden tijdens de conferentie Preserving the Present for the Future Web Archiving, Kopenhagen, 18-19 juni 2001.

<sup>30</sup> <http://www.cs.cornell.edu/wya/LC-web>. De website van het Minerva-project staat op <http://www.loc.gov/minerva> maar is (nog) niet toegankelijk; C. AMMEN, *MINERVA: Mapping the INternet Electronic Resources Virtual Archive -Web Preservation at the Library of Congress*, Lezing gehouden te Darmstadt op 8 sept. 2001.

<sup>31</sup> <http://www.iisg.nl/occasio>; J. QUAST, *Het internetarchief van het IISG*, in *Nederlands Archievenblad*, september 2000, p. 16-17.

<sup>32</sup> <http://www.archipol.nl>. Archipol hanteert een selectieve benadering en gebruikt HT Track en archipol.cgi om websites te archiveren.

<sup>33</sup> J. MASÉNAS, *The BnF-project for web-archiving*, Lezing gehouden te Darmstadt op 8 sept. 2001.

<sup>34</sup> A. RAUBER e.a., *Austrian on-line archive. Current status and next steps*, Lezing gehouden te Darmstadt op 8 sept. 2001.

<sup>35</sup> <http://www.nara.gov/records/websnapshot.html>

<sup>36</sup> *A policy for keeping records of web-based activity in the Commonwealth Government*, januari 2001; *Guidelines for keeping records of web-based activity in the Commonwealth Government*, maart, 2001; S. MCKEMMISH en G. ACLAND, *Assessing essential evidence on the web: towards an Australian recordkeeping metadata standard*,

<sup>37</sup> <http://www.records.pro.gov.uk/documents/prem/18/1/default.asp>; D. RYAN, *Archiving the no. 10 website - the story so far*, Lezing te Londen, 25 april 2002.

Uit deze eerste archiveringservaringen kunnen al een aantal lessen getrokken worden met het oog op het ontwikkelen van een eigen archiveringssysteem. Met uitzondering van *Pandora* en *Minerva* gaan de bibliotheekinitiatieven uit van het systematisch vastleggen van (een deel van) het WWW. Hun eerste opdracht was de ontwikkeling van een op maat gemaakt computerprogramma voor het aanleggen van een bulkverzameling websites<sup>38</sup>. Gezien de omvang van te archiveren websites verlopen alle handelingen automatisch (indexeren, off line plaatsen, beschrijven, metadata extraheren, beschikbaar stellen, enz.) en is de menselijke tussenkomst tot een minimum herleid. Deze computerprogramma's worden harvesters of webspiders genoemd en werken volgens hetzelfde principe als de indexeringsrobots van de internetzoekmachines. Zij starten op één bepaalde URL en volgen vervolgens de hyperlinks in de webpagina's. De harvesters moeten voorzien zijn van een goed uitgebouwde versie- en duplicatiecontrole. Zo wordt vermeden dat meermaals dezelfde website of mirrorsites worden gestockeerd. Mits ingebouwde controles blijft voor het systematisch archiveren van een nationaal domein als dat van Zweden ongeveer 1000 gigabytes opslagcapaciteit vereist. De laatste 'oogst' van het Zweedse web leverde 30 miljoen bestanden en 15 miljoen webpagina's op. Het Zweedse project heeft sinds 1997 al 7 keer een snapshot genomen van de Zweedse webspace.

De keuze voor een bulkarchiveringsoptie heeft een aantal nadelen waardoor deze piste niet aangewezen is voor een websitesarchiveringssysteem voor archiefdiensten en -instellingen. Ten eerste neemt één harvestoperatie veel tijd in beslag. Het downloaden van alle websites duurt al gauw een paar maanden, waardoor de websites slechts met een lage frequentie worden vastgelegd<sup>39</sup>. Deze lage frequentie contrasteert met de vluchtigheid en snelheid van het medium. Voor het archiveren van elke versie van een website kan deze piste niet worden gevolgd. De bibliotheekwereld is zich van dit probleem bewust. In Frankrijk en Oostenrijk bijvoorbeeld is men van plan om naast het bulkproces op voorhand geselecteerde websites met een hogere frequentie vast te leggen (bijv. digitale kranten en tijdschriften).

Ten tweede is er bij bulkarchivering nagenoeg geen kwaliteitscontrole. Bij automatisch verwerving zal de kans op het ontdekken van fouten heel klein zijn en niets garandeert dat websites in hun actieve vorm geen aanpassingen behoeven zodat er een langdurige digitale leesbaarheid is verzekerd. Dit was overigens één van de redenen waarom de werkgroep van het Australische *PANDORA*-project ervoor koos om enkel geselecteerde websites te archiveren<sup>40</sup>. De kans bestaat dat websites in een bulkarchief niet raadpleegbaar zijn. Een controle van de gearchiveerde websites is noodzakelijk. Bovendien is het ook lang niet zeker dat één type harvester met dezelfde instellingen geschikt is om elk type website te

<sup>38</sup> Voortbouwend op de Finse webarchiveringsrobot werd in het kader van het Nedlib-project een nieuwe harvester ontwikkeld. De broncode (C++) van de Nedlib harvester kan vrij gedownload worden vanaf <http://www.csc.fi/sovellus/nedlib>. De harvester moet gekoppeld worden aan een MySQL-databank. De vastgelegde websites worden omwille van twee redenen in een databank opgeslagen: indexering (retrieval) en beheer van miljoenen bestanden. De harvester van de universiteit van Helsinki heeft Solaris als besturingssysteem. Het Oostenrijkse Aolaproject bouwde een eigen harvester. Het bouwde daarvoor verder op de Nedlib-harvester.

<sup>39</sup> Bij grote harvestoperaties weet men wel wanneer de archivering wordt gestart, maar kan slechts heel vaag worden ingeschat wanneer de operatie voltooid zal zijn. Op voorhand kan men moeilijk voorspellen hoe groot een nationaal domein is en hoeveel tijd het vastleggen in beslag neemt. Vooral de grote websites vertragen het proces.

<sup>40</sup> A.R. KENNEY en O.Y. RIEGER, *The National Library of Australia's Digital Preservation Agenda, an interview with C. Webb*, in *RLG-DigiNews*, 15 febr. 2001; W. CATHRO, C. WEBB en J. WHITING, *Archiving the Web: The PANDORA Archive at the National Library of Australia* (<http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html>).

stockeren. Zo bevat de Waybackmachine een aantal versies van de website van het stadsarchief Antwerpen, maar geen enkele versie werkt en zelfs de startpagina kan niet worden bekeken.

Ten derde worden websites waarnaar niet verwezen wordt door een andere website niet vastgelegd. Ten slotte is er ook het probleem van opslagcapaciteit, raadpleging en toegankelijkheid. De hoeveelheid gearchiveerde computerbestanden is heel groot zodat hier een robuuste, maar toch snelle en gebruiksvriendelijke oplossing voor nodig is<sup>41</sup>. Archiefinstellingen en documentatiecentra zullen maar zelden over de nodige infrastructuur beschikken.

Het *Pandora*- en *Minervaproject* vertrekken net als alle andere archiefinitiatieven vanuit een voorafgaande selectie van websites met archiefwaarde. Net zoals bij de bulkbibliotheekprojecten gaat er niet veel aandacht uit naar het deep web. Nochtans is de archivering van het deep web een essentieel onderdeel van de archivering van dynamische en interactieve websites. Websites die in het kader van e-commerce en e-government worden gebruikt, zijn net dynamische en interactieve websites. Opvallend bij deze projecten is ook de keuze voor één bepaalde archiveringsstrategie voor alle websites zonder dat met de website zelf wordt rekening gehouden. De architectuur en aard van de website bepaalt nochtans mee hoe een website voor archivering kan worden vastgelegd.

## V. ARCHIVEREN VAN WEBSITES

Een archiveringssysteem voor websites richt zich op de archiefdocumenten die op het WWW worden gepubliceerd en op de archiefstukken die op basis van de interactie of transactie via een website worden gecreëerd. Deze laatste categorie computerbestanden wordt gearchiveerd vanwege hun contextuele waarde, als bewijsstuk of gewoon om de inhoud van een website te reconstrueren.

Het is dan ook belangrijk om in de eerste plaats duidelijke doelstellingen voor het archiveringssysteem te formuleren. De doelstellingen bepalen immers WAT er van het WWW of een bepaalde website wordt gearchiveerd. Op zijn beurt zal de WAT-vraag uitmaken HOE websites worden vastgelegd. Ook het type website zal deze laatste vraag mee bepalen. Bij opname in het archiveringssysteem moet vervolgens ook worden nagedacht over beschrijving (metadata), de maatregelen die een duurzame archivering waarborgen en het ter beschikking stellen van de gearchiveerde website en de gerelateerde digitale archiefbescheiden.

Voor websites wordt een archivering in digitale vorm nagestreefd. De reden ligt voor de hand. Websites zijn in hun primaire vorm digitaal en latere consultatie gebeurt bij voorkeur op een manier die zo dicht mogelijk de on line versie benadert. Enkel bij digitale archivering bewaart men zoveel mogelijk originele eigenschappen van de website. Er bestaan een paar technieken om websites op

<sup>41</sup> De bestaande projecten gebruiken voornamelijk tapes en harde schijven als opslagmedium. Weinig geraadpleegde bestanden worden op tape bewaard, veel geraadpleegde op harde schijf. Bulkopslag stelt ook bijzondere vereisten voor het besturingssysteem. De conventionele besturingssystemen van personal computers hanteren een beperking op het aantal directories en bestanden dat ze kunnen bevatten.

papier te archiveren (hard copy), maar hierbij gaan te veel originele eigenschappen en functionaliteiten verloren<sup>42</sup>. Bovendien is hard copy archivering enkel op de inhoud gericht.

Het WWW is geëvolueerd tot een dynamisch en interactief medium. Het vastleggen van websites en de gerelateerde computerbestanden contrasteert met haar dynamisch karakter zodat elke poging wel op één of andere manier met informatie- en eigenschappenverlies samengaat. Het web is ook dynamisch in de zin dat het voortdurend in evolutie is. De website van een bepaalde organisatie kan immers evolueren van een statische naar een dynamische en interactieve website. Het archiveringssysteem zal dan ook voortdurend bijsturing behoeven.

In tegenstelling tot andere projecten die met websitesarchivering bezig zijn, wordt hier geen enkelvoudig archiveringssysteem uitgewerkt. Bij elke keuzestap worden verschillende scenario's geschetst. De implementatie van een archiveringssysteem voor websites is immers afhankelijk van de doelstellingen, van het type website en de beschikbare technologische middelen. Een voorbeeld van een mogelijk archiveringssysteem wordt in het volgend hoofdstuk uitgewerkt.

## **A. KWALITEITSEISEN VOOR GEARCHIVEERDE WEBSITES**

Bij de opname in het digitaal depot moeten websites aan een aantal kwaliteitsvereisten beantwoorden. Deze kwaliteitsvereisten zijn dezelfde voor statische websites als voor dynamische en interactieve websites. Een gearchiveerde website moet aan de volgende kwaliteitseisen voldoen:

- ☒ alle bestanden worden gearchiveerd die nodig zijn om een zo getrouw mogelijke reconstructie van de website te maken (tekst, afbeeldingen, stylesheets, scripts, logbestanden, databanken, gebruikersprofielen, enz.).
- ☒ de startbestanden (o.a. default.htm, index.htm, start.htm of welcome.htm) en de submappen van één versie worden in één map binnen het website-archief geplaatst.
- ☒ de bestandenstructuur en de bestandsnamen worden zo getrouw mogelijk van de webserver overgenomen. Bij webpagina's met een vaste inhoud kan identiek dezelfde bestandsnaam worden overgenomen. De bestandsnaam van webpagina's met een dynamische inhoud leunt zo dicht mogelijk aan bij de originele bestandsnaam.
- ☒ voor de interne links worden relatieve pathaanduidingen gebruikt, voor de externe links absolute pathaanduidingen. Past men voor de interne links geen relatieve maar absolute links toe, dan gaat men bij het raadplegen van de achterliggende HTML-pagina op zoek naar de on line versie en raadpleegt men bijgevolg niet meer de gearchiveerde webpagina. Relatieve pathaanduidingen maken ook het beheer gemakkelijker (bijv. verplaatsen versies). Voor de links naar bronnen buiten de eigen website worden wel absolute

---

<sup>42</sup> Men kan de actieve pagina's in een webbrowser afdrukken. Een bijzondere capturetechniek voor websites wordt door Adobe's Acrobatprogramma geboden. Met dit programma is het mogelijk om een website als een PDF-bestand op te slaan, waarbij de hyperlinks in bookmarks worden omgezet. Het spreekt voor zich dat archiveren als hard copy functionaliteits- en vormverlies met zich mee brengt en dat de broncode niet wordt vastgelegd. Deze techniek is wel opmerkelijk sneller dan het off line plaatsen van een website, maar kan eigenlijk enkel worden gebruikt om de informatie op een snelle manier vast te leggen. Dit programma heeft overigens dezelfde algemene tekortkomingen als een off line browser.

pathaanduidingen gebruikt<sup>43</sup>. Links naar virtuele mappen worden zoveel mogelijk omgezet naar relatieve links.

- ☑ actieve elementen zoals datum en bezoekersaantal worden uitschakeld. De datum en het bezoekersaantal moeten die van het ogenblik van de momentopname zijn. Bovendien kan de scripting voor foutmeldingen of oneindige loops zorgen. Beide gegevens worden in de metadata opgenomen.
- ☑ IT-afhankelijkheden (hardware, software, internetprotocollen, enz.) worden tot een minimum beperkt. Er wordt zo systeemafhankelijk mogelijk gearchiveerd. De computerbestanden waaruit een website is samengesteld, zijn zoveel mogelijk gestandaardiseerd. De tags en attributen van de toegepaste markuptaal maken deel uit van de gestandaardiseerde (X)HTML-specificatie.
- ☑ alle onderdelen van een webpagina worden op hetzelfde ogenblik gearchiveerd. De in line afbeeldingen en de webpagina's worden gelijktijdig vastgelegd.
- ☑ de gearchiveerde website en gerelateerde archiefdocumenten worden in het archiveringssysteem van de organisatie opgenomen. De digitale archiefdocumenten worden veilig bewaard en beschreven op basis van hun metadata

## B. SELECTIE: WAT ARCHIVEREN?

De vraag WAT er wordt gearchiveerd bestaat uit twee deelvragen die elk de archiveringsstrategie mee bepalen. De selectieproblematiek komt op twee niveaus aan de orde. In de eerste plaats moet er een acquisitiebeleid worden opgesteld. Welke websites worden gearchiveerd? Nadien wordt van elke website vastgelegd wat er precies in het archief wordt opgenomen. Met het beantwoorden van deze vragen worden de globale doelstellingen van het archiveringssysteem geformuleerd. Op basis hiervan kan dan het archiveringssysteem worden uitgewerkt.

### B.1 Het acquisitiebeleid: welke websites archiveren ?

Aangezien het WWW een vrij medium is, kan elke particulier, vereniging of instelling een website op het web plaatsen. Eind 2001 waren er naar schatting 8,4 miljoen verschillende websites op het WWW<sup>44</sup>. Het vooropstellen van een duidelijk acquisitiebeleid is belangrijk, want het bepaalt de uitgangspositie van het archiveringssysteem. Bij bulkarchivering verloopt het opnameproces volledig

---

<sup>43</sup> Men zou kunnen overwegen om de externe links te documenteren zodat de gebruiker later kan inschatten welke informatie achter deze link schuilde. Men kan dit doen in de vorm van HTML-commentaar, de attributen ALT of LONGDESCR of men kan de link stoppen en naar een afzonderlijke HTML-pagina met meer uitleg leiden. Deze aanpassingen zijn allemaal arbeidsintensief en bovendien overbodig wanneer de basisregels inzake *Web Content Accessibility* worden toegepast. Eén van die regels schrijft voor dat elke link voldoende moet gedocumenteerd worden in de webpagina zelf. Nochtans is dit één van de aanbevelingen van C. Dollar. (C. DOLLAR, *Archival preservation of smithsonian web resources: strategies, principles, and best practices*, 4.3).

<sup>44</sup> <http://wcp.oclc.org> ; <http://www.pandia.com/searchworld/2000-39-oclc-size.html>

automatisch, doet men een beroep op harvesters en beperkt de archivering zich tot de informatie die de webclient ontvangt. De originele webserverbestanden, het ‘deep web’, de logbestanden en de digitale neerslagen van transacties kunnen enkel bij een selectieve benadering worden gearcheveerd.

Bij het vastleggen van een acquisitiebeleid is er bij voorkeur enig overleg tussen bibliotheek en archief. In Australië en Canada bijvoorbeeld is er een taakverdeling. In Australië bewaren de bibliotheken de websites die de status hebben van een digitale publicatie, terwijl de archiefdiensten de websites archiveren die archiefbescheiden zijn, bevatten of genereren<sup>45</sup>. In Canada bewaren de bibliotheken de internetsites en de archieven de intranetsites<sup>46</sup>. De taakverdeling is grotendeels gebaseerd op de functie van websites. In de realiteit zullen veel websites zowel een digitale publicatie als een archiefbescheid zijn. Voor websites met deze gemengde status worden best afspraken tussen bibliotheek en archief gemaakt. Overlappingsen zullen wellicht onvermijdelijk zijn.

In tegenstelling tot de lopende buitenlandse bibliotheekprojecten zullen de Vlaamse documentatiecentra en archiefstellingen eerder vertrekken vanuit een selectief acquisitieprofiel. De selectie zal in grote mate samenvallen met hun algemeen acquisitie- of verzamelbeleid. In het geval van de privaatrechtelijke archiefinstellingen of documentatiecentra ligt het voor de hand dat ze de websites verzamelen van de instellingen die binnen hun onderzoeksterrein actief zijn. De functie van de websites doet voor deze archiefinstellingen weinig terzake, want ze verzamelen ook papieren documentatie en publicaties.

De archiefdiensten van openbare besturen archiveren om te beginnen de websites van hun eigen organisatie of instelling. Deze archiefdiensten zullen zich in veel gevallen niet hiertoe beperken. Naar analogie met papieren publicaties worden ook websites die dienen als bron of als hulpmiddel voor historisch onderzoek gearcheveerd. Voorbeelden van dergelijke websites zijn de verkiezingssites van lokale afdelingen van politieke partijen, een website met relevante heemkundige of genealogische informatie of de websites van personen of evenementen die op één of andere manier met de archiefvormer gelinkt zijn<sup>47</sup>. Het is evenmin denkbeeldig dat een archiefdienst van een gemeentebestuur websites in verband met de geschiedenis van de gemeente en zijn inwoners probeert te archiveren. Bij de selectie van te archiveren websites kan in principe het algemeen acquisitieprofiel van de archiefdienst of het documentatiecentrum worden toegepast.

De selectiebenadering laat naast een grotere kwaliteitscontrole eveneens samenwerking en contact met de archiefvormers en webdesigners toe. Dit is van belang om de context van de website vast te leggen en om op de hoogte te blijven van de evolutie, de versies en de updates. De archiefvormer is beter geplaatst dan de archivaris om dit te volgen en hij kan de archivaris ook de nodige metadata bezorgen. Met het oog op de auteurswetgeving is in België contact met de archiefvormer sowieso vereist. Zonder samenwerking met de archiefvormer kan men evenmin de gerelateerde computerbestanden die niet bereikbaar zijn voor de webclient archiveren. Het nadeel van deze aanpak tegenover bulkarchivering is natuurlijk de arbeidsintensiviteit en de hogere kostprijs per gearcheveerde website.

Een bijzonder geval zijn de websites die een portaalsite zijn. Portaalsites zijn websites die hoofdzakelijk links naar andere websites bevatten. Deze sites bieden geen inhoudelijke meerwaarde en

---

<sup>45</sup> *A policy for keeping records of web-based activity in the Commonwealth Government*, p. 8-10.

<sup>46</sup> D. LÉGER, *Legal Deposit and the Internet: Reconciling Two Worlds*, Lezing gegeven te Darmstadt, 8 sept. 2001.

<sup>47</sup> Voorbeelden toegepast op het stadsarchief Antwerpen hiervan zijn de websites van de Antwerpse Vereniging voor Romeinse archeologie, het Genootschap voor Antwerpse geschiedenis (historische informatie), het Vandyckjaar en Mode2002 Landed | Geland (evenementen) en KAPA (verenigingen).

dragen als probleem met zich mee dat de meeste links na verloop van tijd niet meer werken. De vraag kan dan ook worden gesteld, of het wel de moeite loont om deze sites te archiveren. Voor linkenpagina's die deel uitmaken van een website liggen de zaken anders. Deze linkenpagina's worden best wel gearchiveerd. Deze webpagina's zijn meestal een onderdeel van een website en worden mee gearchiveerd om de goede werking van de website niet te verstoren.

## B.2 Wat archiveren van websites?

Gearchiveerde websites moeten in de eerste plaats raadpleegbaar zijn. Dit houdt in dat de websites reconstrueerbaar en interpreteerbaar moeten zijn. Beide vereisten komen aan de orde bij de vraag WAT er moet worden gearchiveerd.

Het is belangrijk dat zowel de inhoud als de oorspronkelijke “look and feel” van een website opnieuw kan worden gereconstrueerd. Voor een reconstrueerbare website moeten in de eerste plaats alle nodige bestanden worden gearchiveerd waaruit een website wordt opgebouwd. Een eerste moeilijkheid schuilt in het vastleggen van de grenzen van een website. Het geheel van gelinkte computerbestanden waaruit een website bestaat is immers niet altijd even gemakkelijk af te bakenen. De reconstructie is ten tweede afhankelijk van de bestanden waaruit de website is samengesteld. Bestanden met dynamische webscripting kunnen niet op een webclient worden uitgevoerd, zodat het voor de reconstructie geen zin heeft de website in dit origineel formaat te bewaren. Het archiveren van de originele ASP-, PHP- en JSP-bestanden kan wel gemotiveerd worden vanwege verantwoordingsdoeleinden of het belang om originele scripts te bewaren. Naast de eigenlijke website zijn er nog een aantal gerelateerde computerbestanden die mee de inhoud en de interactie van een website weergeven. Hierbij gaat het in de eerste plaats om logbestanden van de webserver en gekoppelde databanken.

Men gaat op voorhand ook best na hoe de downloads van de website in het digitaal depot worden opgenomen: worden ze samen met de website gestockeerd of worden ze afzonderlijk gearchiveerd? Downloads zijn de bestanden die via de website beschikbaar worden gesteld en die doorgaans tekst- of audio- visuele bestanden zijn (PDF, MS Word, MP3, zip, tar). Deze keuze heeft gevolgen voor de nodige opslagcapaciteit voor elke gearchiveerde versie. Een oplossing zou kunnen zijn: de downloads die op de webserver staan samen met de gearchiveerde website bewaren en de downloads in een databank of content managementsysteem op een applicatieserver afzonderlijk archiveren. In dit laatste geval moet in de metadata van de website naar deze informatie worden verwezen en moet duidelijk zijn welke download op welke webpagina beschikbaar was. Bij de archivering van de downloads wordt het digitale archiveringssysteem ook afgestemd op het papieren archiveringssysteem. Nog heel wat publicaties zijn in papieren vorm binnen de organisatie aanwezig en worden als dusdanig gearchiveerd. De archivering van de digitale versie (bijv. in PDF of Word) lijkt dan overbodig, tenzij de digitale versie op de website niet identiek is of extra functionaliteiten bevat.

Om een website te kunnen interpreteren moet men ook zijn context kennen. Een deel van deze informatie vindt men terug in de logbestanden van de webserver of in de metadata. Voor websites die deel uitmaken van een heel informatiesysteem kan het ook wenselijk zijn om hierover documentatie bij te houden (bijv. technische fiches, beschrijving van de functionele vereisten, documenten over de technische ontwikkeling en systeemvereisten, installatiedocumenten, handleiding voor de administrator/webmaster, enz.). Dit geldt eveneens voor de documentatie over gekoppelde databanken met archiefwaarde.

Selectie is bijna enkel mogelijk op het niveau van websites, niet op het niveau van de bestanden die een website vormen. Bij het archiveren van een website wordt met andere woorden best een volledige mirror in het archief opgenomen. Zich beperken tot het (onder)deel van de website met archiefwaarde wordt moeilijk. Zo vergroot men overigens ook de kans dat de gearchiveerde website niet meer werkt. In een overzichtelijk ontworpen site worden onderdelen zoals scripts en afbeeldingen meestal in gemeenschappelijke mappen bewaard zodat het veel werk vraagt om te achterhalen welke bestanden al dan niet worden gekopieerd. Om niet op voorhand uitvoerig te moeten onderzoeken welke bestanden worden gekopieerd of om het nemen van een snapshot geen talloze keren te moeten herhalen, neemt men best een mirror van de volledige website. Zo is men zeker dat alle nodige bestanden aanwezig zijn en de website functioneert. Als bijvoorbeeld de website van een lokale afdeling van een partij een onderdeel is van de algemene website van de politieke partij, dan wordt gemakkelijheidshalve de volledige website gearchiveerd. Anders loopt men het gevaar dat bestanden, frames of framesets ontbreken en dat de gearchiveerde en on line versie grondig van elkaar verschillen. Vroeger werd ook al eens voorgesteld dat het archiveren van de tekstuele informatie het belangrijkste was en dat afbeeldingen slechts op de tweede plaats kwamen (cfr. de cache van zoekrobots). Ondertussen is het grafische aspect van websites zo belangrijk geworden dat afbeeldingen een essentieel onderdeel zijn. Bovendien bevatten veel bestanden met afbeeldingen ook tekstuele informatie.

De selectieproblematiek inzake websitesarchivering toont duidelijk aan dat er al op het tijdstip van creatie een archiefwaardering moet zijn. Websites archiveren met terugwerkende kracht zal zelden mogelijk zijn.

#### B.2.1 Websites met een vaste inhoud

Het archiveren van websites met een vaste inhoud stelt nagenoeg geen problemen, ook niet in de keuze van wat er wordt gearchiveerd. Deze websites hebben maar één inhoud en één vorm die voor iedereen gelijk is. De website op de webserver bestaat hoofdzakelijk uit statische HTML-pagina's, afbeeldingen en stylesheets. Het enige actieve element van dergelijke websites is de navigatie op basis van de vastgelegde hyperlinks. Deze websites bevatten soms wel een aantal dynamische elementen, maar kenmerkend voor deze onderdelen (bv. menubalk, roll-over afbeeldingen, automatische datum, toevoeging aan favorieten, enz.) is dat zij het resultaat zijn van de uitvoering van clientscripts.

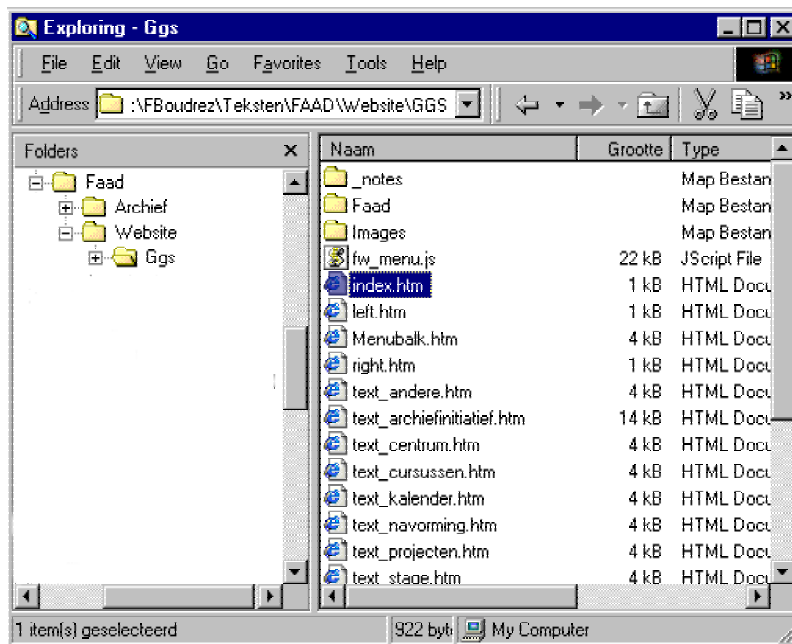
De websites met een vaste inhoud worden gearchiveerd zoals ze on line op de webserver beschikbaar zijn. Alle bestanden in een gestandaardiseerd bestandsformaat worden in hun origineel formaat opgeslagen. Niet gestandaardiseerde bestanden worden bij voorkeur naar een geschikt archiveringsformaat omgezet vooraleer ze in het digitaal depot worden opgenomen. Er kan gewoon een kopie worden gemaakt van de bestanden in hun oorspronkelijke structuur zoals ze op de webserver staan, want de on line en off line versies van deze websites zijn identiek en raadpleegbaar. Een dergelijke kopie wordt een mirror genoemd. De originele bestanden reconstrueren op om het even welk hedendaags computerplatform zowel on line als off line de website.

Het maken van een kopie van de bestanden vanop de webserver zal in de meeste gevallen inhouden dat ook verouderde en niet meer gelinkte bestanden in het archief worden opgenomen. Deze bestanden blijven soms na hun update op de serverschijf staan, maar behoren niet meer tot de on line versie van de website.

De logbestanden van websites kunnen ook archiefwaarde hebben vanwege de contextuele gegevens die ze bevatten. Op basis van de serverlogbestanden kan achterhaald worden wie de website bezocht,

welke bestanden het meest werden geraadpleegd, hoe lang bepaalde webpagina's werden bekeken, welke bestanden werden gedownload, enz. De logbestanden worden bijgehouden door de webserversoftware in de vorm van platte tekstbestanden of rechtstreeks in een databank. Men kan de frequentie van de logbestanden bepalen (dagelijks/wekelijks). De meeste webserver bieden de mogelijkheid om logbestanden op te slaan in een gemeenschappelijk logformaat of in een formaat eigen aan de webserversoftware<sup>48</sup>. Een andere optie om deze contextuele gegevens te bewaren, is de archivering van de statistieken die op basis van de logbestanden werden gecreëerd. Het archiveren van alle logbestanden vereist immers heel wat opslagcapaciteit en is nogal omslachtig. Voor de analyse en het maken van statistieken is er speciale software op de markt. Het gemeenschappelijk logformaat wordt ondersteund door de meeste analysetools. Men kan ook de voornaamste gegevens in de metadata van de gearchiveerde website overnemen.

Websites met formulervelden worden ook tot de statische websites gerekend. De aanpak voor statische websites volstaat om de website en haar formulier(en) opnieuw op het scherm op te roepen. Indien de informatie verstuurd naar een webserver archiefwaarde heeft, dan moet die ook worden bewaard. De verstuurde informatie is in de meeste gevallen bij de ontvanger in de vorm van e-mails of een databank aanwezig. Of het serverscript, het uitvoerbaar programma voor de verwerking van de formuliergegevens en de bijhorende documentatie mee gearchiveerd worden is opnieuw afhankelijk van de eventuele verantwoordingsplicht.



**Afbeelding 4:** Een gearchiveerde versie van de website van de interuniversitaire archiefopleiding, een website met een vaste inhoud. De bestanden worden in hun oorspronkelijk formaat in de originele bestandenstructuur opgeslagen. Alle bestanden die samen de website vormen worden in één map samengesplaatst ('GGS'). Deze map is het vertrekpunt voor de relatieve pathaanduidingen die de bestanden aan elkaar linken.

<sup>48</sup> Er bestaan verschillende formaten voor de logbestanden: Common Logfile, Combined Logfile, NCSA Common, W3C Extended, Microsoft IIS. Een logbestand bevat sequentiële lijnen gevuld met ASCII-karakters. Logfiles hebben doorgaans de extensie log, lf of crlf. Elke lijn bevat een directive (bv. versie, velden, software (browser, versie, resolutie, besturingssysteem), verwijzer, startdatum, einddatum, datum, opmerking: een lijn met een directive begint steeds met #) of een entry. Lijnen met entrygegevens bevatten info over de HTTP-acties. De velden binnen één lijn worden met spaties van elkaar gescheiden. Directives bevatten aanwijzingen over het loggingproces. Velden bevatten een omschrijving van de informatie die in elke entry voorkomt. Het einde van een entry wordt gevormd door een CR of CRLF-aanduiding. Een voorbeeld van een logfile is raadpleegbaar op de DAVID website (cases → websites → lezing Informatie 2001 → Wat en hoe vastleggen?). Er wordt doorgaans een onderscheid gemaakt tussen de unieke bezoeken (teller+1 per computer die zich aanmeldt), pageviews (totaal aantal bekeken webpagina's), hits (teller+1 per opgevraagd bestand).

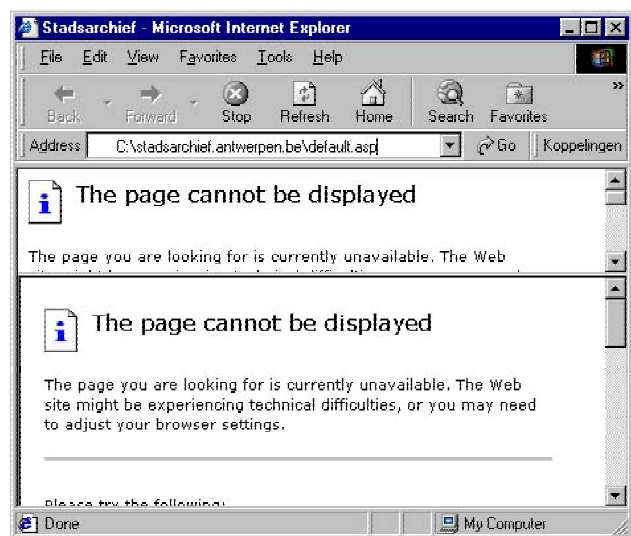
Bij een website met een statische inhoud kan ook een logbestand van het up- en downloaden voor archivering in aanmerking komen. In dit logbestand wordt bijgehouden wanneer en welke bestanden naar de webserver werden gestuurd. Uit dit bestand kan voor een stuk de historie worden afgeleid<sup>49</sup>, en eventueel kan dit logbestand mee helpen aantonen wat de inhoud op een gegeven moment was. Het FTP-logbestand over het up- en downloaden kan eventueel als basis dienen of de webadministrator legt zelf hiervoor een bestand aan<sup>50</sup>.

Met uitzondering van de formulervelden blijven de gearchiveerde statische websites hun originele functionaliteit volledig behouden.

#### B.2.2 Websites met een dynamische inhoud

De selectieproblematiek stelt meer vragen bij het archiveren van websites met een dynamische inhoud. Bij deze websites worden de HTML-pagina's pas na het ontvangen van een HTTP-verzoek op de server samengesteld of wordt de HTML-pagina vanuit een gekoppelde toepassing geleverd. De inhoud van een webpagina kan afhankelijk zijn van factoren zoals de ontvangen zoekopdracht (bijv. raadplegen uurregeling treinverkeer), het gebruikersprofiel of de voorkeuren van de bezoeker (bijv. dmv van een cookie), of op basis van de informatie die op dat ogenblik in het gekoppelde documentbeheersysteem of de achterliggende databank aanwezig is. Hier is het niet alleen de vraag wat van de website wordt gearchiveerd, maar ook of de interactie wordt bewaard. De inhoud is immers in veel gevallen afhankelijk van de server-clientinteractie en de reconstructie van de inhoud is enkel mogelijk als ook de interactie wordt gearchiveerd. Twee aandachtspunten staan centraal: wat moet gearchiveerd worden om de website later nog te kunnen raadplegen en wat moet gearchiveerd worden om de interactie - en dus ook de inhoud - vast te leggen.

Om de meeste websites met dynamische inhoud later nog te kunnen raadplegen volstaat het niet dat er gewoon een kopie van de originele bestanden wordt genomen zoals ze op de webserver staan. Een eerste probleem is de afhankelijkheid van de website aan de webserver en het 'diep web'. Deze websites zijn een onderdeel van een heel digitaal informatiesysteem en functioneren niet



**Afbeelding 5:** Websites met een dynamische inhoud kunnen niet in hun origineel formaat zonder de webserverconfiguratie en -software worden bekeken. Probeert men deze websites toch enkel en alleen met een webbrowser te bekijken, dan verschijnen er op de plaats waar de informatie moet komen enkel maar foutmeldingen.

<sup>49</sup> Op de DAVID-website staat een voorbeeld van een dergelijk logbestand (cases → websites → lezing Informatie 2001 → frequentie). In de logfile wordt bijgehouden: de datum van de up -of download, tijdstip, bronmap, doelmap, up- of download en bestandsnaam.

<sup>50</sup> In het bestand dat men zelf aanlegt, kan men alle nodige gegevens registreren. Men kan zelf kiezen hoe uitgebreid dit logbestand is. In een FTP-logbestand worden maar een beperkt aantal gegevens geregistreerd. Dit bestand kan echter wel ontstaan wanneer bijvoorbeeld ook in de HTML-headers een aantal aanvullende metadata worden opgenomen.

autonoom. De website kan in zijn originele vorm maar op het scherm weergegeven worden wanneer de oorspronkelijke webserverconfiguratie (o.a. virtuele mappen), de nodige software (serverscript en bijhorende module, serversoftware, uitvoerbaar programma) en het gekoppelde bestandensysteem (databanken, documentbeheerssystemen, content managementsysteem) actief blijven. Ook al hoeft de website niet functioneel te blijven, de website kan niet zomaar losgekoppeld worden van het achterliggende informatiesysteem. Bij de dynamische en interactieve websites is de output in een webbrowser wel een statische HTML-pagina, maar in veel gevallen zijn de ontvangen webpagina's het resultaat van een samengesteld en geïntegreerd informatiesysteem. Probeert men dit toch, dan krijgt men in de webbrowser alleen maar foutmeldingen te zien.

Een tweede moeilijkheid vloeit voort uit het feit dat de websites geen vaste inhoud hebben. Elke bezoeker krijgt al naargelang zijn verzoek of voorkeuren een andere inhoud te zien. Wat is dan de inhoud van de website? Hoe weten we dan welke informatie op een bepaalde website werd gepubliceerd?

Voor beide problemen werden al diverse oplossingen gesuggereerd, maar met de eigenlijke uitwerking moet nog worden gestart zodat hierover nog weinig of geen expertise is. Het on line archiveren van deze websites houdt in dat het achterliggende informatiesysteem operationeel moet blijven. Dit is niets meer dan de toepassing van de computermuseumstrategie en is geen goede optie voor archivering op lange termijn. Van zodra één onderdeel van het hele informatiesysteem niet meer beschikbaar is, kan de gearcheverde website niet meer geraadpleegd worden. Emulatie van deze websites zou inhouden dat er niet alleen een emulator voor de website zelf moet zijn, maar ook voor de webserversoftware, de webserverconfiguratie, de scripts, de uitvoerbare programma's en voor de koppeling met de achterliggende databanken. Bovendien zou voor elke gearcheverde website een afzonderlijke emulator moeten worden gemaakt. Dit is onrealistisch. Toepassingsvoorbeelden zijn niet bekend. Emulatie kan daarentegen wel gebruikt worden voor de webbrowsers (zie verder onder Digitale Duurzaamheid). Het Universal Preservation Format<sup>51</sup> (UPF) leek een andere mogelijke piste, maar wegens gebrek aan fondsen werd nog geen prototype ontwikkeld<sup>52</sup>. In Denemarken en Schotland wordt gedacht aan het filmen van de monitor terwijl iemand surft<sup>53</sup>. Dit is een onpraktische oplossing voor grote websites en vraagt veel tijd bij de raadpleging.

Om deze websites te archiveren volgens de vooropgezette kwaliteitsvereisten maken we best een onderscheid tussen de verschillende lagen waaruit de website wordt opgebouwd. De opsplitsing van het hele informatiesysteem in de lagen die in het DAVID-rapport *Het digitaal archiveringssysteem* werden onderscheiden, kan hierbij als uitgangspunt dienen<sup>54</sup>. De drie lagen zijn de inhoud, de structurele of logica elementen en de tools. Elke laag wordt afzonderlijk gearcheveerd.

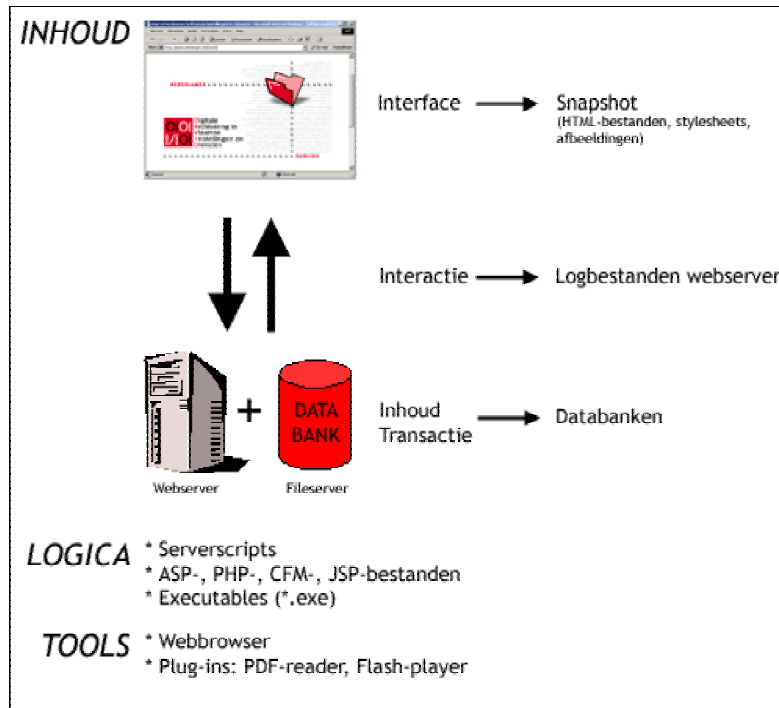
---

<sup>51</sup> Het UPF heeft tot doel een archiveringsformaat voor multimediate bestanden te zijn met onafhankelijkheid op het vlak van platform, applicatie en drager. Een UPF-bestand zou ook zelfbeschrijvend moeten zijn. Alle nodige gegevens (metadata, alle technische specificaties om toegang te krijgen tot de inhoud, ...) wordt aan het bestand toegevoegd: encapsulation (<http://info.wgbh.org/upf/>).

<sup>52</sup> E-mail van Tim Shepard, 23 juli 2001.

<sup>53</sup> B. CHRISTENSEN-DALSGAARD, *Archive Experience, not Data*, Lezing gehouden in Darmstadt, 8 sept. 2001; S. BORDWELL, *Objective-Archival preservation of websites*, Lezing gegeven te Londen, 25 april 2002.

<sup>54</sup> *Het digitaal archiveringssysteem: beheersinventaris, informatielagen en beslissingsmodel als uitgangspunt*, p. 15. De benamingen van de drie lagen in een informatiesysteem werden hier aangepast aan websites (data → inhoud, structurele info → logica, tools).



**Afbeelding 6:** De afzonderlijke archivering van de onderdelen waaruit het hele informatiesysteem is samengesteld. De interface wordt vastgelegd door een snapshot te archiveren. Zo archiveren we de manier waarop de website er uit zag en hoe de informatie op het WWW werd gepresenteerd. Om de inhoud van de website te kennen archiveren we het gekoppelde 'back-office'-systeem en de logbestanden van de webserver. De logica-elementen worden indien nodig rechtstreeks van de webserver gekopieerd. Voor de raadpleging zijn tenslotte nog de passende webbrowser en eventueel de bijhorende plug-ins vereist.

Een snapshot is een momentopname van een website zoals die op een webclient werd aangeleverd. De dynamische webpagina's worden op de webserver in statische HTML-webpagina's omgezet en zo naar de webclient doorgestuurd. Deze statische versie van een website kan in een webbrowser worden getoond zonder de koppeling met de webserver en het 'deep web'. Op die manier blijft de software-afhankelijkheid beperkt tot een webbrowser en eventueel de nodige plug-ins. Net zoals bij websites met een vaste inhoud worden er dus HTML-pagina's, afbeeldingen en stylesheets bewaard. Het enige verschil tussen de actieve en gearchiveerde website is dat interactieve functionaliteiten zoals het opvragen van gegevens uit een databank niet meer beschikbaar zijn. De statische HTML-webpagina's bevatten hoogstens nog clientscripts. Het archiveren van de originele dynamische webpagina's en de serverscriptbestanden heeft geen enkel nut voor de off line reconstructie van de website, maar kan wel belangrijk zijn vanwege de verantwoordingsplicht of hun informatieve waarde op zich. Dit zou inhouden dat er twee verschillende versies van de website worden gearchiveerd: een statische versie voor het raadplegen van de website en een versie met de originele bestanden zoals ze op de webserver stonden.

De inhoud van dynamische websites is bij de actieve versie van de website opgeslagen in databanken, een documentbeheerssysteem of een content management systeem. Deze gekoppelde bestandensystemen vormen het 'deep web'. Tijdens een snapshotoperatie is het zelden mogelijk om inhoud van het 'back-office'-systeem mee vast te leggen. Dit is wel mogelijk bij documenten die vanuit een documentbeheerssysteem beschikbaar worden gesteld, maar bij databanken die worden bevraagd, kan dit niet. Los van de vraag of de inhoud van het achterliggende informatiesysteem mee kan vastgelegd worden tijdens de snapshotoperatie is het zelden opportuun om dit te doen. Vanwege een aantal redenen (o.a. tijdsduur, meerdere keren archiveren van dezelfde informatie wanneer verschillende versies van dezelfde website worden gearchiveerd) is het echter beter om net zoals bij gewone databanken of content management systemen de informatie binnen het back-officesysteem bij te houden en hiervoor een afzonderlijke archiveringsstrategie toe te passen. Binnen een dergelijke 'deep web'-archivering zal versiebeheer in de meeste gevallen aangewezen zijn. Wanneer de inhoud

van een website vanuit een gekoppelde databank of web content managementsysteem wordt gevoed, zal men hier in het (FTP-) logbestand over het up- en downloaden van webpagina's geen sporen van terugvinden. In dit geval moeten er afzonderlijke logbestanden binnen de databank of het web content managementsysteem worden bijgehouden.

Wat een bezoeker uiteindelijk te zien krijgt, kan afhankelijk zijn van diverse factoren. In de meeste gevallen zal hij een de gekoppelde databanken op één of andere manier bevragen. Door de zoekopdracht te archiveren kan aangetoond worden welke informatie beschikbaar werd gesteld. De uitgevoerde HTTP-Get commando's met zoekopdrachten worden in de logbestanden van de webserver bewaard. Een webserver houdt in de logbestanden die hij standaard aanlegt onder meer volgende gegevens bij: bezoekers (IP-adres of domeinnaam), datum en tijd, bezochte pagina's, uitgevoerde acties, gebruikte webbrowser, enz. De logbestanden zijn niet in de eerste plaats bedoeld voor het vastleggen van een interactie of een transactie. Logbestanden worden doorgaans bijgehouden voor de noden van de webdesigners en -beheerders en zijn niet overzichtelijk of gemakkelijk te ontcijferen. Momenteel wordt volop werk gemaakt van de standaardisatie van een formaat voor logbestanden dat uitbreidbaar is volgens de noden van de instelling<sup>55</sup>. Men vergewist zich best op voorhand van de mogelijke archiefwaarde van logbestanden. Ten eerste is het belangrijk om de interactie met archiefwaarde te definiëren zodat de essentiële gegevens zeker in de logbestanden worden geregistreerd. Ten tweede moet er ook een archiveringsstrategie voor deze logbestanden worden vastgelegd, anders is de kans groot dat ze verloren gaan. Het bijhouden van de logbestanden met archiefwaarde vertrouwt men best niet aan een derde partij (bijv. een externe provider) toe. De logbestanden die door webadministrators standaard worden bijgehouden, zullen zelden volstaan.

Bij de recentste generatie websites kan de inhoud ook afhankelijk worden gemaakt van het gebruikersprofiel, de software waarover de bezoeker beschikt (webbrowser + versie), cookies, enz. Als deze gegevens van belang zijn, dan moeten ze ook samen met de website worden gearchiveerd. De gebruikte webbrowser wordt in de meeste webserverlogbestanden automatisch bijgehouden. Het archiveren van de cookies van de bezoekers kan een aantal moeilijkheden opleveren. De cookies worden in principe enkel op de harde schijf van de webclient bewaard. Er is wel een formaat voor logbestanden die de cookies als bijkomende informatie kan opnemen, maar hierdoor wordt het logbestand heel groot en de verwerking bijgevolg heel moeilijk.

Het afzonderlijk archiveren van de lagen waaruit een door een databank aangestuurde website is opgebouwd, heeft wel het nadeel dat een aantal oorspronkelijke functionaliteiten van de website verloren gaan. Raadpleging van of toevoeging aan een (documenten)databank via de website is niet meer mogelijk. Door de afzonderlijke onderdelen van elkaar te scheiden, gaat de integratie en functionaliteit verloren. Die integratie en functionaliteit is op software en logica gebaseerd en beide operationeel houden conflicteert met de doelstelling om zo systeemafhankelijk mogelijk te archiveren<sup>56</sup>.

---

<sup>55</sup> Zie hiervoor <http://www.w3.org/TR/WD-logfile>.

<sup>56</sup> In de Australische richtlijn voor het archiveren van websites is het operationeel houden van de functionaliteit een belangrijke vereiste. De overheden worden opgelegd om dynamisch gegenereerde on line bronnen in functionele staat te archiveren (*Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government*, p. 12 en 26). De richtlijn blijft wel vaag over de uitvoering.

## B.2.3 Besluit

Bij website-archivering komt meer kijken dan louter webpagina's in het digitaal depot opnemen. Voor elke website met archiefwaarde moet nagegaan worden welke bestanden in welke vorm worden gearcheveerd. Hierbij mag men zich niet beperken tot de website zelf: ook scriptbestanden, logbestanden, gebruikersprofielen en het 'diep web' kunnen voor archivering in aanmerking komen. Het gaat immers niet alleen om de reconstructie van de website zelf maar ook om de vastlegging van de nodige metadata en het naleven van een verantwoordings- en bewijsplicht.

**Tabel 1:** Samenvatting: Wat archiveren van websites ? De computerbestanden met archiefwaarde die niet voor een webbrowser bereikbaar zijn (cursief), moeten vanop de web- of applicatieserver worden gekopieerd.

| <b>LAGEN</b>        | <b>STATISCHE INHOUD</b>                                                                                                      |                                                         | <b>DYNAMISCHE INHOUD</b>                                                                                                                                         |                                   |
|---------------------|------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
|                     | INPUT<br>WEBSERVER                                                                                                           | OUTPUT<br>WEBBROWSER                                    | INPUT<br>WEBSERVER                                                                                                                                               | OUTPUT<br>WEBBROWSER              |
| <b>INTERFACE</b>    | HTML, XML<br>GIF, JPEG, TIFF, PNG<br>CSS, XSL                                                                                | HTML, XML<br>GIF, JPEG, TIFF, PNG<br>CSS, XSL           | HTML, ASP, PHP<br>GIF, JPEG, TIFF,<br>PNG, XML<br>CSS, XSL                                                                                                       | HTML, XML<br>GIF, JPEG, TIFF, PNG |
| <b>INTERACTIE</b>   | clientscripts ingebed<br>in HTML of aparte<br>bestanden                                                                      | clientscripts ingebed<br>in HTML of aparte<br>bestanden | client- en<br>serverscripts<br><i>uitvoerbare<br/>programma's</i>                                                                                                | clientscripts                     |
| <b>INHOUD</b>       | HTML, txt, PDF, doc,<br>rtf, xls, mdb, zip                                                                                   | HTML, txt, PDF, doc,<br>rtf, xls, mdb, zip              | HTML, txt, PDF, doc,<br>mdb, <i>databanken<br/>document- en inhoud<br/>beheerssystemen,<br/>logbestanden<br/>webserver en<br/>databanken</i>                     | HTML, txt, PDF, doc,<br>mdb       |
| <b>TRANSACTIE</b>   | HTML-formulieren<br><i>e-mails<br/>databanken<br/>logbestanden server</i>                                                    | HTML-formulieren                                        | HTML-formulieren<br><i>e-mails en<br/>databanken met<br/>formuliergegevens,<br/>logbestanden<br/>webserver</i>                                                   | HTML-formulieren                  |
| <b>DOCUMENTATIE</b> | <i>beschrijving van<br/>functie website<br/>binnen het<br/>werkproces, van het<br/>functioneren van de<br/>website, enz.</i> | /                                                       | <i>beschrijving van<br/>functie website<br/>binnen het<br/>werkproces, van het<br/>functioneren van de<br/>website, van de<br/>gekoppelde<br/>databank, enz.</i> | /                                 |

## C. HOE WEBSITES VASTLEGGEN VOOR ARCHIVERING ?

### C.1 Websites met een vaste inhoud

Voor websites met een statische inhoud wordt een mirror gearcheveerd. Een mirror is een identieke kopie van de bestanden in het bestandsformaat, met de bestandsnamen en in de structuur zoals ze op

de webserver staan. De gearcheverde website is raadpleegbaar voor iedereen die over de vereiste webbrowser beschikt.

Een eerste methode is het rechtstreeks kopiëren van de bestanden vanop de webserver. Hiervoor is toegang tot de harde schijf van de webserver vereist. Men kan dit doen door op de webserver zelf een kopie te maken van alle bestanden of door een FTP-programma te gebruiken. Voor het maken van een kopie vanop de webserver is de actieve medewerking van de archiefvormer vereist. De archiefvormer draagt de mirror aan de archivaris over (push: tape, CD<sup>57</sup>) of geeft de archivaris FTP-toegang tot de webserver. Met een FTP-programma kan men vanop elke clientcomputer de bestanden van een webserver lokaal plaatsen. Hiervoor zijn wel bijzondere toegangsrechten vereist. Deze methode zal bijna altijd voor gevolg hebben dat er teveel computerbestanden worden gearcheveerd. De harde schijven van websites zijn zelden een toonbeeld van efficiënt en rationeel bestandenbeheer. Verouderde en recente bestanden staan er meestal door elkaar en bij het maken van een kopie zullen dus ook bestanden waar geen links (meer) naar verwijzen worden gearcheveerd. Twee belangrijke voorwaarden voor de toepassing van deze methode zijn het gebruik van relatieve links voor de interne links en het overnemen van de mappenstructuur van de webserver<sup>58</sup>. Vooral bij de oudste statische websites werden ook absolute pathaanduidingen in de interne links gebruikt. Een automatische omzetting van interne absolute naar relatieve links zal zelden mogelijk zijn<sup>59</sup>. Als dit niet mogelijk is, dan zal de tweede weg meer aangewezen zijn.

Een tweede manier is dat de archivaris volledig zelfstandig te werk gaat en de nodige bestanden kopiëert met behulp van een off line browser (cfr. infra). De off line browser kan indien nodig absolute links naar relatieve links omzetten. De archivaris kan met behulp van een off line browser zelf een website archiveren (pull). In tegenstelling tot FTP-toegang kan een archivaris met een off line browser een on line website niet wijzigen. Een off line browser heeft wel enkel toegang tot de bestanden die voor webclients bereikbaar zijn. Wanneer geen gekoppelde databanken of logbestanden worden gearcheveerd, kan een off line browser dus volstaan.

Websites in Flash worden tot websites met een vaste inhoud gerekend. In principe kunnen beide archiveringsmethoden worden gebruikt, maar de archivering van enkele Flash-websites door het Stadsarchief Antwerpen wees echter uit dat het maken van een mirror met behulp van een off line browser niet altijd even succesvol verliep. Bij een aantal websites was het kopiëren van de bestanden vanop de webserver de gemakkelijkste en snelste oplossing. Overigens is bij Flash-websites de kans groot dat men voor archivering toch op de archiefvormer is aangewezen. Absolute interne links kunnen immers niet gemakkelijk naar relatieve worden omgezet. Huidige off line browsers hebben geen toegang tot de binaire \*.swf-bestanden. Handmatig de links aanpassen is veelal onmogelijk, want bij het publiceren van \*.fla-bestanden als een \*.swf-bestand worden deze laatste doorgaans tegen

---

<sup>57</sup> Het gebruik van een CD kan wel een aantal moeilijkheden inhouden. Uitwisselbare CD's voldoen best aan de ISO-9660. Deze standaard houdt beperkingen in inzake het aantal karakters in bestands- en mapnamen. Verkorten van de bestandsnamen betekent ook dat alle links aanpassing behoeven. Het gebruiken van uitbreidingen op ISO-9660 (Joliet, Rock Ridge) of CD's niet als lange termijndrager gebruiken, kan een oplossing bieden.

<sup>58</sup> Vanwege de noodzaak om identiek dezelfde mappenstructuur van de webserver over te nemen, zou men kunnen overwegen om een statische website naar een TAR-bestand om te zetten en zo over te dragen. TAR-bestanden zijn wel besturingssysteemafhankelijk en moeten na neerlegging onmiddellijk opnieuw worden omgezet. Bij extractie wordt de mappenstructuur en de afzonderlijke bestanden hersteld.

<sup>59</sup> Niet werkende interne absolute links kunnen wel automatisch worden opgespoord. Het omzetten naar relatieve links echter is handenarbeid. Een optie zou kunnen zijn om de op CD neergelegde website tijdelijk terug op een webserver te plaatsen en vervolgens met een off line browser te kopiëren.

veranderingen of importeeracties beveiligd. De archiefvormer dient bijgevolg zelf interne absolute linken aan te passen of de archivaris het paswoord door te geven en eventueel de originele \*.fla-bestanden te bezorgen.

On line webbrowsers (IE Explorer, Netscape, Opera, enz.) kunnen wel gebruikt worden om één webpagina off line te plaatsen (met de opdracht 'opslaan als ...'), maar voldoen niet om een volledige website te archiveren<sup>60</sup>. Backups van de website volstaan evenmin als archiveringswijze. Backupprocedures maken deel uit van de hedendaagse beheers- en veiligheidsprocedures voor websites, maar zijn niet bruikbaar als archiveringssysteem<sup>61</sup>.

## C.2 Websites met een dynamische inhoud

### C.2.1 Snapshots

Om interactieve websites zodanig te archiveren dat ze zonder hun oorspronkelijk webserverconfiguratie en webserversoftware raadpleegbaar zijn, kan er niet enkel een kopie van de originele bestanden worden gearchiveerd. In de plaats van de originele ASP-, PHP- of JSP-bestanden worden statische HTML-bestanden gearchiveerd. Voor het vastleggen van interactieve websites in de vorm van HTML-bestanden worden hiervoor bijzonder ontworpen computerprogramma's gebruikt: off line browsers.

Off line browsers zijn computerprogramma's voor het off line plaatsen van op voorhand geselecteerde websites. Voor het vastleggen van een website wordt een afzonderlijk project geopend. Binnen elk project kunnen de instellingen worden gedefinieerd (startURL, diepte van de linken, uitsluitingen, enz.). Off line browsers waren initieel ontworpen om langdurig te raadplegen websites lokaal te plaatsen zodat men deze in de toekomst zonder bijkomende telefoonkosten kon raadplegen. De eerste generatie off line browsers dienden enkel om een kopie te maken van de bestanden van de website zoals ze op de webserver stonden. De websites van toen waren immers overwegend statische websites. Ondertussen hebben de jongste versies van de off line browsers zich aangepast aan de evolutie van websites en beschikken ze over de mogelijkheid om dynamisch samengestelde webpagina's als statische HTML-pagina's te bewaren. De oorspronkelijke extensies zoals ASP, PHP en JSP worden in HTML omgezet zodat de pagina's off line op om het even welke computer met een webbrowser kunnen worden bekeken<sup>62</sup>. Een bestand met de naam 'default.asp' wordt dan als

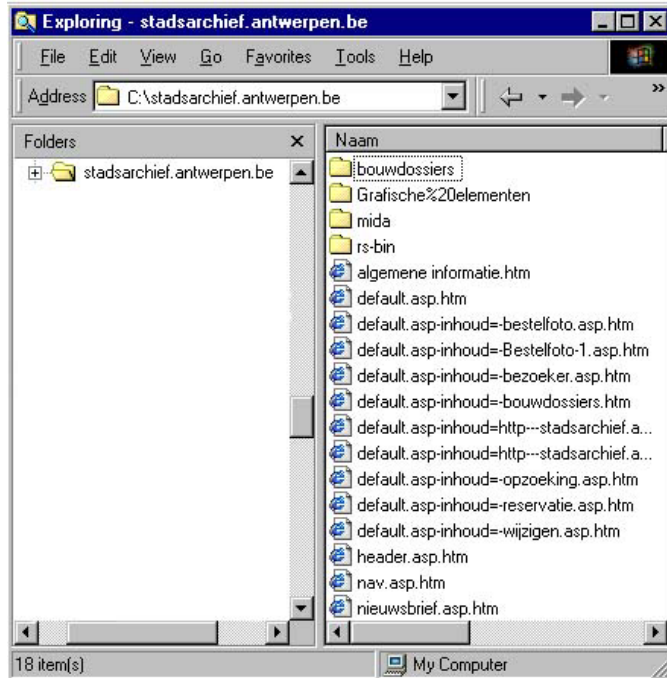
---

<sup>60</sup> Enkel de HTML-pagina's die in het actieve venster zijn geopend worden lokaal opgeslagen. Hierdoor wordt slechts één webpagina tegelijk off line geplaatst. Voor het vastleggen van een website die uit meerdere pagina's bestaat moet elke pagina afzonderlijk off line worden geplaatst. Soms worden enkel de frameset bewaard, en niet de webpagina's met inhoud. Er wordt niet aan de vooropgestelde kwaliteitsvereisten voldaan: interne links zijn absoluut, gemeenschappelijke afbeeldingen worden iedere keer opnieuw bewaard, de tweede laag van roll-over afbeeldingen worden niet mee opgeslagen, het verband tussen de webpagina's van een website gaan verloren, en de originele bestandenstructuur en bestandsnamen worden niet overgenomen.

<sup>61</sup> De ervaringen met de eerste versies van de website van de stad Antwerpen illustreren dit nadrukkelijk. De eerste twee versies stonden nog op backuptape. De archivering van deze websites was een kritische onderneming die gelukkig succesvol kon worden afgerond. De derde versie kon helemaal niet gerecupereerd worden. (F. BOUDREZ, *Van backup tot gearchiveerde website. De archivering van de eerste versies van de Digitale Metropool Antwerpen*, Antwerpen, 2002).

<sup>62</sup> ASP-, PHP- of JSP-bestanden kunnen niet in een gewone webbrowser off line worden bekeken. Hiervoor is bijzondere software vereist zoals MS Personal Web Server. Dit verhoogt opnieuw de software-afhankelijkheid en deze oplossing dient helemaal niet voor een grote hoeveelheid gearchiveerde websites.

'default.asp.htm' of als 'default.htm' opgeslagen. Een aantal programma's geven de bestanden een nieuwe bestandsnaam: de oorspronkelijke bestandsnaam wordt vervangen door de URL die in de adresbalk van een on line browser verschijnt. Statische HTML-pagina's op de webserver ondergaan geen wijzigingen en worden gewoon in hun originele vorm off line als HTML-bestanden geplaatst.



**Afbeelding 7:** De website van het stadsarchief Antwerpen is een website met veel webpagina's die pas op het ogenblik van het ontvangen van een HTTP-request worden samengesteld. De dynamische webpagina's zijn gekoppeld aan een documentbeheersysteem en geven toegang tot documenten die voor het publiek beschikbaar zijn. Deze pagina's worden als statische HTML-pagina's gearcheveerd. Bij het archiveren van de website werden alle ASP- bestanden als statische HTML-pagina's opgeslagen. De originele bestandsnaam werd behouden maar kreeg de extensie .htm. De gearcheveerde webpagina's tonen welke bestanden op het moment van de snapshot beschikbaar waren. De documenten zelf werden niet in de snapshot opgenomen, maar blijven binnen het documentbeheersysteem bewaard.

Met een harvester of een off line browser wordt via het netwerk een snapshot genomen. Het uitvoeren van een dergelijke operatie verloopt niet altijd even gemakkelijk en gaat soms met moeilijkheden gepaard. Er zijn een aantal problemen die mits de goede keuze van off line browser kunnen opgelost worden, maar er zijn ook een aantal structurele problemen:

- ➡ het is moeilijk om de grenzen van een website vast te leggen. De meeste programma's bieden de mogelijkheid om een snapshot te beperken tot alle bestanden die zich binnen dezelfde startURL bevinden, maar mappen die er buiten staan worden dan niet mee vastgelegd. Een virtuele map met bijvoorbeeld een gemeenschappelijke banner voor alle websites van dezelfde instelling wordt niet mee opgenomen in de snapshot wanneer als startURL de URL van één instelling wordt opgegeven. Een ander probleem zijn de 'redirects' die in veel websites worden gebruikt. De meeste programma's vragen ook een opgave van het aantal niveaus (diepte van de linken) die moeten worden vastgelegd. Vooraleer men een exact aantal opgeeft, moet men op voorhand op de hoogte zijn van het aantal niveaus, anders wordt een deel van de website niet mee gearcheveerd. In de meeste gevallen zullen er meer bestanden worden gearcheveerd dan noodzakelijk. Ook bestanden die niet tot de website behoren worden mee off line geplaatst. Die overbodige bestanden verwijdert men best achteraf.
- ➡ niet alle websites kunnen door middel van een off line browser of harvester worden gearcheveerd. Deze manier van werken is beperkt tot publiek toegankelijke (delen van) websites. Intranetsites vormen of de afgeschermdede delen van een website, kunnen niet off

- line worden geplaatst, tenzij de archivaris over toegangsrechten beschikt.
- ➡ enkel de websites worden gearcheveerd. Serverlogbestanden, gekoppelde databanken ('deep web') en logbestanden over het up- en downloaden van de website kunnen niet mee worden vastgelegd.
  - ➡ er kunnen enkel snapshots worden genomen van de websites of de bestanden die actief zijn. Men beschikt enkel over een snapshot genomen op bepaalde tijdstippen. De versies van websites die meerdere malen zijn gewijzigd tussen twee snapshots in, worden niet gearcheveerd.
  - ➡ de tweede laag van roll-overafbeeldingen, server-sided image maps, DTD's en XSL-stylesheets worden niet altijd vastgelegd. Websites die een toepassing zijn van Flash kunnen soms moeilijk met de huidige off line browser lokaal worden geplaatst.
  - ➡ virtuele directories: een onderdeel van een website kan in een virtuele directory worden opgeslagen. De meeste off line browsers zullen de bestanden in de virtuele directory niet mee archiveren wanneer een andere servernaam in het webadres wordt gebruikt. Een ander probleem is de absolute pathaanduiding waarmee de bestanden in de virtuele directory worden bereikt.
  - ➡ het nemen van een snapshot is tijdrovend en neemt voor een grote website al vlog een paar uren in beslag. Dit levert vooral problemen op bij heel dynamische sites (bijv. van een krant die voortdurend wordt aangepast). De kans is dan groot dat tijdens de vastleggingsoperatie de website wordt aangepast en dat men een versie archiveert die in de praktijk nooit bestaan heeft. De eerste en laatste vastgelegde webpagina kunnen tot verschillende versies behoren<sup>63</sup>.
  - ➡ het nemen van snapshots gaat in veel gevallen met fouten gepaard: hyperlinks werken niet, bestanden zijn niet beschikbaar, overbelasting van de server, time-out van de server, scripts veroorzaken eindeloze loops, ... . Als de site wordt aangepast terwijl er een snapshot wordt genomen, is de kans groot dat er een onherstelbare fout optreedt<sup>64</sup>.

De nadelen van het gebruik van een off line browser zijn dus legio, maar bij een selectieve benadering is er momenteel geen alternatief om "on the fly" gegenereerde webpagina's vast te leggen. De nadelen kunnen opgevangen worden door ten eerste een goede off line browser te kiezen en ten tweede de fouten in een snapshot te verbeteren.

De keuze van off line browser is heel belangrijk. Op het internet is een heel gamma off line browsers als freeware of shareware beschikbaar<sup>65</sup>. Een goede off line browser biedt minstens volgende mogelijkheden:

- ☒ keuze tussen het maken van een zo identiek mogelijke kopie (voor websites met een statische inhoud), een reconstrueerbaar snapshot (voor websites met een dynamische

<sup>63</sup> Een bekend voorbeeld hiervan is de aankondiging van een gebeurtenis in de toekomst op de frontpagina van een digitale krant, terwijl verder in de gearcheveerde website al een verslag staat.

<sup>64</sup> J. HAKALA, *Collecting and Preserving the Web: Developing and Testing the NEDLIB Harvester*, in *RLG-DigiNews*, April 15, 2001, vol.5, nr. 2.

<sup>65</sup> Een overzicht van off line browsers wordt bijgehouden op: <http://www.tucows.nl>. Zie ook <http://www.davecentral.com/browse/67/> voor een overzicht en korte besprekingen.

- inhoud) en het off line plaatsen van één specifieke webpagina<sup>66</sup>.
- ☑ beperking van het off line plaatsen van bestanden tot het opgegeven startdomein. Off line browsers volgen de links om de volgende bestanden te kopiëren. Beperkt men de browser niet tot het opgegeven domein dan worden ook de externe links gevolgd en worden er andere websites gekopieerd.
  - ☑ volgen correct de re-directs binnen een website.
  - ☑ omzetting van de absolute pathaanduidingen voor links binnen de eigen website naar relatieve pathaanduidingen.
  - ☑ kopiëren de originele bestandenstructuur van de webserver.
  - ☑ laten selectie toe van het type bestanden die off line worden geplaatst. Bij de instellingen van de off line browsers kan gedetailleerd worden opgegeven welke type bestanden worden gekopieerd en welke niet. Zo kunnen bijvoorbeeld PDF-bestanden en pagina's met uitvoerbare codes achterwege gelaten worden, wanneer deze afzonderlijk worden gearhiveerd.
  - ☑ een aantal off line browsers voegen in de vorm van HTML-commentaar de naam van het gebruikte browserprogramma aan het HTML-script toe. Als er commentaar wordt toegevoegd, dan zijn dat bij voorkeur gegevens over de snapshotoperatie (metadata: datum, tijdstip, titel website, ...).
  - ☑ voorzien in een versie- en duplicatiecontrole zodat indien nodig de actieve bestanden met de reeds gekopieerde bestanden worden vergeleken. De meeste programma's laten toe dat alleen de pagina's gewijzigd na een bepaalde datum worden gearhiveerd.
  - ☑ rapporteren over de fouten bij het nemen van een snapshot. Het nemen van een snapshot gaat soms met fouten gepaard. De fouten worden in een logbestand opgeslagen.
  - ☑ kunnen zich bij webserveren aanmelden als verschillende on line browsers. Een aantal websites zijn beveiligd en weren off line browsers of robots. Aan de website kan een browsercheck verbonden zijn, waardoor het wenselijk is zich als een welbepaalde browser aan te melden.
  - ☑ spreiden de verzoeken aan de webserver en laten de gebruiker toe om het aantal bestanden op te geven die gelijktijdig gevraagd worden. Om overbelasting en robots exclusion te vermijden, wordt het aantal bestanden die op hetzelfde ogenblik worden aangevraagd best beperkt.

Het gebruik van een off line browser die aan deze vereisten voldoet, is op zich nog geen voldoende garantie voor een kwaliteitsvolle archivering. Achtereenvolgende snapshotoperaties met dezelfde off

---

<sup>66</sup> Met een off line browser kan geen 100 % identieke kopie van dynamische webpagina's worden gemaakt. Een off line browser ontvangt immers alleen maar HTML-bestanden, en bijvoorbeeld geen ASP- of PHP-bestanden. Deze laatste bestanden kunnen bij het off line plaatsen wel hun oorspronkelijke extensie behouden, maar zijn in wezen de HTML-versies van het oorspronkelijk ASP-, PHP- of JSP-bestand. De bestanden bevatten immers geen scripting meer.

line browser kunnen soms verschillende resultaten opleveren. Factoren zoals overbelasting van de server en het tijdstip spelen een rol.

Snapshots moeten gecontroleerd worden op fouten en anomalieën. Voor deze kwaliteitscontrole kan men een beroep doen op computerprogramma's. Een goede off line browser of harvester zal zelf een rapport geven van het totaal aantal vastgelegde bestanden en de fouten die optraden. Een dergelijke logbestand is al een belangrijke indicator, maar alleen hierop mag men niet betrouwen. Er bestaan specifieke programma's en on line diensten voor de controle van websites en deze kunnen ook gebruikt worden voor een kwaliteitscontrole van gearcheeerde of te archiveren websites<sup>67</sup>. In hun controle kunnen deze bijzondere programma's heel ver gaan: de geldigheid van de interne en externe links, de bestandsnamen, de HTML-syntax en geldige attributen, de aanwezigheid van alle nodige computerbestanden, de compatibiliteit met bepaalde webbrowsers, de samenstelling van formulieren, ongedefinieerde ankers, ...<sup>68</sup>. Vooral het verder functioneren van de links is heel belangrijk. Op basis van de links wordt de website gereconstrueerd, wordt de band tussen twee computerbestanden vastgelegd, en krijgt men toegang tot de overeenstemmende computerbestanden. Zonder correcte links is de mirror of het snapshot van de website niet functioneel. De verbetering van foute links is dan ook prioritair. Met deze controlerende tools kunnen de courant voorkomende problemen worden opgespoord en misschien zelfs opgelost. E-mailadressen, formulieren en externe links kunnen desgewenst inactief gemaakt worden. Indien nodig verwijzen deze onderdelen van de gearcheeerde website naar de nog actieve on line versie. Een laatste controle wordt best door de archivaris uitgevoerd.

Een aantal problemen kunnen enkel handmatig worden opgelost. Bij de archivering van versie 6 van de David-website werd het snapshot op vier punten gecorrigeerd: 1. de tweede laag van de roll-overafbeeldingen in het navigatiepaneel werd handmatig naar de overeenstemmende map gekopieerd, 2. de DTD's en XSL-stylesheets van de voorbeelden van het gearcheeerde kiezersregister en e-mail werden bij de XML-bestanden geplaatst, 3. de verandering van de mapnaam 'Grafische elementen' door de off line browser in 'Grafische%20elementen' werd ongedaan gemaakt zodat de links weer werkten, 4. de teller op de startpagina werd uitgeschakeld: de broncode voor de teller bleef op zijn plaats staan, maar werd als HTML-commentaar gedefinieerd.

Het vastleggen van de originele ASP- of PHP-bestanden is enkel mogelijk wanneer men toegang heeft tot de webserverschijf (kopie, FTP-toegang).

#### C.2.2 Logbestanden

De logbestanden worden bijgehouden als platte tekstbestanden of in een databank op de web/fileserver en worden als zodanig gearcheeerd. Deze digitale archiefdocumenten worden eveneens in het archiveringssysteem opgenomen.

---

<sup>67</sup> De meeste HTML-editors beschikken over een linkchecker. Voorbeeld on line diensten zijn: <http://validator.w3.org/checklink>; <http://www.cast.org/bobby/>

<sup>68</sup> Een voor de hand liggende foutmelding zullen afgebroken externe links zijn wanneer er geen Internet-verbinding is of wanneer downloads niet samen met de gearcheeerde website worden bewaard.

### C.2.3 Databanken

Archivering van gekoppelde (document)databanken is een problematiek op zich die hier niet uitvoerig kan behandeld worden. Men kan de versies van een databank archiveren door in een object-georiënteerde databank een historiek van elk object aan te leggen, snapshots van de (relationele) databank te nemen of in logbestanden de wijzigingen ten aanzien van de initiële versie bij te houden. Voor (document)databanken met een permanentie archiefwaarde moet eveneens voldaan worden aan de vereisten inzake systeemafhankelijke archivering. Gewone backups komen hiervoor niet in aanmerking.

## D. FREQUENTIE

### D.1. Bepalen van de frequentie

Websites hebben als voordeel dat ze heel snel aanpasbaar zijn. De verschillende versies van een website kunnen elkaar heel snel opvolgen. Bij de archivering van websites moet men hier rekening mee houden. De gemiddelde levensduur van een webpagina wordt op 75 à 100 dagen geschat<sup>69</sup>, maar met dit gegeven kan men uiteindelijk maar weinig aanvangen. Algemeen kan men wel stellen dat een statische website minder wijzigingen ondergaat dan een dynamische website.

Bij het bepalen van de frequentie waarmee websites worden gearcheveerd, is de grote moeilijkheid dat veranderingen aan websites eerder onregelmatig gebeuren, dan met een vaste regelmaat. Voor elke website moet uiteindelijk afzonderlijk worden onderzocht met welke frequentie een versie van de website wordt gearcheveerd. Belangrijke indicatoren zijn de aard, het doel en natuurlijk de frequentie van aanpassingen aan de website. Het archiveren van een dagelijkse of misschien zelfs tweedagelijkse versie van de website van een presidentskandidaat is voor later historisch onderzoek naar de gevoerde campagne geen overbodige luxe. De website van een politieke partij zal in de periode voor en na de verkiezingen meer wijzigingen ondergaan dan anders. De website van een krant wijzigt dagelijks of zelfs meerdere keren per dag. Vanwege de documentaire of historische waarde kan de archivering met een hoge frequentie nodig zijn. Dit zal zeker het geval zijn wanneer websites ingeschakeld zijn in een bedrijfsproces waarover men verantwoording moet afleggen.

De frequentie wordt niet alleen mee bepaald door de website zelf maar ook door het acquisitiebeleid- en profiel van de archiverende instelling. De frequentie evolueert uiteindelijk van het occasioneel archiveren van momentopnames tot het archiveren van elke versie. Is de archivering van een website omwille van de verantwoordingsplicht vereist, dan moet in ieder geval elke versie worden bewaard of reconstrueerbaar zijn.

---

<sup>69</sup> D. SHENK, *The world wide library*, in *Hotwired*, 2 sept. 1997 ([http://hotwired.lycos.com/synapse/feature/97/35/shenk1a\\_text.html](http://hotwired.lycos.com/synapse/feature/97/35/shenk1a_text.html)); L. DEMBART, *Go Wayback. Internet Archive stores pages long gone from Web*, in *International Herald Tribune*, p. 13.

## D.2. Hoe versies archiveren?

Als vertrekpunt archiveert men best de volledige website. Bij het archiveren van de wijzigingen heeft men in principe de keuze: men kan zich beperken tot het archiveren van de wijziging of men kan opnieuw de volledige website archiveren. In het eerste geval moet men heel goed weten welke bestanden al dan niet werden gewijzigd. Off line browsers kunnen wel achterhalen welke bestanden in vergelijking met een vorige versie werden gewijzigd, maar men moet oppassen dat oudere bestandsversies niet met recentere worden overschreven. Men moet ook afwegen of het nemen van een volledig nieuwe mirror of nieuw snapshot niet efficiënter is. De voornaamste afweging parameters bij deze afweging zijn beschikbare opslagcapaciteit en arbeidsuren. Aangezien de bestandsomvang doorgaans geen onoverkomelijk probleem is, zal wellicht in de meeste gevallen voor deze optie worden gekozen. Bij de archivering van momentopnames met een lage frequentie is het aangewezen om de volledige website te archiveren.

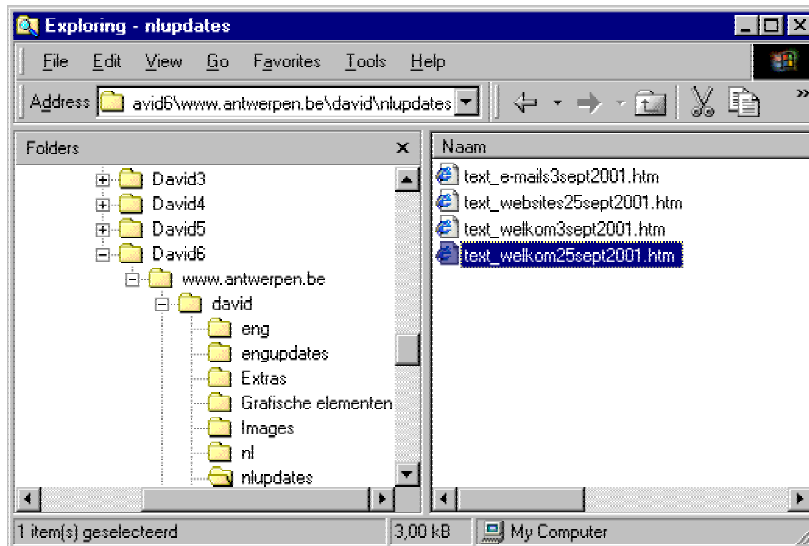
Het bijhouden van alle versies van de website is onmogelijk zonder de actieve inbreng van de archiefvormer. Overigens is het in veel gevallen ook zijn taak. De archiefvormer weet wanneer de website wijzigingen ondergaat. Technisch gezien is het mogelijk dat men vanop afstand met een off line browser, een harvester of een spider alle versies archiveert, maar dit is zeer belastend voor het netwerk en de kans is reëel dat informatie verloren gaat. Het is veel efficiënter dat de archiefvormer zelf een historiek bijhoudt of de archivaris op wijzigingen attendeert. De archiefdienst van een openbare instelling zal dit wellicht gemakkelijker kunnen bereiken voor het archiveren van de eigen website(s) dan de archiefinstellingen en het documentatiecentra die websites van andere instellingen of verenigingen archiveren. Het Nederlands documentatiecentrum voor politieke partijen kan voor de archivering van de websites van politieke partijen niet op de actieve medewerking van de partijen rekenen en moet dus op eigen initiatief mirrors maken of snapshots nemen<sup>70</sup>.

### D.2.1 De wijzigingen aan websites met een vaste inhoud

Bij websites met een vaste inhoud volstaat het om de gewijzigde originele bestanden in het digitaal archiveringssysteem op te nemen. Men kan deze toevoegen aan de map waarin de volledige versie wordt bijgehouden. Om te vermijden dat de nieuwe versies oude bestanden overschrijven, moet men zorgen voor een unieke bestandsnaam of de bestanden in een aparte map plaatsen.

---

<sup>70</sup> Mededeling Gerrit Voerman, Documentatiecentrum Nederlandse Politieke Partijen



Afbeelding 8: Het afzonderlijk archiveren van de updates aan versie 6 van de DAVID website. De map David 6 bevat een snapshot van de volledige website die werd genomen onmiddellijk nadat de nieuwe versie on line werd geplaatst (3 juli). De bestanden die sindsdien werden gewijzigd, worden in de mappen 'nlupdates' en 'engupdates' geplaatst. Aan de originele bestandsnaam wordt de datum toegevoegd waarop de updates on line werden geplaatst. De mappen 'nlupdates' en 'engupdates' staan op hetzelfde niveau als de mappen 'nl' en 'eng' zodat de linken met de afbeeldingen in de map 'Images' blijven werken en dezelfde afbeelding geen tweemaal wordt gearcheveerd. Aangezien de updates door de datum toch een unieke bestandsnaam hebben, zou ook kunnen overwogen worden om ze in de map met de originele bestanden te plaatsen.

#### D.2.2 De wijzigingen aan websites met een dynamische inhoud

Een website die heel frequent wordt gewijzigd, wordt doorgaans op een dynamische wijze samengesteld. Het aanpassen van statische webpagina's is veel te tijdrovend voor dagelijkse of wekelijkse aanpassingen. De inhoud wordt aangeleverd vanuit een 'back-office'-informatiesysteem. Bij het archiveren van verschillende versies moet ook nagedacht worden aan een versiebeheer en een archivering van dit 'back-office'-informatiesysteem. In de praktijk zal het meestal gaan om de toepassing van één of andere vorm van databankarchivering.

Rekening houdende met de snelheid waarmee dynamische websites worden aangepast, lijkt het inbouwen van een automatisch versie- of historiekbeheer in het 'back-office'-informatiesysteem de meest interessante optie. Op die manier kan de arbeidsintensiviteit worden beperkt. Een dergelijke module kan zelden achteraf in een bestaand systeem worden ingebouwd, zodat het opnieuw belangrijk is dat men bij het ontwerp al de nodige stappen onderneemt. Vandaag de dag wordt hier veel aandacht aan besteed. Het bijhouden van de wijzigingen aan een website is overigens één van de basisprincipes van *web content management*.

De website zelf zal maar gearcheveerd moeten worden wanneer er iets aan de website als interface wordt gewijzigd. Dit kan bijvoorbeeld een verandering aan de afbeeldingen, schermopbouw of stylesheets zijn.

## E. DIGITALE DUURZAAMHEID

Voor het verzekeren van de leesbaarheid en raadpleegbaarheid van gearcheverde digitale documenten kunnen in principe verschillende strategieën worden toegepast<sup>71</sup>. De duurzame digitale archivering van websites toont aan dat deze strategieën elkaar niet uitsluiten, maar naast elkaar en zelfs in combinatie met elkaar kunnen worden gebruikt.

Websites zijn een verzameling gelinkte computerbestanden in diverse bestandsformaten. Om de digitale duurzaamheid op lange termijn te verzekeren, moeten de HTML-tags en attributen correct worden geïnterpreteerd, de scripts uitvoerbaar blijven en de gekoppelde bestanden (afbeeldingen, animatie, documenten) leesbaar blijven. Dit zal in de eerste plaats afhangen van de compatibiliteit met de gangbare webbrowsers en de plug-ins<sup>72</sup>.

Uit de Zweedse en Finse harvestervaring van hun respectieve webspaces blijkt dat er na een harvest meer dan 440 verschillende bestandsformaten in het website-archief zijn opgenomen. Dit cijfer geeft echter een vertekend beeld. In de praktijk is er in grote mate standaardisatie op het vlak van gebruikte bestandsformaten voor websites. Immers, webdesigners houden rekening met het feit dat zoveel mogelijk gebruikers de website vanop de meest uiteenlopende platformen moeten kunnen consulteren. Dit kan enkel bereikt worden door bestandsformaten met de status van standaard te gebruiken. De courant gebruikte bestandsformaten waaruit websites bestaan (zie tabel 1) staan allemaal bovenaan in de hiërarchie van bestandsformaten (zie afbeelding 9). De niet-gestandaardiseerde bestandsformaten komen dus maar een zeldzaam aantal keren voor. Deze bestanden worden best omgezet naar een gestandaardiseerd formaat dat zo goed mogelijk de originele eigenschappen overneemt. Onder de niet-gestandaardiseerde bestanden zitten ook veel bestanden die op basis van hun extensie of hun MIME-type niet herkenbaar zijn.

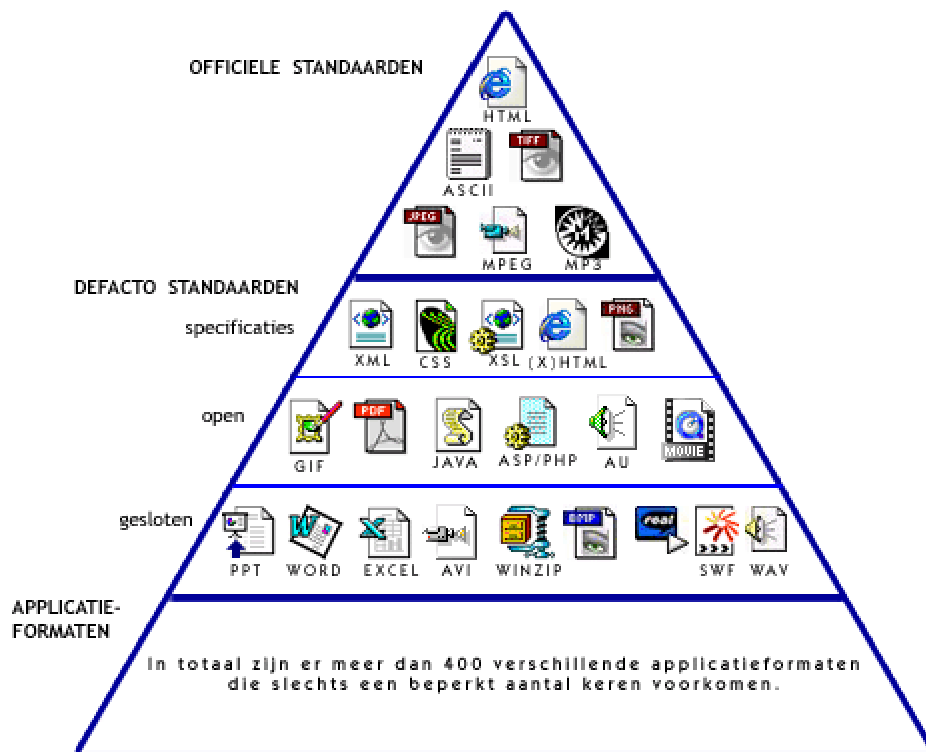
**Tabel 2:** De meest courante bestandsformaten waaruit gearcheverde websites bestaan

| TYPE                | BESTANDSFORMATEN                                                      |
|---------------------|-----------------------------------------------------------------------|
| <b>HYPERTEKST</b>   | htm/html, shtm/shtml, xhtml, stm/stml, xml,                           |
| <b>AFBEELDINGEN</b> | gif, tif/tiff, jpeg/jpg, bmp, png, xbm, fif                           |
| <b>GELUID</b>       | au, wav, mid/midi, mp2/3, riff,                                       |
| <b>VIDEO</b>        | mpg/mpeg, avi, ani, qt, mov, fli, flc                                 |
| <b>ANIMATIE</b>     | animated gif, swf,                                                    |
| <b>DOCUMENTEN</b>   | doc, pdf, txt/text, wrl, xml, xls, xlw                                |
| <b>DOWNLOADS</b>    | zip, rar, arj                                                         |
| <b>SCRIPTS</b>      | ASP, JAVAservlets, -applets, -script (.js, .class), PHP, VB, cfm, jsp |

<sup>71</sup> De algemene archiveringsstrategieën worden beschreven in: *Het digitale archiveringssysteem: beheersinventaris, informatielagen en beslissingsmodel als uitgangspunt*, p. 7-11.

<sup>72</sup> De bibliotheekwereld is druk bezig met het vastleggen van website-archieven, maar heeft nog geen sluitende oplossing voor het verzekeren van de digitale duurzaamheid. Men is zich bewust van de problematiek, maar men hinkt nog op verschillende gedachten. Feitelijk komt het erop neer dat men denkt alle digitale archiveringsstrategieën toe te passen: bewaren van originele hard- en software, migratie en emulatie.

Afbeelding 9: Hiërarchie van de frequent voorkomende bestandsformaten.



Het bewaren van de webpagina's met een dynamische inhoud als statische HTML-bestanden vergemakkelijkt de zaken nog. Hierdoor worden er geen ASP- of PHP-bestanden maar HTML-pagina's in het archief opgenomen en vermijdt men platformgebondenheid. Het raadplegen van deze originele bestanden zal toch in een gewone teksteditor gebeuren. Uit de Zweedse en Finse harvestoperaties blijkt dat de HTML-, GIF- en JPEG-bestanden alleen al meer dan 95 % van het totaal aantal bestanden vormen.

Tabel 3: Aandeel van de bestandsformaten in aantal en in bestandsomvang in het Zweedse websitearchief (2000)<sup>73</sup>.

| TYPE         | AANTAL | OMVANG |
|--------------|--------|--------|
| HTML         | 52 %   | 13,9 % |
| GIF          | 24 %   | 6,8 %  |
| JPEG         | 20 %   | 15,9 % |
| TIFF         |        | 3,5 %  |
| PNG          | 0,3 %  |        |
| PDF          | 1,3 %  | 10,3 % |
| ASCII        | 2 %    | 9,8 %  |
| Octet-stream | 0,9 %  | 11,5 % |
| Postscript   | 0,3 %  | 3,8 %  |

<sup>73</sup> K. PERSSON, *The Kulturarw3 Project - The Swedish Royal Web Archiw3e*, Lezing gehouden in Svetlogorsk, aug. 2000. <http://kulturarw3.kb.se/html/statistik.html>. Het aandeel van de HTML-, GIF- en JPEG-bestanden bedraagt in deze tabel 94 %. In latere harvestoperaties is het aandeel van deze bestanden toegenomen tot ca. 97 à 98 %. De statistieken van de archivering van de Finse en Oostenrijkse webspace bevestigen deze cijfers. (A. ARVIDSON, *Harvesting the Swedish webspace*; J. HAKALA, *Harvesting the Finnish Web space - practical experiences*; A. ASCHENBRENNER, *Long-Term Preservation of Digital Material*, p. 76).

|                   |        |       |
|-------------------|--------|-------|
| <b>MSWord</b>     | 0,3 %  |       |
| <b>Real audio</b> | 0,2 %  |       |
| <b>WAV</b>        | 0,07 % |       |
| <b>Zip</b>        | 0,4 %  | 6,5 % |
| <b>MPEG</b>       |        | 2 %   |

Of de gearchiveerde websites in de toekomst in hun huidige formaat nog geraadpleegd kunnen worden, staat en valt met de webbrowsers waarover men beschikt<sup>74</sup>. De jongste generatie webbrowsers zijn achterwaarts HTML-compatibel, ondersteunen XML, geven afbeeldingen weer en voeren clientscripts uit. HyperText Markup Language is voortdurend in ontwikkeling. HTML is een markuptaal waarmee de presentatie van een webpagina wordt gedefinieerd. HTML-bestanden bevatten inhoud en opmaakgegevens. De standaarden voor de HTML-taal worden gedefinieerd door het W3C aan het MIT Laboratory for Computer Science. HTML is de standaard voor het publiceren van hypertext op het internet. Van HTML bestaan verschillende versies. De eerste versie werd verspreid vanaf maart 1993. Weidverspreide versies zijn de 2.0 (1994), 3.2 (1996), 4.0 (1997), en 4.01 (1999, ISO 15445:2000). De recentste versies van HTML zijn Dynamic HTML (DHTML) en XHTML 1.0. XHTML is een herformulering van HTML in XML<sup>75</sup>. DHTML is de marketingterm voor een mengeling van HTML, stylesheets, DOM en scripting. DHTML is niet formeel gestandaardiseerd.

De jongere HTML-versies zijn overwegend compatibel met de voorgaande versies. Bij elke versieverbetering van HTML worden echter ook nieuwe tags en attributen geïntroduceerd en een aantal bestaande tags of attributen als verouderd bestempeld of zelfs verwijderd<sup>76</sup>. De ontwikkeling van de stylesheets zit hier voor veel tussen (CSS, XSL). De HTML-syntax werd meer en meer op het gebruik van stylesheets afgestemd. HTML en stylesheets werden complementair zodat het principe van scheiding van markup en lay-out beter wordt bereikt. De webbrowsers worden voortdurend aan de HTML-en stylesheetevolutie aangepast, en omgekeerd. In de meeste gevallen zijn de nieuwe HTML-versies een verfijning en uitbreiding van de vroegere versies en kunnen de webbrowsers HTML-pagina's in een oudere versie nog steeds inlezen. De kans bestaat echter dat een aantal tags of attributen niet meer worden uitgevoerd of problemen veroorzaken. Dit probleem doet zich ook voor wanneer HTML-pagina niet gestandaardiseerde, en dus in bepaalde gevallen niet ondersteunde, HTML-tags bevat. Dit is bijvoorbeeld het geval wanneer een HTML-editor code genereert die enkel door de browser van dezelfde producent wordt omgezet (bijv. Netscape: <CENTER>, Internet Explorer:<MARQUEE>). Deze niet-ondersteunde of niet-gestandaardiseerde HTML-tags zijn op termijn een probleem voor de leesbaarheid van de websites. Men zou dit kunnen verhelpen door de

<sup>74</sup> Een goede website voor het volgen van de evolutie van websites en de nodige plug-ins is <http://browserwatch.internet.com>. Hier vindt u ook een goede opsomming van de software die standaard al dan niet samen met de webbrowser wordt geïnstalleerd. Deze informatie is doorgaans ook beschikbaar in de helpfunctie van de webbrowser.

<sup>75</sup> Voor meer achtergrondinformatie over het gebruik van XML als taal voor webtoepassingen: <http://www.w3.org/MarkUp/>.

<sup>76</sup> Bijvoorbeeld: in HTML 4.0 werd de tag <PRE> opgenomen die de tags <XMP>, <PLAINTEXT> en <LISTING> verving. Andere nieuwe tags in HTML 4.01 zijn: <ABBR>, <ACRONYM>, <BDO>, <BUTTON>, <COL>, <COLGROUP>, <DEL>, <FIELDSET>, <FRAME>, <FRAMESET>, <IFRAME>, <INS>, <LABEL>, <LEGEND>, <NOFRAMES>, <NOSCRIPT>, <OBJECT>, <OPTGROUP>, <PARAM>, <SPAN>, <TBODY>, <TFOOT>, <THEAD> en <Q>. Voorbeelden van deprecated HTML-tags zijn: <APPLET>, <BASEFONT>, <CENTER>, <DIR>, <FONT>, <ISINDEX>, <MENU>, <S>, <STRIKE>, <U>. Ook de attributen kunnen 'deprecated' zijn. In HTML 4.01 zijn de volgende attributen van het element BODY deprecated: "background", "text", "link", "vlink", "alink".

verouderde of niet gestandaardiseerde tags te vervangen door nieuwe tags of tags met dezelfde functie. In feite komt dit neer op een migratie van de HTML-bestanden.

Migratie betekent overwegend nog het herschrijven of herprogrammeren van de HTML-syntax. Aangezien dit proces nog nauwelijks of niet automatisch kan verlopen, is het duidelijk dat de migratiepiste niet op grote schaal kan worden toegepast. Migratie komt meestal neer op handmatig verbeterwerk, vereist grondige (X)HTML-kennis en is arbeidsintensief, ook al kan er hiervoor een tool worden ingeschakeld<sup>77</sup>. Omzetting van HTML-tags lijkt aangewezen bij het aanpassen van niet-gestandaardiseerde tags en verkeerde attributen van HTML-bestanden vooraleer ze in het website-archief worden opgenomen<sup>78</sup>. In de praktijk kan men dit nog wat uitstellen: de meest gangbare webbrowsers zijn heel relax in het toepassen van de HTML-syntaxregels en kunnen webpagina's met foutieve code nog correct omzetten.

Emulatie van de nodige webbrowsers lijkt daarentegen meer geschikt te zijn als antwoord op de HTML-evolutie op lange termijn. Met emulatoren behoeven HTML-bestanden met verouderde tags geen aanpassing en kunnen ze in hun originele vorm worden bekeken. Voor het ogenblik is er nog geen nood aan emulatoren, want de gangbare webbrowsers (Internet Explorer 6.x, Netscape 6.x) kunnen ook websites gebaseerd op de oude HTML-specificaties weergeven. Met de komst van XHTML en de verdere ontwikkeling van stylesheets is het niet ondenkbaar dat de HTML-ondersteuning binnen afzienbare tijd wordt afgebouwd. Een emulator van de recentste HTML-browser die ook vroegere versies ondersteunt, zou dan moeten volstaan.

De huidige generatie webbrowsers zijn grafische webbrowsers die in staat zijn om afbeeldingen in diverse bestandsformaten (GIF, TIFF, JPEG, PNG) op het scherm weer te geven. De webbrowsers volgen ook de evolutie inzake de bestandsformaten voor afbeeldingen en moeten de overeenstemmende decompressies kunnen uitvoeren (bijv. JPEG2000: najaar 2001). Er is dus een verband tussen de browserprogramma's enerzijds en de HTML-versies en de afbeeldingen die webpagina's bevatten anderzijds.

Een vierde vereiste voor de webbrowser heeft betrekking op de uitvoering van scripts die aan clientzijde plaats vinden. Oudere webbrowsers kunnen geen JAVA- of VBscript uitvoeren. De nodige software voor het uitvoeren van scripts kan standaard in een webbrowser aanwezig zijn of kan als plug-in bij de webbrowser worden geïnstalleerd.

Tenslotte moet de webbrowser ook webpagina's gelinkt aan stylesheets correct kunnen weergeven. De evolutie is immers gericht op een scheiding van HTML-markup en lay-out (CSS, XSL). Bij het correct toepassen van de recente (X)HTML- en XML-standaarden is het gebruik van stylesheets onvermijdbaar geworden.

De digitale duurzaamheid van de website plug-ins is een ander paar mouwen. De plug-ins zijn meestal systeemafhankelijk en producentgebonden (bijv. Flash Player, Real Player, Live 3D, Shockwave, Real Audio, Acrobat Reader, enz.). De gevallen waarbij een producentgebonden plug-in naar neutrale plug-ins kunnen overgezet worden, zullen meer uitzondering dan regel zijn. Zolang de nieuwe versies van de plug-in voor achterwaartse ondersteuning zorgen, is er geen probleem. Op het ogenblik dat die achterwaartse ondersteuning wegvalt, moet er voor een alternatief worden gezorgd.

---

<sup>77</sup> Binnen de schoot van het *World Wide Web Consortium* heeft een groep vrijwilligers een tool ontwikkeld voor het opsporen en verbeteren van fouten in de HTML-syntax. Meer uitleg, de broncode en downloads zijn beschikbaar op <http://www.w3.org/People/Raggett/tidy/>.

<sup>78</sup> Een analyse van de websitecollectie van het Pandoraproject wees uit dat dit een reële bedreiging is: 7 miljoen niet-gestandaardiseerde HTML-tags en 14 miljoen HTML-tags met verkeerde attributen (W. CATHRO, C. WEBB en J. WHITING, *Archiving the web: the pandora archive at the National Library of Australia*).

De websites die een Flash-toepassing zijn, zullen wellicht een heel eigen aanpak vergen. Voor het raadplegen van deze websites is een webbrowser alleen onvoldoende, maar dient men ook over de specifieke Flash Player te beschikken. De heel specifieke Flashfunctionaliteit op het vlak van animatie en vectorafbeeldingen en zijn producentgebondenheid maakt de archivering van websites in Flash geen gemakkelijke opgave. Emulatie van de plug-in zou één oplossing kunnen zijn. Een andere mogelijkheid is het omzetten van de Flash-bestanden naar een sequentie van meer gestandaardiseerde bestandsformaten. In hoeverre beide pistes haalbaar zijn, is voor het ogenblik moeilijk te voorspellen. Aangezien de meeste websites in Flash ook nog een statische HTML-broer hebben zou men het zekere voor het onzekere kunnen kiezen en de HTML-versie archiveren.

De downloads van een website met archiefwaarde worden bij voorkeur naar een geschikt archiveringsformaat omgezet. Hierbij moet men eraan denken dat bij omzetting de bestanden een nieuwe extensie krijgen en dat de links moeten aangepast worden. In de meeste gevallen is de digitale informatie on line in een bestandsformaat met de status van standaard opgeslagen (bijv. \*.doc, \*.pdf). Men kan zich wel de vraag stellen of het niet beter is om de audi-visuele streamingbestanden naar gewone audio-visuele bestanden om te zetten. Deze laatste bestandsformaten zijn doorgaans beter gestandaardiseerd en hierdoor zou opnieuw een stukje software-afhankelijkheid worden vermeden.

## **F. BEHEER VAN DE GEARCHIVEERDE WEBSITES**

### **F.1 Beschrijven**

In de metadata van de gearchiveerde websites worden naast de klassieke metadata van digitale archiefbescheiden een aantal gegevens specifiek voor websites opgenomen:

#### **ALGEMEEN**

- ▶ titel
- ▶ archiefvormer
- ▶ onderwerp/abstract
- ▶ functie/doelstelling
- ▶ on line versies
- ▶ teller:
  - ❑ telleraantal bij on line plaatsing
  - ❑ telleraantal op het tijdstip van archivering
  - ❑ telleraantal bij off line plaatsing
- ▶ openbaarheid

#### **WEBSERVER**

##### *PLATFORM*

- ▶ hardware
- ▶ besturingssysteem
- ▶ webserverconfiguratie en -software
- ▶ serverscripts

- ☐ CGI
- ☐ ASP
- ☐ JSP
- ☐ PHP
- ☐ andere
- ▶ uitvoerbare programma's
- ▶ koppeling met applicaties

#### *LOGBESTANDEN*

- ▶ inhoud
- ▶ frequentie

#### *DOCUMENTATIE*

### **MIRROR/SNAPSHOT**

#### *ALGEMEEN*

- ▶ gearcheiverde versie
- ▶ versienummer
- ▶ webmaster
- ▶ webdesign
- ▶ inhoudsverantwoordelijken
- ▶ talen
- ▶ aanpassingen
- ▶ ontbrekende bestanden

#### *TECHNISCHE GEGEVENS*

- ▶ URL / IP-adres
- ▶ startpagina
- ▶ sitemap
- ▶ aantal bestanden
- ▶ aantal mappen
- ▶ totale bestandsomvang
- ▶ paswoord(en)

#### *BESTANDSFORMATEN EN VERSIES*

- ☐ (X)HTML
- ☐ Tekstbestanden:
  - ASCII
  - PDF
  - MS WORD
  - WordPerfect
  - andere
- ☐ Audiobestanden:
  - au
  - wav
  - mp3
  - midi

- andere
- ❑ Videobestanden:
  - mpeg
  - mov/qt
  - avi
  - andere
- ❑ Afbeeldingsbestanden:
  - GIF
  - JPEG
  - TIFF
  - PNG
  - andere
- ❑ Animatietoepassingen:
  - animated gifs
  - Shockwave Flash Player / Movie
  - andere
- ❑ Uitvoerbare programma's
- ❑ Clientscripts
  - Java
  - ActiveX
  - webbrowser(s)
  - plug-ins

#### *HISTORIEK*

- ▶ datum on line beschikbaar
- ▶ wijzigingen/updates
- ▶ datum verwijdering
- ▶ datum snapshot en opname in het archiveringssysteem

#### *SOFTWARE BIJ RAADPLEGING*

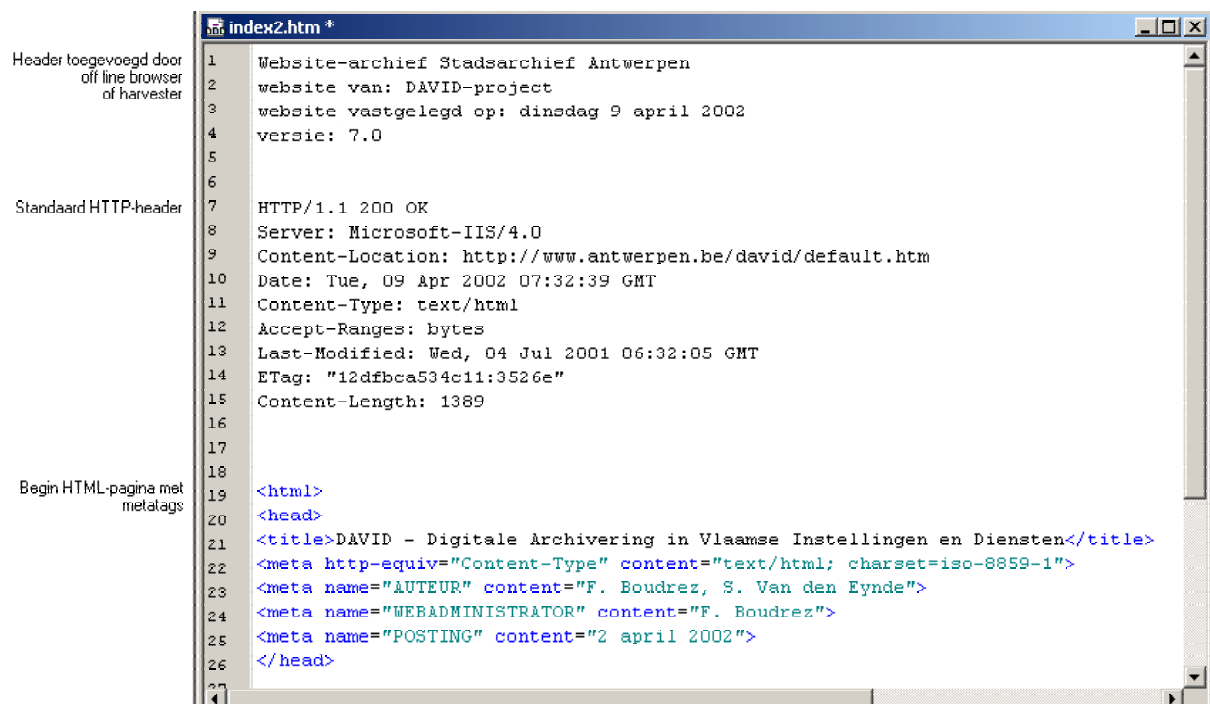
- ▶ webbrowser
- ▶ plug-ins

#### *FOUTEN EN OPMERKINGEN*

- ▶ fouten / gebreken
- ▶ opmerkingen

De hierboven opgesomde metadata kunnen bijna enkel mits medewerking van de archiefvormer worden aangelegd. Deze metadata velden achteraf invullen is voor de archivaris tijdrovend en soms onmogelijk. Een aantal metadata kunnen enkel door de archiefvormer meegedeeld worden, en kunnen niet uit de neergelegde bestanden worden afgeleid. Gemakkelijkheidshalve kan men de archiefvormer een (digitaal) formulier laten invullen. Wanneer de archivaris toch zelf de metadata moet aanleggen dan gebruikt hij hiervoor het best software die wordt gebruikt voor het testen van de website. Een aantal technische gegevens (aantal webpagina's, gebruikte bestandsformaten, enz.) worden door deze programma's in statistieken bijgehouden en kunnen gemakkelijk in de metadata worden overgenomen..

De metadata kunnen op meerdere plaatsen worden opgeslagen. Er zijn verschillende opties. Aangezien men aan HTML-pagina's headerinformatie kan toevoegen, is het mogelijk om de metadata in de header van de webpagina als HTML-metatags op te nemen. Een andere optie is het vastleggen van de metadata in een afzonderlijk computerbestand (bijv. een XML-bestand<sup>79</sup>). Een combinatie van beide is evenzeer mogelijk. In beide gevallen legt men de metadata op het niveau van de website vast. Wil men daarentegen van elk bestand afzonderlijk een aantal metadata vastleggen, dan biedt het multipart MIME-object (rfc 2045-2049<sup>80</sup>) een alternatief. Hierdoor kan aan elk type bestand een header met een aantal metadagegevens worden toegevoegd. Een gelijkaardige mogelijkheid is het uitbreiden van de standaard HTTP-header door met de webharvester of off line browser bijkomende metadata toe te voegen. Het computerbestand bevat dan gegevens over het archiveringsproces, gegevens over het object en ten slotte de data van het archiefobject zelf. Bij de keuze van de opslagplaats van de metadata dient men niet enkel met een veilige, efficiënte en duurzame bewaring rekening te houden, maar ook met het gebruikersmak. De metadata moeten voor de archiefgebruikers gemakkelijk en overzichtelijk raadpleegbaar zijn.



Afbeelding 10: Gearchiveerde webpagina met een aantal metadata in de header.

In de metadata worden ook de niet meer werkende toepassingen of functionaliteiten beschreven. Dit kan een koppeling met een bepaalde applicatie, met andere websites of verouderde plug-in zijn.

Twee soorten van metadata die specifiek zijn voor websites, verdienen bijzondere aandacht: de authenticiteitsgegevens en de datum waarop de website on line was.

<sup>79</sup> Een voorbeeld en downloadbaar model XML-bestand is samen met een bijhorende stylesheet beschikbaar op de DAVID-website.

<sup>80</sup> <http://www.rfc-editor.org/rfc/rfc2045-2049.txt>

## a) Authenticiteitsgegevens

De informatie over de authenticiteit van de te archiveren webpagina moet worden gearchiveerd. De relatief lage kosten die met webcreatie gepaard gaan en het gemak waarmee bestaande webpagina's kunnen gekopieerd worden, maken dat er op het WWW soms professioneel uitziende websites verschijnen die de surfer de indruk geven dat ze op de webstek van een bepaald bedrijf, organisatie, enz. zijn aanbeland. Op die manier proberen cybercriminelen de nummers van de kredietkaarten van nietsvermoedende cybertoeeristen te bemachtigen. Een unieke *identifier* is voor de meeste websites daarom geen overbodige luxe.

De op *Public Key Infrastructure* gebaseerde techniek van de digitale handtekening is een veel gebruikte methode om uitsluitsel te geven over de origine van een website. De titularis ondertekent zijn website digitaal. Een *server ID* bevestigt de identiteit van de titularis van de website en de eigendom van de bewuste domeinnaam.<sup>81</sup> Aan de hand van het *server ID* kan de bezoeker de digitale handtekening verifiëren en kan hij er zeker van zijn dat hij zich bevindt op de echte website van de beoogde persoon of organisatie. Een *server ID* is een digitaal certificaat dat de titularis op zijn website plaatst en dat wordt uitgereikt door een certificatedienstverlener. Alvorens zo'n server-certificaat uit te reiken, zal de certificatedienstverlener de documenten opvragen die bewijzen dat de organisatie niet onder een valse naam opereert en gerechtigd is om de domeinnaam te voeren waar de te authenticeren website achter schuilgaat.<sup>82</sup> Met een digitale handtekening kan men niet alleen de website zelf authenticeren, maar eveneens de downloads die op de website aangeboden worden, zoals software.

Er bestaan verschillende manieren om te achterhalen of een website uitgerust is met een server-certificaat. Meestal staat er een icoontje op de website dat aangeeft dat het gaat om een beveiligde site (zie afbeelding x).<sup>83</sup> Door er op te klikken, verkrijgt men informatie over het certificaat op het scherm.<sup>84</sup>



Afbeelding 11: Een "Secure Site Seal" icoontje dat aangeeft dat er voor deze website een server-certificaat werd uitgereikt door Verisign

Men kan ook via zijn browser checken of er voor een bepaalde website een certificaat werd uitgereikt. In Microsoft Explorer dient men rechts te klikken in de pagina van de website die men wil verifiëren en "Properties" te selecteren. In het Properties-venster klikt men vervolgens de "certificates button" aan. Men verkrijgt dan alle informatie over het certificaat o.a. de geldigheidsduur en de

<sup>81</sup> De registratie van domeinnamen van het type dot-be gebeurde tot eind 2000 door de vzw DNS België. Sinds 11 december 2000 registreert DNS niet zelf meer de .be domeinnamen. Er werd een netwerk van agenten opgericht die een contract hebben afgesloten met DNS om domeinnamen te registreren (bijv. Planet Internet). Nadat de agent de registratieprocedure succesvol doorlopen heeft, en de registratievergoeding werd betaald, kent DNS een exclusieve licentie toe aan de titularis van de domeinnaam om de aangevraagde domeinnaam te mogen gebruiken. Derden kunnen via de website van DNS ( <http://www.dns.be> ) nagaan of een domeinnaam geregistreerd is en aan wie die toebehoort.

<sup>82</sup> Zie bijvoorbeeld de *Serversign Certificate<sup>TM</sup> Procedure* die bij Globalsign moet gevolgd worden om een Serversign<sup>TM</sup> certificaat te bekomen [http://www.globalsign.net/digital\\_certificate/serversign/index.cfm](http://www.globalsign.net/digital_certificate/serversign/index.cfm)

<sup>83</sup> Bijv. <http://www.ignite.com/application-services/products/verisign/news/news/ns011.html>

<sup>84</sup> Bijv. <https://digitalid.trustwise.com/secureServer/cgi-bin/haydn.exe>

uitgever van het certificaat. Het venster van Netscape Navigator bevat links onderaan een icoontje dat een hangslot voorstelt. Wanneer de site beveiligd is, is het hangslot gesloten. Door erop te klikken, wordt er een scherm geopend met *security information*, o.a. over eventuele webcertificaten. Men kan tenslotte ook op de website van de certificatedienstverlener checken of er voor een bepaalde website door deze certificatedienstverlener een certificaat werd uitgereikt.<sup>85</sup>

b) Datum waarop de website on line beschikbaar was

Eveneens van groot belang is de datum waarop de website on line beschikbaar was. Een aantal projecten<sup>86</sup> geven elke website een time-stamp bij de archivering. De datum is een zeer belangrijk gegeven omdat het discussies kan helpen beslechten over de vraag wie een bepaalde wetenschappelijke vondst eerst publiceerde of over inbreuken op het auteursrecht. Ook naar aanleiding van een octrooiaanvraag kan een website-archief, uitgerust met de time stamp functionaliteit, doorslaggevend zijn bij de beslissing om het octrooi al dan niet toe te kennen. De wet op de uitvindingsoctrooiën van 28 maart 1984 (B.S. 9 maart 1985) bepaalt immers dat een octrooi slechts kan worden verleend indien de uitvinding *nieuw* is. Een uitvinding wordt als nieuw beschouwd indien zij geen deel uitmaakt van de stand van de techniek. De stand van de techniek wordt gevormd door al hetgeen vóór de datum van indiening van de octrooiaanvraag openbaar toegankelijk is gemaakt door een schriftelijke of mondelinge beschrijving, door toepassing of op enige andere wijze.<sup>87</sup> Een publicatie van de uitvinding op Internet (zelfs op een Chinese website in het Chinees) vóór de datum van de octrooiaanvraag heeft dus tot gevolg dat de uitvinding niet meer nieuw is en dus niet in aanmerking komt voor octrooiering. De nieuwheidstoets is immers gebaseerd op wereldwijd verspreide kennis in verband met uitvindingen.

Wat is dat nu, zo'n time-stamp en hoe wordt die aan de te archiveren website toegevoegd? Na de selectie van de website berekent het archiveringssysteem een hashing-code voor deze website. Vervolgens voegt het archiveringssysteem de juiste datum en eventueel de tijd toe aan de hashing-code en ondertekent het geheel met zijn private sleutel. Het resultaat is de time-stamp. Time-stamping is gebaseerd op de idee dat de getijdstempelde elektronische informatie (bijv. een website) moet bestaan hebben ten laatste op het moment dat de time-stamp tot stand komt. Dit is logisch, vermits de time-stamp berekend wordt op basis van de inhoud van de getijdstempelde informatie.

## F.2 Veilige bewaring

Bij het beheer van de gearchiveerde websites moet ook nagedacht worden over de veilige bewaring van de websites. De opname van de snapshots in een courant document- of archiefbeheerssysteem ligt niet voor de hand. De kans is immers groot dat de hyperlinks niet meer werken. De identificatie van de bestanden in een document- of archiefbeheerssysteem gebeurt niet op basis van de bestandsnamen, maar op basis van de ID's van de bestanden. Met het oog op latere terbeschikkingstelling is het verder functioneren van de mirror of het snapshot een must. Alternatieven zijn de bewaring van de snapshots

---

<sup>85</sup> <http://secure.globalsign.net/phoenixng/services.cfm?id=1413967734&reset=yes>

<sup>86</sup> Bijv. The Nordic Web Archive

<sup>87</sup> Art. 5 van de wet op de uitvindingsoctrooiën.

op de server in beveiligde mappen (firewall, beperkte toegangsrechten) of op een read-only drager. Het is ook mogelijk om de snapshot in het document- of archiefbeheerssysteem op te nemen en een kopie voor raadpleging op de server ter beschikking te houden.

### F.3 Opslag op informatiedragers

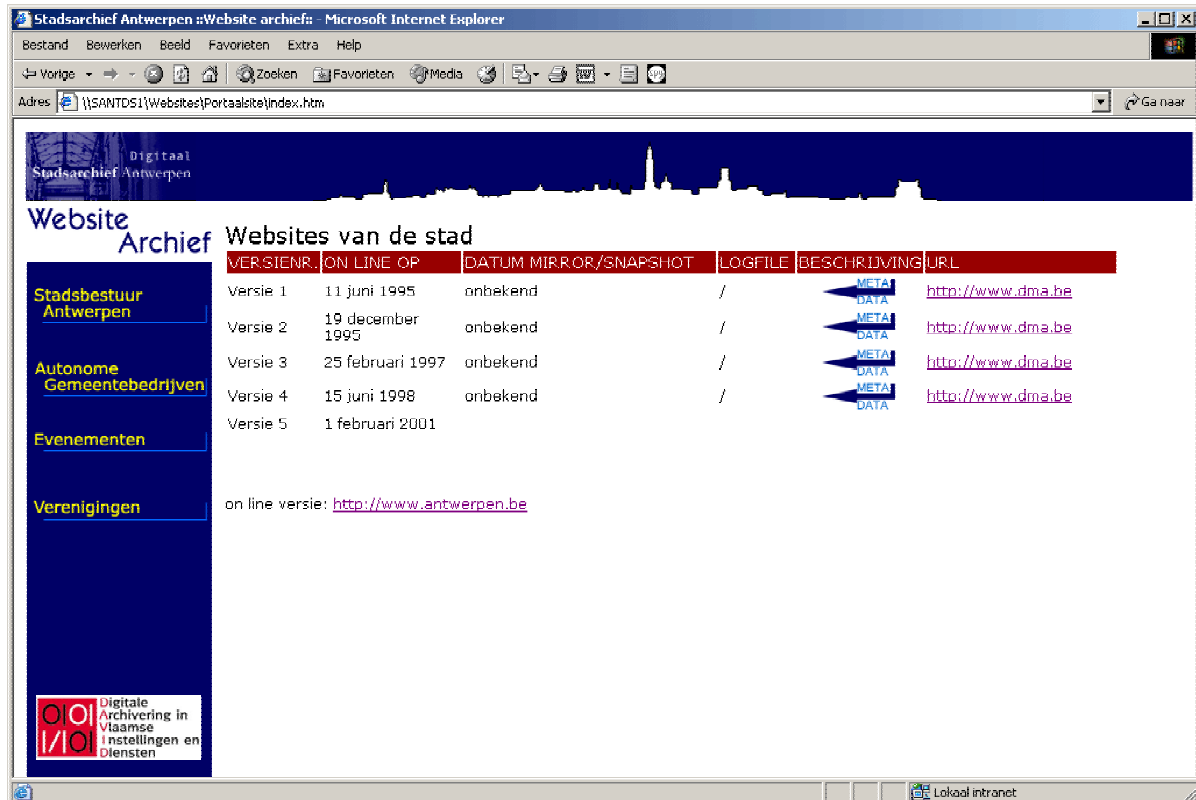
Gearchiveerde mirrors en snapshots kunnen op harde schijven, magnetische banden en optische schijven worden bewaard. Bij magnetische banden en optische schijven is het vanwege de digitale duurzaamheid belangrijk om zoveel mogelijk standaarden toe te passen inzake formaten, type dragers, bestandssysteem, enz. Een belangrijk aandachtspunt bij het gebruik van CD-ROMS als dragers voor gearchiveerde websites zijn de map- en bestandsnamen. De ISO-9660 standaard voor CD-ROMS houdt immers een aantal beperkingen in inzake toegestane karakters en lengte van map- en bestandsnamen. Aanpassen van map- en bestandsnamen brengt met zich mee dat ook alle links aanpassing behoeven. Bij het gebruik van CD-ROMS als informatiedrager is het bijgevolg aangewezen dat de off line browser alle map- en bestandsnamen omzet naar het 8+3 compatibele systeem. Harde schijven en optische dragers zijn sneller toegankelijk en raadpleegbaar dan magnetische banden.

## H. TER BESCHIKKING STELLEN

De gearchiveerde websites worden bij voorkeur digitaal ter beschikking gesteld. Voor later onderzoek of toekomstige verantwoording worden websites het best bekeken zoals ze in hun primaire vorm beschikbaar waren op het WWW.

Een mogelijke manier om websites te ontsluiten en ter beschikking te stellen is via een portaalsite. Vanuit de portaalsite is elke gearchiveerde website met bijhorende metadata, sitemap en logbestanden raadpleegbaar. Men kan zijn website-archief beschikbaar stellen vanop bepaalde dragers, de harde schijf van een computer of via een netwerk. Als men over een webserver beschikt, kan men probleemloos alle gearchiveerde websites on line ter beschikking stellen. Een koppeling tussen webserver en archiefbeheerssysteem lijkt de aangewezen oplossing voor on line ter beschikking stelling. In dit geval neemt men best ook een aantal voorzorgsmaatregelen zodat de bezoeker weet dat het om een gearchiveerde versie gaat. Men moet er immers van uitgaan dat ook zoekrobots de gearchiveerde on line websites zullen indexeren en dat bezoekers rechtstreeks toegang zullen hebben zonder eerst langs de website van het archiefdienst te passeren.

Bij het terbeschikking stellen en uitbouwen van een archief voor on line bronnen dient niet alleen nagedacht te worden over de raadpleging van gearchiveerde websites. Ook de gerelateerde (digitale) archiefdocumenten en de metadata moeten toegankelijk en leesbaar zijn.



**Afbeelding 12:** Een portaalsite voor het website-archief van het Stadsarchief Antwerpen. Vanuit de portaalsite zijn de verzamelde websites en hun metadata raadpleegbaar. Op de overzichtspagina van elke website staan de verschillende versies, de data waarop de versies on line werden geplaatst, de data van de snapshots, een link naar het logbestand van de snapshotoperatie, de URL met een link naar de startpagina van elke gearchiveerde versie, een link naar de metadata van elke versie en ten slotte een link naar de versie die on line staat. Een on line versie van deze portaalsite is beschikbaar op de DAVID website.

## VI. UITBOUWEN VAN HET ARCHIVERINGSSYSTEEM

### A. ROL VAN DE ARCHIVARIS

De archivaris speelt een belangrijke rol in het archiveringsproces van websites. Binnen de meeste organisaties en instellingen zal er nog maar weinig aandacht aan websites-archivering zijn besteed. Het is zijn/haar taak de archivering van websites en gerelateerde archiefdocumenten op de agenda te plaatsen en erop toe te zien dat het archiveringssysteem aan de vooropgezette kwaliteitsvereisten beantwoordt.

De archivaris zorgt voor:

- ➡ naleving van de geldende wet-en regelgeving met betrekking tot het auteursrecht en de

bescherming van de persoonsgegevens<sup>88</sup>

- ➡ de archiefwaardering van websites en gerelateerde archiefdocumenten
- ➡ de definiëring en identificatie van de archiefbescheiden in het totale informatiesysteem: alleen deze archiefdocumenten worden in het archiveringssysteem opgenomen
- ➡ bewustmaking van de betrokken diensten en IT-verantwoordelijken van de archiefwaarde van bepaalde websites en hun gerelateerde archiefdocumenten (o.a. voorkomen van onrechtmatige vernietigingen, registreren van de nodige gegevens)
- ➡ het bijhouden van de nodige metadata
- ➡ de afstemming van de websitesarchiveringsstrategie op het globale archiveringssysteem van de organisatie
- ➡ het naleven van de vooropgezette kwaliteitsvereisten in het archiveringssysteem
- ➡ de evaluatie of implementatie van archiveringsmodules in de bestaande informatiesystemen, of bij ontstentenis het oprichten van lokale website-archieven
- ➡ de verspreiding van het archiveringssysteem binnen de organisatie
- ➡ een kwaliteitscontrole van de mirrors/snapshot alvorens ze in het digitaal depot worden opgenomen
- ➡ maatregelen om de raadpleegbaarheid (toegankelijkheid en leesbaarheid) op lange termijn te garanderen: migratie, emulatie, opslag op veilige en duurzame informatiedragers, toepassing van zoveel mogelijk standaarden, periodieke controle van de informatiedragers en overzetting van de digitale informatie naar nieuwe gestandaardiseerde dragers.

Het eindresultaat is een archiveringssysteem met duidelijke procedures, routines en afbakening van de verantwoordelijkheden. Bij website-archivering zijn de inhoudsverantwoordelijken (meestal de diensten zelf) en de webmasters/administrators betrokken. Wie dat precies zijn, hangt af van de organisatiestructuur van de instelling en de procedure voor het aanpassen en beheren van de website. In kleine organisaties neemt meestal dezelfde persoon beide functies op zich. In grotere instellingen zullen beide taken meer gescheiden zijn.

De archivaris zal als trekker van het archiveringssysteem functioneren. Wanneer de archivaris websitesarchivering binnen zijn organisatie ter sprake brengt, zal ongetwijfeld van zijn kant enige inbreng worden verwacht. Concrete richtlijnen voor de creatie, het beheer en de archivering kunnen als vertrekpunt voor het uitbouwen van een archiveringssysteem dienen. Hieronder worden voorbeelden van dergelijke richtlijnen gegeven. De richtlijnen mbt ontwerp en beheer zijn universeel toepasbaar en gelden voor alle types websites. Deze richtlijnen kunnen hier al vrij concreet uitgewerkt worden. Bij de archiveringsrichtlijnen is dit minder het geval. De archiveringsstrategie is immers afhankelijk van het acquisitiebeleid, de organisatie, het type website(s) en de beschikbare IT-infrastructuur. De archiveringsrichtlijn beperkt zich in dit rapport tot de elementen die er deel van uit moeten maken. Beide richtlijnen werden meer geconcretiseerd in *Digitale Archivering: Richtlijn & advies nr. 5 'Websitesbeheer voor archivering'*<sup>89</sup>.

<sup>88</sup> Zie hiervoor de hoofdstukken VII en VIII in dit rapport.

<sup>89</sup> Dit document is beschikbaar op de DAVID-website.

## B. EFFICIËNT ONTWERP EN BEHEER

Websitesarchivering is gemakkelijker te realiseren wanneer bij het ontwerp en het beheer van websites al met archivering rekening wordt gehouden.

Het toepassen van een aantal eenvoudige regels mbt efficiënt ontwerp en beheer van de websites zorgt er ten eerste voor dat de aanpassingen op het archiveringsmoment en tijdens het verdere beheer van de gearchiveerde website minimaal zijn. Deze regels hebben betrekking op het maken van goed gestructureerde (X)HTML-pagina's, het respecteren van de (X)HTML-syntaxregels en het gebruik van gestandaardiseerde (X)HTML-tags en bijhorende attributen. De website-ontwerper heeft er overigens alle belang bij om deze regels toe te passen, want zo wordt vermeden dat de website voor een deel van de potentiële bezoekers niet toegankelijk is. Bij het ontwerpen en beheren van webpagina's is het ten tweede ook belangrijk om het aanmaken, wijzigen en posten van de websites expliciet te documenteren, zodat deze gegevens later als metadata bruikbaar zijn. Duidelijke richtlijnen en regels moeten ook voorkomen dat digitale bestanden met archiefwaarde onrechtmatig worden vernietigd of overschreven.

De richtlijn met betrekking tot creatie en beheer is deels gebaseerd op de *Web Content Accessibility*<sup>90</sup> regels van het *World Wide Web Consortium*. Het verzekeren van de uitwisselbaarheid van webinformatie vertoont veel gelijkenissen met de maatregelen voor digitale duurzaamheid. In deze richtlijn zitten anderzijds ook de kwaliteitsvereisten en de eerste ervaringen van het stadsarchief Antwerpen inzake websites-archivering vervat.

### Algemeen

- ▶ Maak duidelijke afspraken over het beheer van de website:
  - ❑ Wie is verantwoordelijk voor het ontwerp en de inhoud?
  - ❑ Wie wijzigt de inhoud van de website? Wie zorgt er voor de updates?
  - ❑ Hoe zorg je ervoor dat de informatie up-to-date is?
  - ❑ Wat gebeurt er met de verouderde informatie?
  - ❑ Wie houdt welke documentatie over het ontwerp en het beheer van de website bij?
- ▶ Houd rekening met de mogelijke archiefwaarde bij het ontwerpen van een website. Denk niet alleen aan de website zelf, maar ook aan gerelateerde e-mails, logbestanden, databanken of documentbeheerssystemen. Leg op voorhand de archiefwaarde van gerelateerde documenten vast en voorkom dat ze worden vernietigd of overschreven.

---

<sup>90</sup> Binnen de schoot van het W3C buigt de Web Accessibility Initiative-werkgroep zich over de toegankelijkheidsproblematiek van het web. WAI stelde onder meer de *Web Content Accessibility Guidelines 1.0* en een bijhorende checklist op. (<http://www.w3.org/WAI/>). In Canada en de Verenigde Staten zijn overheidsadministraties verplicht deze regels toe te passen (Canada: *Government of Canada Internet Guide* [http://www.cio-dpi.gc.ca/ig-gi/index\\_e.asp](http://www.cio-dpi.gc.ca/ig-gi/index_e.asp); VS: *Electronic and Information Technology Accessibility Standards Section 508*).

### Lange termijn planning

Denk op lange termijn bij het opstarten van een website. Zorg enerzijds voor een duidelijke en beheersbare procedure voor het up-to-date houden van de website. Zorg er anderzijds ook voor dat oudere informatie beschikbaar blijft. Houd hier rekening mee bij het uitwerken van een mappenstructuur en het maken van links.

### Mappenstructuur

- ▶ ontwerp een duidelijke mappenstructuur voor de bestanden die de website vormen
- ▶ plaats alle bestanden en submappen die de website vormen in één gemeenschappelijke rootmap
- ▶ denk vooruit en plan onmiddellijk de verdere uitbouw van de website in de toekomst. Houd hier rekening mee bij het uitwerken van de mappenstructuur zodat deze niet bij elke versie wordt gewijzigd. Zorg voor consistentie in de pathaanduidingen (functionaliteit links!).

### Bestanden

- ▶ Gebruik zoveel mogelijk gestandaardiseerde bestandsformaten. Een toepassing hiervan zou kunnen zijn:

Tekst: XML, (X)HTML, PDF, Word

Afbeeldingen: JPEG, GIF, PNG

Animaties: GIF, Flash

Video: MPEG

Geluid: MP3, WAV

Stylesheets: CSS, XSL

- ▶ Geef bestanden en hun versies zo uniek mogelijke bestandsnamen.

### Externe webpagina's

Vermijd of beperk zoveel mogelijk het gebruik van webpagina's of afbeeldingen van andere websites. Je hebt geen controle op de beschikbaarheid, de inhoud en het functioneren van deze webpagina's.

### Links

- ▶ Interne links: gebruik relatieve pathaanduidingen
- ▶ Externe links: gebruik absolute pathaanduidingen
- ▶ Documenteer de link door duidelijk zijn doel aan te geven

## Webpagina's

- ▶ Gebruik markup en stylesheets waarvoor ze eigenlijk zijn bedoeld. Houd structuur en opmaak van elkaar gescheiden. Gebruik de (X)HTML-tags voor de markup van de tekst en stylesheets voor de lay-out.
- ▶ Bouw overzichtelijke webpagina's. Breng een duidelijke structuur aan in de webpagina's. Bouw de webpagina in drie delen op:
  - ❑ eerste lijn: verduidelijking van de toegepaste (X)HTML-versie (en eventueel specificatie van de DTD): `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN "http://www.w3.org/TR/html4/strict.dtd">` of `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "DTD/xhtml11-strict.dtd">`
  - ❑ header: `<HEAD></HEAD>`
  - ❑ body: `<BODY></BODY>`
- ▶ het rootelement is altijd: `<HTML></HTML>`
- ▶ ken binnen de header aan elke webpagina een titel toe: `<TITLE></TITLE>`

## Documenteer elke versie: Metadata

- ▶ Beschrijf de gearcheiverde websites, hun versies en hun wijzigingen. Houd de metadata per versie in een afzonderlijk bestand bij of centraliseer alle metadata.
- ▶ Zorg voor een vastgelegd metadataschema. Stel dit schema beschikbaar in de vorm van een formulier, XML-bestand of databank.
- ▶ Leg vast wie de metadata bijhoudt.

## (X)HTML-tags en attributen

- ▶ gebruik enkel de gestandaardiseerde (X)HTML-tags. Vermijd het gebruik van afgekeurde tags en attributen. De standaarden zijn: HTML 4.01 (ISO/IEC-15445:2000) en XHTML 1.0.
- ▶ respecteer de volgorde van de (X)HTML-elementen
- ▶ plaats de elementen in de juiste volgorde: orden de (afsluitende) tags correct
- ▶ sluit de elementen altijd af

### Bij gebruik van:

### Zorg er voor dat:

- ▶ niet-tekstuele info: er een tekstueel alternatief is via de attributen "ALT" of "LONGDESC". Voorbeelden van niet textuele elementen zijn afbeeldingen, animaties, image maps, applets, grafische knoppen.
- ▶ links: het doel van de link duidelijk is
- ▶ stylesheets: de HTML-bestanden ook zonder de stylesheets leesbaar zijn.
- ▶ image maps: er zoveel mogelijk client image maps worden gebruikt, vermijd het gebruik van server image maps
- ▶ tabellen: de headers van rijen en kolommen duidelijk geïdentificeerd zijn

- ▶ frames: elk frame geïdentificeerd wordt aan de hand van een betekenisvolle naam en verduidelijkt de relatie tussen de frames wanneer dit niet uit de naam kan worden afgeleid
- ▶ dynamische inhoud: er een alternatief is wanneer de interactie niet functioneert
- ▶ scripts / applets: de webpagina nog werkt wanneer de scripts/applets afstaan of niet worden ondersteund. Als dit niet mogelijk is, verwijst dan naar een gelijkaardige webpagina met dezelfde informatie

## C. ARCHIVERING

Het archiveringssysteem is afhankelijk van de organisatie en de beschikbare technologische infrastructuur. De meest optimale situatie voor systematische website-archivering is de integratie van het websitebeheer met het recordsmanagements- en archiveringssysteem van de organisatie. Zo bekomt men dat websites, de op websites gebaseerde archiefdocumenten, de aanpassingen en updates doeltreffend worden bijgehouden. Een dergelijke situatie is voor het ogenblik bij Vlaamse instellingen en diensten echter meer uitzondering dan regel. De statische of dynamische websites staan meestal los van het recordmanagementsysteem en het gros van de in gebruik zijnde informaticasystemen zullen niet de nodige functionaliteiten bieden, zodat aanpassingen en updates aan hun controle en versiebeheer ontsnappen. De archivering van websites en de gerelateerde archiefdocumenten staat nog maar in zijn kinderschoenen. Het is de komende jaren een uitdaging voor de archivaris om archiveringsmodules aan de bestaande informatiesystemen te koppelen of erin te laten opnemen.

In afwachting en bij ontstentenis van een gestructureerde archivering ontsnapt de organisatie niet aan de archiveringsplicht en moet er een (tussentijdse) oplossing worden gezocht. Het enige alternatief is dan het opzetten van een off line website-archief waarin de website-versies, hun wijzigingen en de gerelateerde digitale archiefdocumenten worden bijgehouden<sup>91</sup>. Deze oplossing kan even goed een definitieve archiveringsoptie zijn, maar is arbeidsintensief zodat bij websites met een hoge updatefrequentie een meer geautomatiseerde aanpak wenselijk is.

Voor het bijhouden van zo'n lokaal off line website archief zijn er in principe twee opties. Men kan de verantwoordelijkheid voor het archiveren naar de betrokken administraties of diensten delegeren. In grote organisaties kan dit de meest aangewezen optie zijn. Het archiveren van websites zou een onderdeel van de routine bij het aanpassen en updaten van de website moeten worden. De andere optie is het centraal archiveren van alle websites. In dit geval blijft de verantwoordelijkheid van de administraties of de diensten beperkt tot het melden van wijzigingen. Voor het archiveren van de websites zelf kan in beide gevallen dezelfde procedure worden gevolgd. Het archiveringsmoment laat men best zo dicht mogelijk aansluiten bij het tijdstip van creatie, en niet bij het tijdstip waarop een nieuwe versie online gaat.

Een procedure of richtlijn voor het archiveren van websites zou volgende onderdelen moeten bevatten:

---

<sup>91</sup> Zie ook: *Managing Internet and Intranet Information for Long-term Access and Accountability. Implementation Guide*, 1999.

## Algemeen

- ▶ Maak duidelijke afspraken over het archiveren van de website: leg de verschillende verantwoordelijken vast: wie meldt een wijziging? wie archiveert de website en de wijzigingen?
- ▶ laat het archiveringsmoment onmiddellijk aansluiten bij het updaten van de website. Zorg ervoor dat de archivering een vast onderdeel van de routine bij het aanbrengen van wijzigingen wordt.
- ▶ leg op voorhand de archiefwaarde van de website en de gerelateerde archiefdocumenten vast
- ▶ vertrouw je niet op backuptapes bij het archiveren van websites
- ▶ verwijder bestanden op de webserver pas wanneer je zeker bent dat ze al zijn gearchiveerd

## Waar wordt het website-archief bijgehouden?

- ▶ Zorg voor een afgeschermd bewaarplaats voor gearchiveerde websites:
  - ❑ worden de websites centraal of decentraal bijgehouden?
  - ❑ in welke mappen worden de gearchiveerde websites en hun versies bijgehouden? op welke drager? Welke mapnamen worden gebruikt?
  - ❑ beperk de toegangsrechten tot deze mappen
  - ❑ houd deze map duidelijk gescheiden van de werkmappen waarbinnen de website wordt ontworpen
  - ❑ zorg regelmatig voor een veiligheidskopie van deze map

## Wat archiveren?

- ▶ De website:
  - ❑ Is het uitgangspunt een volledig snapshot of mirror van elke nieuwe website?
  - ❑ Bij websites met een dynamische inhoud: worden er enkel snapshots gearchiveerd of worden ook de originele serverbestanden (ASP-, PHP- of JSP-bestanden) bewaard?
  - ❑ Welke documentatie over de website heeft archiefwaarde?
- ▶ De serverscripts:
  - ❑ Worden de serverscripts gearchiveerd? Wordt elke versie bijgehouden?
  - ❑ Waar worden de serverscripts bijgehouden?
  - ❑ Hoe worden de serverscripts gearchiveerd?
- ▶ De downloads:
  - ❑ Hebben de downloads archiefwaarde?
  - ❑ Worden de downloads op andere plaatsen gearchiveerd of bij de mirrors/snapshots?
- ▶ De databanken:
  - ❑ Hebben de gekoppelde databanken archiefwaarde? Welke databanken of tabellen worden gearchiveerd?
  - ❑ Hoe wordt de informatie in gekoppelde databanken of documentbeheerssystemen gearchiveerd?
- ▶ De e-mails:

- ☐ Worden de ontvangen e-mails gearchiveerd?
- ☐ Hoe worden de e-mails gearchiveerd?
- ▶ De logbestanden:
  - ☐ Hebben de logbestanden archiefwaarde?
  - ☐ Worden logbestanden zelf of de analyserapporten bewaard?
  - ☐ Wordt er informatie uit de logbestanden in de metadata opgenomen?
  - ☐ Hoe worden de logbestanden gearchiveerd?
  - ☐ Wat is de bewaartermijn van de logbestanden?

#### Hoe archiveren?

- ▶ Gaat het om een website met een vaste of dynamische inhoud? Worden er mirrors of snapshots gearchiveerd?
- ▶ Met welke toepassing wordt een snapshot van een website met dynamische inhoud gemaakt?
- ▶ Waaruit bestaat de kwaliteitscontrole van mirrors of snapshots?

#### Versies / Frequentie

- ▶ Worden alle wijzigingen/versies gearchiveerd?
- ▶ Wordt na elke wijziging de volledige website gearchiveerd of enkel de wijziging?
- ▶ Met welke frequentie worden momentopnames gearchiveerd?
- ▶ Hoe en in welke mappen worden de versies bewaard? Hoe worden de bestandsnamen aangepast?
- ▶ Hoe worden de verschillende versies in de metadata gedocumenteerd?

#### Metadata

- ▶ Welke bijkomende metadata worden over de gearchiveerde website bijgehouden?

#### Digitaal depot

- ▶ Wanneer worden de gearchiveerde websites in het digitaal depot opgenomen?
- ▶ Hoe gebeurt de neerlegging?

## VII. HET AUTEURSRECHT: EEN OBSTAKEL VOOR DE ARCHIVERING VAN WEBSITES

### A. INLEIDING

De digitalisering van het documentenverkeer heeft als logisch gevolg dat men deze documenten ook in digitale vorm gaat bewaren. In het vorige deel van dit rapport (deel IV) werd er uitvoerig uiteengezet welke technische problemen er zich allemaal voordoen bij het archiveren van een relatief nieuw soort digitaal document: websites. Er werden verschillende archiveringsstrategieën aangereikt waarvan de implicaties allemaal verschillend zijn: acquisitiebeleid, periodiciteit, hoe gaat men de website vastleggen voor archivering enz. De keuzes die door een organisatie gemaakt worden om een archiveringsstrategie uit te tekenen zullen grotendeels ingegeven worden door de bijzondere noden van deze organisatie.<sup>92</sup>

Bij het uittekenen van een archiveringsstrategie voor websites zal de organisatie ook rekening moeten houden met de juridische implicaties van het auteursrecht. Bibliotheken luiden reeds sinds geruime tijd de noodklok.<sup>93</sup> Het huidige auteursrecht biedt immers onvoldoende mogelijkheden om reproducties te maken van publicaties zonder de toestemming van de auteur. Niet alleen bibliotheken, maar alle instanties die te maken krijgen met de reproductie van informatie, ervaren het auteursrecht als een obstakel. Heel wat informatie wordt immers auteursrechtelijk beschermd, waardoor de reproductie ervan aan bepaalde voorwaarden is onderworpen. Vooral ten aanzien van websites tekent het probleem ten aanzien van het auteursrecht zich scherp af. Hoe immaterieel het Internet ook mag lijken, bewerkingen op Internet brengen bijna steeds met zich mee dat er kopies gemaakt worden van digitale gegevens. Surfen op het Internet, FTP, ‘uploading’ en ‘downloading’ impliceren eindeloze kopieerbewerkingen.<sup>94</sup>

Aanvankelijk werd cyberspace, vooral door de Internetgemeenschap, als een juridisch vacuüm beschouwd. Men ging ervan uit dat het bestaande rechtssysteem niet van toepassing zou zijn op Internet. Deze stelling was natuurlijk vooral ingegeven door het feit dat Internet geen landsgrenzen kent. De nieuwe ‘Internetruimte’ werd vaak vergeleken met de open zee, waarop geen enkele staat aanspraak kan maken.<sup>95</sup> Niets is minder waar. De inwoners van cyberspace zijn burgers van verschillende landen, elk met eigen rechten en plichten, waaronder ook de rechten en plichten van de auteurswetten dienen te worden begrepen.<sup>96</sup> De inhoudelijke bestanddelen van het Internet, nl. websites en hun inhoud ontsnappen dus niet aan de toepassing van het auteursrecht, enkel en alleen omdat zij aanwezig zijn op en verspreid worden via het Internet.

<sup>92</sup> Zie o.a. het deel over het acquisitiebeleid.

<sup>93</sup> ‘Auteurswet vormt hinderpaal voor bibliotheekarchief’, *NRC Handelsblad*, 3 maart 1998; PRINS, J.E.J., ‘Digitale duurzaamheid: een verloren geschiedenis?’, 3, <http://infolab.kub.nl/till/data/topic/digiduur.html>

<sup>94</sup> SPOOR, J., ‘The copyright approach to copying on internet: (over)stretching the reproduction right?’, in *The future of copyright in a digital environment*, HUGENHOLTZ, P. (ed.), Antwerpen, Kluwer, 1996, 67 en 69.

<sup>95</sup> BAINBRIDGE, D., *Introduction to Computer Law*, Gosport, Ashford Colour Press, 2000, 77; CLAPPAERT, W., ‘Auteursrecht en Internet’, in *Telecom en Internet. Recht in beweging*, DE POORTER, B. (ed.), Gent, Mys en Breesch, 1999, 359.

<sup>96</sup> WESTERBRINK, B.N., *Juridische aspecten van het Internet*, Amsterdam, Otto Cramwinkel, 1996, 88-89.

## B. AUTEURSRECHT IN EEN NOTENDOP

In België wordt het auteursrecht beheerst door twee reglementeringen: de Conventie van Bern van 9 september 1886 voor de bescherming van letterkundige en kunstwerken (goedgekeurd bij Wet van 26 juni 1951, *B.S.* 13 oktober 1951) en de Auteurswet van 30 juni 1994 betreffende het auteursrecht en de naburige rechten (*B.S.* 27 juli 1994, verder afgekort als ‘AW’). De Conventie van Bern heeft enkele minimum standaarden van auteursrechtelijke bescherming in het leven geroepen, die België in zijn auteurswet heeft overgenomen.

### B.1. Beschermde werken

#### a) Werken van letterkunde of kunst

De AW beschermt de rechten van de auteur ten aanzien van ‘werken van letterkunde of kunst’. De wet zelf geeft geen definitie van deze term. De Conventie van Bern geeft wel een niet-limitatieve omschrijving van voorbeelden van werken van letterkunde of kunst.<sup>97</sup> Uit deze opsomming blijkt alvast dat men veel meer heeft willen beschermen dan wat er volgens het dagelijkse taalgebruik onder deze term verstaan wordt.

Om te weten of een werk al dan niet onder de bescherming van het auteursrecht ressorteert, is het overigens niet belangrijk om te weten om welk soort werk het gaat.<sup>98</sup> Van zodra een werk in de ruimste zin van het woord voldoet aan een aantal intrinsieke kenmerken, is het een werk dat onder de bescherming van het auteursrecht valt. Het is in principe voldoende dat het werk *origineel* is en *in een bepaalde vorm* is gegoten.

‘Originaliteit’ wil niet zeggen dat de idee achter het werk nieuw moet zijn. Wanneer men een boek schrijft over digitale archivering, dan kan dit werk auteursrechtelijk beschermd zijn, ook al werden er al tientallen boeken over hetzelfde onderwerp geschreven. ‘Originaliteit’ betekent enkel dat het werk het resultaat moet zijn van een intellectuele inspanning van een mens, en een eigen, persoonlijke inbreng van de maker moet bevatten.<sup>99</sup> Een werk dat door middel van volledig machinaal- of technologie gestuurde procédés is tot stand gekomen, wordt dus niet beschermd. De mate van intellectuele inspanning is echter niet van belang; ook de kwaliteit van het werk is niet doorslaggevend.

<sup>97</sup> Art. 2.1 Conventie van Bern: ‘*De uitdrukking “letterkundige en kunstwerken” omvat alle voortbrengselen op het gebied der letterkunde, wetenschap en kunst, welke ook de wijze of de vorm van uitdrukken zij, zoals boeken, brochures en andere geschriften; voordrachten, toespraken, preken en andere werken van die aard; toneelwerken of dramatisch-muzikale werken, choreografische werken en pantomimes, waarvan de wijze van opvoering in schrift of op andere wijze is nedergelegd; muzikale composities met of zonder woorden; cinematografische werken en werken verkregen door een met cinematographie overeenstemmend procédé; werken van teken-, schilder-, bouw-, beeldhouw-, graveer- en lithographeerkunst; fotografische werken en werken verkregen door een met fotografie overeenstemmend procédé; werken van toegepaste kunsten; illustraties en aardrijkskundige kaarten; tekeningen, schetsen en plastische werken met betrekking tot de aardrijkskunde, de topografie, de bouwkunde of de wetenschappen.*’

<sup>98</sup> VOORHOOF, D., ‘Multimedia en auteursrecht. Afschermen en beschermen van informatie. Juridische problemen rond de beschikbaarheid en de reproductie van informatie op één drager’, in *Multimedia, Interactiviteit, Kennisspreiding*, verslagboek, Symposium van de Stichting Boek, 14 en 15 november 1995, LUC Diepenbeek, 100.

<sup>99</sup> Cass., 25 oktober 1989, *Pas.*, 1990, I, 239; Cass., 2 maart 1993, *Pas.*, 1993, 234.

Het werk moet bovendien uitgedrukt zijn in een zintuiglijk waarneembare vorm. Een idee of een concept wordt op zich dus niet beschermd. Wanneer men een idee heeft voor een boek dat men wil gaan schrijven (bijv. een scenario voor een spannend verhaal), dan zal men daarvoor geen auteursrechtelijke bescherming verkrijgen. Ideeën, hoe uniek ze ook mogen zijn, worden niet beschermd.<sup>100</sup>

b) En websites?

Het auteursrecht geldt zoals gezegd ook op het Internet. Het Internet is een wereldomspannend netwerk van computers (servers) waarop afgebakende gehelen van digitale documenten, zgn. websites, ter beschikking worden gesteld aan de hele internetgemeenschap. Een website is een geheel van gelinkte bestanden, waaronder tekstbestanden (al dan niet voorzien van hyperlinks), geluidsbestanden, databanken, scripts enz. Een website is een nieuw soort van kunstwerk dat tot enkele jaren geleden niet voorkwam. De auteur van een website wordt nochtans op dezelfde manier beschermd door het auteursrecht als de auteur van een klassiek literair werk.<sup>101</sup>

Wanneer we bovenstaande principes over originaliteit en waarneembaarheid toepassen op een website, dan moeten we vaststellen dat een website vanwege zijn karakter onder het regime van het auteursrecht kan vallen:

- Ook al zijn webpagina's geen tastbare documenten, dit doet niets af aan de toepassing van het auteursrecht. Om de auteursrechtelijke bescherming te kunnen genieten, volstaat het dat het werk is uitgedrukt in één of andere vorm, zelfs al is dat een immateriële vorm.
- Om de originaliteit van een website te beoordelen, zijn de opbouw en de samenstelling ervan de doorslaggevende criteria. Het is de wijze waarop informatie via een website voorgesteld wordt, die auteursrechtelijke bescherming geniet. Heel wat websites hebben ongeveer dezelfde indeling (bijv. welkom, voorstelling, contact, links...). Dit is gemeengoed geworden op het Internet en het overnemen van deze indeling maakt dan ook geen schending uit van het auteursrecht. Het overnemen van de wijze waarop deze indeling naar voren wordt gebracht, kan eventueel wel een inbreuk zijn op het auteursrecht. Websites worden vaak gebouwd door de combinatie van reeds bestaande 'formats', die de makers ervan op Internet plaatsen (let op: deze formats worden soms ook ten onrechte op het Internet gepubliceerd). Soms komen ze ook tot stand met behulp van computerprogramma's: in plaats van de code zelf te programmeren, klikt men op het scherm aan hoe de website er moet uitzien. In deze gevallen is er geen sprake van een eigen, originele creatie van de maker van de website.

Het is van belang om in te zien dat men een onderscheid moet maken tussen de website zelf als auteursrechtelijk beschermd werk, en de bestanddelen van de website, die elk op zich ook van de bescherming door het auteursrecht kunnen genieten, bijv. foto's, tekeningen, artikels, logo's, downloads... Bij publicaties op het Internet zijn er dus vaak verschillende auteursrechthebbenden: de

---

<sup>100</sup> TRIAILLE, J.P., 'La protection des idées. Les modes non contractuels de protection des idées en droit belge', *J.T.*, 1994, (797), 799.

<sup>101</sup> CLAPPAERT, W., 'Auteursrecht en Internet', in *Telecom en Internet. Recht in beweging*, DE POORTER, B. (ed.), Gent, Mys en Breesch, 1999, 361; SPOOR, J., *o.c.*, 69.

maker van de foto, de schrijver van het artikel, de ontwerper van het logo, de auteur van het rapport dat op de website kan gedownload worden, de ontwerper van de website (voorzover de website blijkt geeft van enige originaliteit) enz.

Of een website blijkt geeft van voldoende originaliteit, is een feitenkwestie die de archivaris moet beoordelen naar aanleiding van de archivering ervan. In laatste instantie zal eventueel de rechter hiervoor moeten oordelen.

c)        Hyperlinks

Vaak brengt de maker van een website een verzameling van koppelingen aan op de eigen website. Deze zgn. *hyperlinks* zijn visuele elementen op een website, die aanvankelijk gebruikt werden als een middel om wetenschappelijke teksten te publiceren. Door te klikken op een *hyperlink*, wordt de URL die erachter schuil gaat automatisch in het browservenster geladen. Op die manier kan de gebruiker snel naar andere verwante of nuttige websites surfen. Zo wordt het web van het Internet geweven. Het plaatsen van een hyperlink naar een andere website is niet verboden.<sup>102</sup> Men kan dit vergelijken met het plaatsen van een voetnoot in een tekst, waarmee dan bedoeld wordt dat er in de bron waarnaar verwezen wordt meer informatie over het onderwerp kan gevonden worden. Verwijzen naar plaatsen waar auteursrechtelijk beschermde werken zich bevinden, maakt geen schending uit van het auteursrecht. Men moet natuurlijk wel voorzichtig zijn met verwijzingen naar sites waar auteursrechtelijk beschermde werken ten onrechte worden toegankelijk gemaakt (bijv. illegaal gekopieerde en aangeboden software). Een verzameling van hyperlinks die op een website wordt geplaatst, is op zichzelf auteursrechtelijk beschermd op voorwaarde dat de verzameling voldoende origineel is bijv. door de ordening of door de volledigheid ervan. De inhoud van sommige portaalsites blinkt werkelijk uit door originaliteit.

d)        Uitzondering voor officiële akten van de overheid

Art. 8 §2 AW bepaalt dat de officiële akten van de overheid buiten het auteursrecht vallen. De teksten van wetten, de verslagen van de werkzaamheden in parlementaire commissies, vonnissen en arresten e.d. zijn dus vrij bruikbaar. Andere overheidspublicaties die door de werknemers van de overheid tot stand worden gebracht, maar niet kunnen gekwalificeerd worden als ‘officiële akte’, vallen gewoon onder de auteursrechtelijke regels bijv. een overheidsbrochure die de toekomstige afschaffing van het Kijk- en Luistergeld toelicht of rapporten die door ambtenaren geschreven worden. Gelet op de opmerking hierboven, zal een website van de overheid, indien hij voldoende origineel is, in ieder geval steeds onder de bescherming van het auteursrecht vallen. De website zelf is immers geen officiële

---

<sup>102</sup> Volgens sommigen is het plaatsen van een ‘link’ naar auteursrechtelijk beschermde informatie een vorm van secundaire openbaarmaking waarvoor de toestemming van de auteur is vereist: BRUNT, G., ‘Auteursrecht in DigiWorld’, in *Intellectueel eigendom in digitaal perspectief*, GIJATH, S. e.a. (ed.), Alphen aan den Rijn: Samson, 1996, 35-36, aangehaald door HUGENHOLTZ, P., *Recht en Internet*, Handelingen Nederlandse Juristen-Vereniging, Deventer, Tjeenk Willink, 1998-I, 212, voetnoot 36.

publicatie, maar een auteursrechtelijk beschermd werk dankzij zijn unieke en door de ontwerper geïnspireerde opbouw.<sup>103</sup>

## B.2. Wie is drager van de auteursrechten?

De vraag wie als auteur wordt aangemerkt van een auteursrechtelijk beschermd werk is belangrijk. Deze persoon verkrijgt immers de auteursrechten hierop. De AW stelt uitdrukkelijk dat het de persoon is die het werk daadwerkelijk tot stand brengt of bracht. Er geldt een weerlegbaar vermoeden dat diegene wiens naam op het werk wordt vermeld, de auteur is.<sup>104</sup>

Organisaties en overheden die hun aanwezigheid op het Web willen verzekeren door een eigen website, laten de ‘bouw’ ervan vaak over aan een gespecialiseerd bedrijf. Wie de opdracht geeft om een website te laten ontwikkelen, moet er rekening mee houden dat de ontwikkelaar principieel alle op de website rustende auteursrechten verkrijgt. Dit is zelfs zo wanneer de ontwikkelaar voor zijn diensten betaald wordt.<sup>105</sup> De opdrachtgever riskeert dus dat het ontwerp dat voor zijn website werd gebruikt, doorverkocht wordt aan anderen. Om dit te vermijden kan men een contract afsluiten met de ontwikkelaar waarin deze zijn auteursrechten afstaat. Dit moet dan wel gebeuren volgens de wettelijk voorgeschreven regels.<sup>106</sup>

Dezelfde regels gelden ook wanneer een werknemer of een ambtenaar de website van de organisatie tot stand brengt in de uitvoering van de taken die uit zijn arbeidsovereenkomst of uit het statuut voortvloeien. Wil de werkgever of de administratie toch de auteursrechten zelf verwerven, dan zal dit schriftelijk moeten vastgelegd worden. Dit kan eventueel vooraf gebeuren in de arbeidsovereenkomst of in het statuut.

Art. 6 lid 1 stelt uitdrukkelijk dat enkel een mens, natuurlijke persoon, auteur kan zijn. Dit is ook logisch aangezien enkel een mens de intellectuele inspanning kan leveren die voor de toepasselijkheid van het auteursrecht vereist wordt. Een rechtspersoon (organisatie, overheidsinstelling...) kan dus in beginsel niet de auteur zijn van de eigen website. De auteursrechten berusten dan bij de natuurlijke persoon die de organisatie of de overheidsinstelling vertegenwoordigt.

De vraag wie de drager is van de auteursrechten op een website, en op de inhoud van de website (artikels, tekeningen enz.) is van belang voor de archivaris. Het is immers deze persoon die zijn toestemming moet geven voor reproductiehandelingen, wijzigingen en openbaarmakingen. Op een website die van enige originaliteit getuigt, en bijgevolg door het auteursrecht beschermd wordt, staat vaak aangegeven wie verantwoordelijk is voor het webdesign. Hieruit blijkt echter niet noodzakelijk

<sup>103</sup> De meeste teksten die op de website van de overheid gepubliceerd worden, zullen meestal ook niet als een officiële akte kunnen aangemerkt worden, ook al gaat het om wetgeving of rechtspraak. Disclaimers op overheidswebsites geven immers vaak aan dat de gepubliceerde gegevens niet noodzakelijk de exacte weergave is van de officieel goedgekeurde tekst of dat enkel de wetgeving die in de papieren versie van het Belgisch Staatsblad is bekendgemaakt, rechtsgeldig is. Zie: [http://portal.vlaanderen.be/http://docs.portal.vlaanderen.be/channels/hoofdmnu/afw\\_van\\_aansprakelijkheid.doc](http://portal.vlaanderen.be/http://docs.portal.vlaanderen.be/channels/hoofdmnu/afw_van_aansprakelijkheid.doc)

<sup>104</sup> Art. 6 lid 2 AW.

<sup>105</sup> VANHEES, H., *Auteursrecht in een notendop*, Leuven, Garant, 1998, 24.

<sup>106</sup> De overdracht behoort o.a. schriftelijk te gebeuren voor bewijdsdoeleinden. Zie art. 3 §3 AW.

dat de auteursrechten bij deze persoon berusten.<sup>107</sup> De auteursrechten<sup>108</sup> kunnen immers zijn overgedragen aan een natuurlijke persoon/werkgever-bestuurder-ambtenaar van de organisatie of van de overheid die titularis is van de domeinnaam. De archivaris contacteert dus best eerst de titularis, om te kunnen checken bij welke natuurlijke persoon de auteursrechten berusten. Op de website van DNS kan je nagaan wie het domein van bijvoorbeeld “antwerpen.be” heeft geregistreerd en zijn diens contactgegevens beschikbaar. Bij deze persoon kan je als archivaris nagaan bij wie de auteursrechten op de website berusten.

Archiefdiensten van openbare besturen beperken zich in vele gevallen niet tot de archivering van de websites van de eigen organisatie. Het Stadsarchief Antwerpen bijvoorbeeld archiveert ook websites van andere organisaties die een bron zijn voor historisch onderzoek. Voor de toepassing van het auteursrecht maakt het echter geen verschil of de archivaris een website van de eigen organisatie wenst te archiveren, dan wel een website van een andere organisatie. Zelfs voor de websites van, in dit geval de stad Antwerpen, zal het Stadsarchief strikt genomen de toestemming van de auteursrechthebbende moeten vragen alvorens tot archivering ervan te mogen overgaan.

### B.3. Wat houden de auteursrechten in voor de auteursrechthebbende?

Een auteur bezit twee soorten rechten: vermogensrechten en morele rechten. De vermogensrechten stellen een auteur in staat om zijn werk te exploiteren en om uit dit werk inkomsten te halen. De morele rechten beschermen daarentegen eerder de persoon van de auteur; zij willen de “intieme band” tussen de auteur en zijn werk beschermen.<sup>109</sup> Het feit dat de auteur contracten heeft gesloten over de economische exploitatie van zijn werk, doet niets af aan de bescherming van het werk op moreel vlak.

#### a) Vermogensrechten

Het belangrijkste vermogensrecht is het *reproductierecht*.<sup>110</sup> Het reproductierecht houdt in dat de auteur het absolute en exclusieve recht heeft om zijn werk te reproduceren of te laten reproduceren. Onder reproduceren verstaat men ‘verveelvoudigen’, of met andere woorden ‘het tot stand brengen van materiële exemplaren’. Materieel wil hier niet zeggen dat de reproductie tastbaar moet zijn, wel dat ze op de éne of andere manier moet vastgelegd zijn. Het louter visualiseren van een werk op een computerscherm kan niet als een reproductiehandeling gekwalificeerd worden. Het opslaan in het computergeheugen is wel een vorm van reproductie.<sup>111</sup> De techniek die gebruikt wordt, speelt geen rol. Het kan gaan om een grafische, een mechanische, een optische, een elektronische... reproductie.

<sup>107</sup> Zie verder: hoe verkrijgt men auteursrechtelijke bescherming. De naam van de webdesigner staat doorgaans op de website vermeld uit commerciële overwegingen.

<sup>108</sup> NL. de vermogensrechten: zie verder.

<sup>109</sup> CORBET, J., ‘Auteursrecht’, in *Algemene Praktische Rechtsverzameling*, Antwerpen, Kluwer, 1997, 54, nr. 143.

<sup>110</sup> Art. 1 §1 lid 1 AW.

<sup>111</sup> CORBET, J., l.c., nr. 115.

Voor iedere reproductie van een auteursrechtelijk beschermd werk, is de toestemming van de drager van de vermogensrechten vereist, behoudens de wettelijk voorziene uitzonderingen.<sup>112</sup> Dit is in principe de maker van het werk, tenzij de vermogensrechten werden overgedragen aan een andere persoon. Het doel van de reproductie speelt geen rol. Zelfs wanneer dit gebeurt met de bedoeling om het werk te archiveren, is in principe de toestemming van de auteur nodig.

Het reproductierecht omvat volgens art. 1 AW ook het ‘exclusieve recht om toestemming te verlenen tot het bewerken of vertalen van het werk’. Het adaptatie- en vertalingsrecht toont aan dat ook ingrijpende wijzigingen bij het reproduceren onder het meesterschap van de auteur blijven, zolang de vorm van het oorspronkelijke werk herkenbaar blijft.<sup>113</sup>

Een ander voor het archiveren van websites belangrijk vermogensrecht, is het recht van de auteur om zijn werk volgens ongeacht welk procédé aan het publiek mee te delen.<sup>114</sup> Een auteur kan dus vrij beslissen of hij zijn werk zal op- of uitvoeren of zal laten op- of uitvoeren. Onder het begrip ‘op- of uitvoeren’ of ‘mededelen’ moet elke reproductie verstaan worden in een niet-duurzame en kortstondige vorm, zoals het zingen van een lied of het vertonen van een film.<sup>115</sup> Van een publieke mededeling is er sprake wanneer er tussen de aanwezigen bij een op- of uitvoering geen intieme band bestaat.

Kan de ontwerper van een website (of diegene aan wie de vermogensrechten werden overgedragen) zich nu ook beroepen op dit exclusieve recht om het werk aan het publiek mee te delen? De ruime bewoordingen waarin het mededelingsrecht werd geformuleerd, laten toe om de nieuwe technologische mededelingstechnieken onder te brengen in het auteursrecht:

Mededelingen van een werk kunnen op een persoonlijke manier gebeuren doordat het werk ‘live’ wordt gebracht door uitvoerende kunstenaars voor een lijfelijk aanwezig publiek. Hiervoor wordt meestal de term ‘op- of uitvoering’ gebruikt. De wetgever heeft echter de term ‘mededeling’ gekozen om duidelijk te maken dat het gaat om meer dan op- of uitvoeringen, en eveneens onpersoonlijke mededelingen kan omvatten zonder dat het publiek lijfelijk aanwezig is.<sup>116</sup> Het bekijken van een website gebeurt via elektronische transmissie van digitale informatie (HTML pagina’s) die op individuele vraag bij de eindgebruiker terecht komt.

Bij het consulteren van een website is het ‘publiek’ ook niet gelijktijdig en op dezelfde plaats aanwezig. Onder de AW kunnen echter individuele ontvangers van een mededeling ook dan samen een publiek vormen wanneer zij de meegedeelde boodschap niet gelijktijdig waarnemen.<sup>117</sup> Het WIPO Copyright Treaty van 1996 stelt dit letterlijk in zijn art. 8: ‘*Authors of literary and artistic works shall enjoy the exclusive right of authorising any communication to the public of their works, by wire or*

<sup>112</sup> De exclusieve vermogensrechten van de auteur gelden niet absoluut. Er bestaan een aantal uitzonderingen o.a. voor de reproductie van geluids- en audiovisuele werken in en voor familiekring (home-taping, art. §1 5° AW) en voor het kopiëren voor privé-gebruik of voor didactische gebruik (de zgn. reprografieregeling, art. 22 §1 4° AW). Archiefinstellingen kunnen evenwel geen gebruik maken van deze dwanglicenties voorzien in de wet, om reproducties te maken van websites (of van auteursrechtelijk beschermde werken in het algemeen) met de bedoeling om deze te archiveren: *Auteursrechten, privacy en royalties in musea, archieven en documentatiecentra*, Handelingen van het symposium gehouden te Brussel op 23.11.97 en op 15.12.97, Vlaamse museumvereniging, Werkgroep Auteursrechten, CUYPERS, J. (ed.), Antwerpen, 1999, 53 en 60.

<sup>113</sup> GOTZEN, F., *Auteursrecht, tekeningen en modellen*, KUL Faculteit Rechtsgeleerdheid, 1998, 67.

<sup>114</sup> Art. 1 §1 laatste lid AW: ‘*Alleen de auteur van een werk van letterkunde of kunst heeft het recht om het werk volgens ongeacht welk procédé aan het publiek mede te delen.*’

<sup>115</sup> VANHEES, H., o.c., nr.63.

<sup>116</sup> Verslag De Clerck, *Gedr. St.*, Kamer, B.Z., 1991-92, nr. 473/33, 64.

<sup>117</sup> GOTZEN, F., o.c., 82.

wireless means, including the making available to the public of their works in such a way that members of the public may access these works from a place and at a time individually chosen by them'.<sup>118</sup>

Verder is het ook nog maar de vraag of een website ooit daadwerkelijk zal bekeken worden. De rechtspraak heeft hierover gezegd dat het gaat om de *mogelijke* toegang voor meerdere mensen.<sup>119</sup>

**Opmerking:** Men heeft niet alleen de toestemming nodig van de ontwerper voor het ter beschikking stellen van een website via Internet<sup>120</sup> (bijv. het geval van een organisatie die een website laat ontwerpen door een bureau voor webdesign), maar ook voor reproductiehandelingen nl. wanneer de eindgebruiker (bijv. de archivaris) op het moment van het binnenhalen van de HTML-pagina's overgaat tot het realiseren van een kopie die dan in het digitale archiveringssysteem zal terecht komen.

→ Het aanbieden van informatie via het Internet is een totaal nieuwe vorm van mededeling<sup>121</sup> die dank zij de ruime bewoordingen van de AW onder het mededelingsrecht valt.

Naast het reproductierecht en het publiek mededelingsrecht, beschikt de auteur nog over enkele andere vermogensrechten<sup>122</sup>, die we hier niet verder behandelen, aangezien ze niet van belang zijn in de discussie rond de archivering van websites (en van digitale documenten in het algemeen).

#### b) Morele rechten

In tegenstelling tot de vermogensrechten, kunnen de morele rechten van de auteur niet te gelde worden gemaakt, noch volledig worden afgestaan. Zij worden *morele* rechten genoemd omdat zij de band tussen de auteur en zijn werk beschermen. In een auteursrechtelijk beschermd werk komt immers de persoonlijkheid van de auteur tot uiting.

De AW verleent vier morele rechten aan de auteur: het recht om het werk (voor het eerst) bekend te maken, het recht om het vaderschap van het werk op te eisen, het recht op eerbied voor het werk (art. 1 §2 AW) en het recht op toegang tot het werk (art. 3 §1 derde lid, *in fine* AW).

Het voor digitale archivering zeer belangrijke recht op eerbied voor het werk, of het 'recht op de integriteit van het werk' zoals het ook wel vaak wordt genoemd, maakt het voor de auteur mogelijk om zich te verzetten tegen *elke* wijziging van het werk, zonder dat hij daarbij enige schade moet aantonen.<sup>123</sup>

<sup>118</sup> <http://clea.wipo.int/PDFFILES/English/WO/WO033EN.PDF>

<sup>119</sup> Rb. Brussel, 16 oktober 1996, *Auteurs en Media*, 1996, 426.

<sup>120</sup> Het in eerste orde aanbieden via het Internet van een op bestelling gemaakte website zal uiteraard meestal geen problemen opleveren omdat een website nu eenmaal bedoeld is om via dit medium verspreid te worden. Voor alle zekerheid wordt het mededelingsrecht toch best schriftelijk overgedragen aan de opdrachtgever.

<sup>121</sup> Die niet valt onder de bijzondere bepalingen die de AW heeft ingesteld voor drie vormen van mededeling nl. de tentoonstelling (art. 9 AW), de satellietomroep (art. 48 e.v. AW) en de kabel distributie (art. 51 e.v. AW).

<sup>122</sup> Nl. het bestemmingsrecht, het verhuur- en uitleenrecht en het tentoonstellingsrecht.

<sup>123</sup> De Belgische auteurswet is op dit punt strenger dan de Conventie van Bern, die slechts de wijzigingen verbiedt die afbreuk doen aan de reputatie of de eer van de auteur (art. 6bis tekst Parijs 1971).

Met betrekking tot archivering rijst de vraag in welke mate de archivaris auteursrechtelijk beschermde websites (en de achterliggende broncode) mag wijzigen met de bedoeling om ze statisch te maken of om ze bestand te maken tegen technologische veroudering (zie verder).

#### B.4. Hoe verkrijgt men auteursrechtelijke bescherming?

Het tot stand brengen van een werk dat voldoet aan de voorwaarden voor auteursrechtelijke bescherming, brengt automatisch met zich dat het werk auteursrechtelijk beschermd is. In België moeten er geen formaliteiten vervuld worden voor het verkrijgen van auteursrechtelijke bescherming. Het aanbrengen van auteursrechtvoorbehoudtekens (bijv. een copyright notice ‘©’, gevolgd door de naam van de rechthebbende) is van geen enkele invloed op de toepasselijkheid van het auteursrecht.<sup>124</sup> Het gebrek aan auteursrechtelijke informatie op een website heeft dus niet voor gevolg dat de website niet auteursrechtelijk beschermd zou zijn en dat de archivaris er een kopie mag van maken zonder voorafgaandelijke toestemming.

### C. ARCHIVERING VAN WEBSITES: PROBLEEMSTELLING

Eén van de argumenten die aan het bestaan van het auteursrecht ten gronde ligt, is de gedachte dat het auteursrecht het culturele erfgoed van een land moet beschermen. Door exclusieve rechten toe te kennen aan auteurs van originele werken, worden ze geïnspireerd om in de toekomst nog andere werken tot stand te brengen. Op die manier dragen ze bij aan het gemeenschappelijke culturele erfgoed.<sup>125</sup>

Digitale creaties worden zoals gezegd net zo goed beschermd door het bestaande auteursrecht als creaties op papier, zoals boeken, pamfletten e.d. In het huidige auteursrecht schuilt er een enorme paradox met betrekking tot het archiveren van websites, en van digitale auteursrechtelijk beschermde werken in het algemeen. Reproductiehandelingen zijn inherent aan het archiveren van digitale documenten, maar het zijn precies deze handelingen die het auteursrecht voorbehoudt aan de auteur. Zonder toestemming van de auteur (van de website en/of van de inhoud van de website) is het dus in principe onmogelijk om websites digitaal te archiveren. Het zijn echter niet alleen reproductiehandelingen die bij gebrek aan toestemming van de auteur conflicteren met het auteursrecht, zoals we verder zullen zien.

Ook voor de preservatie van het papieren erfgoed zijn reproductiehandelingen (nl. kopiëren op microfilm) essentieel op het moment dat het papier dreigt te vergaan.<sup>126</sup> Musea en bibliotheken mogen

---

<sup>124</sup> In België bestaat er wel een regeling inzake ‘wettelijk depot’, maar deze regeling heeft niets te maken met auteursrecht. De wetgeving op het wettelijk depot verplicht de neerlegging in de Koninklijke Bibliotheek van België van de in België uitgegeven publicaties en van de in het buitenland uitgegeven publicaties waarvan de auteur of een van de auteurs Belg is en in België gedomicileerd is. (Wet 8 april 1965, *B.S.* 18 juni 1965).

<sup>125</sup> HUGENHOLTZ, P., *DIPPER Digital Intellectual Property Practice Economic Report – Copyright aspects of caching*, Institute for Information Law, University of Amsterdam, September 1999, 7.

<sup>126</sup> Het auteursrecht blijft gedurende zeventig jaar bestaan na het overlijden van de auteur ten voordele van zijn erfgenamen. Ook ten aanzien van papieren documenten zal de archivaris het auteursrecht dus als een obstakel ervaren, omdat de auteursrechten nog minstens zeventig jaar blijven gelden na de totstandkoming van het werk.

de elektronische publicaties waarover zij beschikken niet zomaar naar andere informatiedragers kopiëren wanneer blijkt dat de informatiedragers onderhevig zijn aan technologische veroudering. Archieven nemen echter een bijzondere positie in met betrekking tot de auteursrechtelijke problemen van preservering. Voor de archivering van digitale documenten, en met name voor websites stelt het probleem zich veel frequenter en veel vroeger, omdat de archivaris eerst een kopie moet maken van het stuk om in het bezit te komen ervan. De auteursrechtelijke problematiek beperkt zich dus niet tot digitale duurzaamheid.

In de rechtsleer werden er verschillende pogingen gedaan om een ‘recht op informatie’ te verdedigen ten aanzien van instellingen of personen die informatie zoeken en daarvoor toegang willen krijgen tot die informatie zonder toestemming van de auteur (zoals archieven of een verzameling kunstwerken).<sup>127</sup> Op Europees niveau bestaan er enkele bepalingen die in de richting van een recht op informatie lijken te gaan. Zo is er bijvoorbeeld art. 3 bis lid 1 van de Richtlijn over de uitoefening van televisie- en omroepactiviteiten<sup>128</sup>, dat aan de lidstaten toelaat om te bepalen dat omroepuitzendingen van bepaalde evenementen die van aanzienlijk belang voor de samenleving zijn (lees: sportevenementen) vrij en kosteloos voor het publiek toegankelijk moeten zijn. De toelating om een dergelijk recht op informatie in te voeren, lijkt gezien het opschrift van de richtlijn enkel betrekking te hebben op televisie, met uitsluiting van Internet.<sup>129</sup> Instellingen van algemeen belang, zoals publiekrechtelijke archiefinstellingen worden dus geconfronteerd met een auteursrecht dat de rechten vooral bij de auteur situeert, en veel te weinig rekening houdt met de problemen waarmee de tussenpersonen geconfronteerd worden, die de informatie bewaren en ter beschikking stellen.

Hieronder gaan we na welke problemen het auteursrecht in concreto oplevert met betrekking tot het archiveren van websites.

### C.1. Reproductiehandelingen

Men kan zich afvragen of het downloaden door de archiefinstelling van een website van de webserver een reproductiehandeling uitmaakt waarvoor de toestemming van de auteur vereist is.

Wanneer de archivaris de te archiveren webpagina opvraagt, laadt hij de inhoud daarvan in het werkgeheugen van zijn computer. Het laden van webpagina's noemt men ‘browsen’. Niemand twijfelt eraan dat dit browsen gepaard gaat met reproductiehandelingen.<sup>130</sup> Er bestaat onder juristen eensgezindheid over het feit dat de opslag van een beschermd werk op een digitaal medium als een

<sup>127</sup> HUGENHOLTZ, P., *Recht en Internet*, Nederlands Juristenblad, Zwolle, Tjeenk Willink, 1998, 201; DOMMERING, P., ‘Internet: een juridische plaatsbepaling van een nieuw communicatieproces’, in *Volatilisering in de economie*, DE JONG, W., Wetenschappelijk Raad voor het Regeringsbeleid, Den Haag, SDU Uitgevers, 1997, 150, aangehaald door BUYDENS, M., *Auteursrechten en Internet – problemen en oplossingen voor het creëren van een online databank met beelden en/of tekst*, Brussel, Federale diensten voor wetenschappelijk, technische en culturele aangelegenheden, 1998, 18.

<sup>128</sup> Richtlijn 97/36/EG van het Europees Parlement en de Raad van 30 juni 1997 tot wijziging van Richtlijn 89/552/EEG van de Raad betreffende de coördinatie van bepaalde wettelijke en bestuursrechtelijke bepalingen in de lidstaten inzake de uitoefening van televisie- en omroepactiviteiten, Pb. EG nr. L 202, 3.7.1997, 60.

<sup>129</sup> BUYDENS, M., *Auteursrechten en Internet – problemen en oplossingen voor het creëren van een online databank met beelden en/of tekst*, Brussel, Federale diensten voor wetenschappelijk, technische en culturele aangelegenheden, 1998, 18.

<sup>130</sup> HUGENHOLTZ, P., *Recht en Internet*, 211.

reproductie in de zin van de auteurswet moet gekwalificeerd worden.<sup>131</sup> In het auteursrecht mag de term ‘reproductie’ echter niet als een technisch concept opgevat worden. Het is in de eerste plaats een juridische notie.<sup>132</sup> Deze juridische notie gaat niet zo ver dat *iedere* reproductie in technische zin een reproductie is in de zin van de wet. De *Legal Advisory Board* van de Europese Commissie stelt in zijn reactie op de *Green Paper on Copyright in the Information Society*:

*“The notions of ‘reproduction’ (...) are only fully understood if they are interpreted not as technical, but as normative (man-made) notions, i.e. they are not in a simple sense descriptive but purpose-oriented and used to define and delimit existing proprietary rights in a sensible and acceptable way. Thus, if the use of a protected work transmitted over a computer network causes (parts of the work) to be intermediately stored, this technical fact does not, in itself, justify the conclusion that an exclusive reproduction right is potentially infringed.”*<sup>133</sup>

Dit is wat men noemt de normatieve interpretatie van het begrip ‘reproductie’. Het loutere ‘consumeren’ van informatie werd traditioneel buiten het auteursrecht gehouden.<sup>134</sup> Browsen op Internet kan men vergelijken met het bladeren in een encyclopedie. Diegene die bladert, verricht geen exploitatiehandeling, maar slechts een gebruikshandeling. Er is dus geen sprake van een reproductie waarvoor de toestemming van de auteur vereist is, wanneer het maken van een kopie van de betreffende website nodig is om deze website aan anderen te kunnen communiceren.<sup>135</sup> Dit wordt bevestigd in de nieuwe Europese richtlijn betreffende het auteursrecht.<sup>136</sup> Art. 5 lid 1 van de richtlijn voorziet een door de lidstaten verplicht in te voeren uitzondering op het reproductierecht voor tijdelijke reproductiehandelingen die van voorbijgaande of incidentele aard zijn, en die een integraal en essentieel onderdeel vormen van een technisch procédé en die worden toegepast met als enig doel de doorgifte in een netwerk tussen derden via een tussenpersoon mogelijk te maken. Overweging 23 van de richtlijn suggereert dat browsen onder art. 5 lid 1 begrepen wordt.<sup>137</sup>

Deze uitzondering is niet van toepassing op archivering. Zowel de statische als de dynamische websites moeten door middel van het maken van kopies van de webserver gehaald worden door de archiefinstelling. Deze kopies zijn niet tijdelijk van aard. Integendeel, de bedoeling is dat ze voor onbepaalde tijd in het digitale archiveringssysteem worden opgenomen. De kopie wordt dus niet louter

<sup>131</sup> Dit wordt trouwens in de Conventie van Bern bevestigd met niet zo veel woorden: art. 9 (Tekst Parijs 1971): ‘*Authors of literary and artistic works protected by this Convention shall have the exclusive right of authorizing the reproduction of these works in any manner or form.*’

<sup>132</sup> HUGENHOLTZ, P., *DIPPER Digital Intellectual Property Practice Economic Report – Copyright aspects of caching*, 14.

<sup>133</sup> <http://europa.eu.int/ISPO/legal/en/ipr/reply/reply.html>

<sup>134</sup> HUGENHOLTZ, P., ‘Convergence and Divergence in Intellectual Property Law: The Case of the Software Directive’, in *Information Law towards the 21<sup>st</sup> Century*, DOMMERING, E. (ed.), Deventer, Kluwer, 1992, 323.

<sup>135</sup> KOHLER, J., *Das Autorrecht*, Jena 1880, 230, aangehaald door SPOOR, J., *Scripta Manent: de reproductie in het auteursrecht*, Groningen, Tjeenk Willink, 1976, 137-138.

<sup>136</sup> Waarover verder meer. Richtlijn 2001/29/EG van het Europees Parlement en de Raad van 22 mei 2001 betreffende de harmonisatie van bepaalde aspecten van het auteursrecht en de naburige rechten in de informatiemaatschappij, <http://www.ivir.nl/wetten/eu/2001-29-EG.pdf>

<sup>137</sup> HUGENHOLTZ, P., *Recht en Internet*, 211, voetnoot 33, BUYDENS, M., ‘La nouvelle directive du 22 mai 2001 sur l’harmonisation de certains aspects du droit d’auteur et des droits voisins dans la société de l’information : le régime des exceptions’, *Auteurs en Media*, 2001/4, 434.

binnengehaald om de website lokaal te kunnen bekijken op het scherm. *Caching* zonder toestemming van de auteur zou nog te verantwoorden zijn. Het gaat dan immers om een hulpmiddel dat er in bestaat om frequent geraadpleegde websites geheel of gedeeltelijk, maar steeds tijdelijk op te slaan op de server van de eindgebruiker met als bedoeling de doorgifte van informatie via het Internet sneller te laten verlopen. De computer van de eindgebruiker moet dan geen verbinding meer maken met de originele server bij het opvragen van die websites. Het archiveren of op een andere manier permanent bewaren van websites is echter niet toegelaten zonder toestemming van de auteur.<sup>138</sup>

Wat het downloaden van de website voor archivering betreft, is er geen verschil tussen statische en dynamische websites. Beiden moeten door de archivaris gekopieerd worden om ervan in het bezit te komen. Het exclusieve reproductierecht is van toepassing telkens wanneer de beschermde werken die opgeslagen zijn op een digitaal medium op een webserver worden geplaatst of van een webserver worden gehaald.<sup>139</sup> Reproductiehandelingen voor archiveringsdoeleinden vereisen dus de toestemming van de rechthebbende.

Er zijn weliswaar enkele uitzonderingen op het exclusieve reproductierecht voorzien in de auteurswet.<sup>140</sup> Art. 22 §1 4° bepaalt o.a. dat de auteur zich niet kan verzetten tegen de gedeeltelijke of gehele reproductie van artikelen of van werken van beeldende kunst, of van korte fragmenten uit werken die op een grafische of soortgelijke drager zijn vastgelegd, wanneer die reproductie uitsluitend bestemd is voor privé-gebruik of voor didactisch gebruik en geen afbreuk doet aan de uitgave van het oorspronkelijke werk ( het zgn. *reprorecht*). Men kan zich afvragen of archiefinstellingen deze bepalingen niet als een ontsnappingsroute kunnen gebruiken om niet steeds de toestemming van de rechthebbende te moeten vragen wanneer zij auteursrechtelijk beschermde werken wensen te archiveren. De reproducties worden toch uitsluitend gemaakt voor *eigen gebruik* nl. om ze te archiveren.

Het is echter niet dit soort gebruik dat de wetgever met deze notie bedoeld heeft. In het Verslag van de Senaat wordt er over de notie “eigen gebruik” het volgende gezegd: “*Er is ook sprake van eigen gebruik met betrekking tot de reprografie: daar wordt het kopiëren voor eigen gebruik door een natuurlijke persoon (zelfs met professionele doeleinden) of voor intern gebruik door een rechtspersoon, van korte fragmenten van werken als wettig beschouwd.*”<sup>141</sup> Deze definitie is op zich zeer ruim, zeker als men haar vergelijkt met de interpretatie in de Franse auteurswet, waar enkel strikt persoonlijk gebruik door een natuurlijke persoon is toegelaten, en ieder collectief gebruik, bijv. in het kader van een rechtspersoon is uitgesloten. Toch zullen archiefinstellingen geen beroep kunnen doen op het *reprorecht* om reproducties te maken met de bedoeling om deze te archiveren. Er is immers geen sprake van persoonlijk gebruik wanneer de reproductie bestemd wordt om gebruikt te worden door een derde, zelfs al is dat gebruik dan persoonlijk.<sup>142</sup> Bij archivering is dat nochtans het geval. De reproductie wordt gemaakt opdat derden het archiefstuk in de toekomst nog zouden kunnen raadplegen, en niet voor interne consultatie door het personeel van de archiefinstelling. Het *reprorecht*

<sup>138</sup> HUGENHOLTZ, P., *DIPPER Digital Intellectual Property Practice Economic Report – Copyright aspects of caching*, 46.

<sup>139</sup> HUGENHOLTZ, P., *DIPPER Digital Intellectual Property Practice Economic Report – Copyright aspects of caching*, 9.

<sup>140</sup> Zie ook voetnoot 112.

<sup>141</sup> Verslag aan de Koning bij het K.B. betreffende de vergoeding verschuldigd aan auteurs en uitgevers voor het privé-gebruik of didactisch gebruik van werken die op grafische of soortgelijke wijze zijn vastgelegd, *B.S.* 7 november 1997, 29 877.

<sup>142</sup> DUBUISSON, F., ‘L’exception de reproduction d’œuvres fixées sur un support graphique ou analogue dans un but privé ou didactique’, *Journal des Tribunaux*, 1997, 657.

zou bijvoorbeeld wel gelden voor de reproductie van een auteursrechtelijk beschermd werk gemaakt door de archiefinstelling met de bedoeling om de interne werking van het archief te verbeteren, bijv. een kopie van een handleiding “digitaal archiveren”.

Hetzelfde geldt voor de notie “didactisch gebruik”. Met uitsluitend didactisch gebruik wordt de reproductie in het kader van de illustratie bij onderwijs of wetenschappelijk onderzoek bedoeld. Wanneer iemand een reproductie maakt voor de eigen studie- of onderzoeksbehoeften, is er persoonlijk gebruik van de kopie en dus privé-gebruik. Bedoeld is hier de situatie van bibliotheken die op verzoek van derden kopies maken bestemd voor gebruik bij studies of bij onderzoek. De finaliteit van de kopie is daarbij doorslaggevend. Een kopie gemaakt voor professioneel gebruik door een advocaat, is geen didactisch gebruik.<sup>143</sup> Een kopie gemaakt door een archiefinstellingen die eventueel voor didactische doeleinden zal gebruikt worden in de toekomst, maar eventueel ook niet (of voor andere doeleinden), is ook geen kopie voor didactisch gebruik.

Het reprorecht is in België bovendien veel enger geformuleerd dan in de andere Europese landen. Er bestaan twee moeilijkheden met het Belgische reprorecht ten aanzien van digitale archivering: 1) reproductie voor eigen gebruik is beperkt tot korte fragmenten, terwijl de meeste archieven enkel zullen geïnteresseerd zijn in de volledige werken; 2) de verkoop van originele werken zou kunnen geschaad worden, omdat niet-digitale archieven het bezit vereisen van verschillende originelen terwijl voor digitale archieven enkel het bezit van één origineel vereist is.<sup>144</sup> Het maken van reproducties voor het archiveren van digitale werken door archiefinstellingen kan bijgevolg niet gekwalificeerd worden als het maken van kopies voor eigen gebruik.

De bestaande uitzonderingen bieden dus geen ruimte voor reproductie ten behoeve van archivering zonder toestemming van de auteur.<sup>145</sup>

## C.2. Het aanbrengen van wijzigingen

Zoals in het technische deel werd beschreven, kunnen bij de archivering van dynamische websites niet de originele bestanden in hun oorspronkelijke bestandenstructuur gearcheveerd worden. In tegenstelling tot de statische websites, moet de broncode van dynamische websites vooraf gewijzigd worden. In de plaats van de originele ASP- of PHP-bestanden moeten statische HTML-bestanden gearcheveerd worden. Dit is de enige manier om ze achteraf op een platformafhankelijke manier te kunnen reconstrueren.<sup>146</sup> De omzetting gebeurt aan de hand van speciale daarvoor ontworpen computerprogramma's nl. off line browsers. Pagina's die al in HTML zijn opgemaakt van bij de oorsprong, hoeven geen wijzigingen te ondergaan.

<sup>143</sup> Hoe de bibliothecaris die de kopie maakt op verzoek, kan nagaan of de kopie werkelijk voor didactische doeleinden zal aangewend worden, is niet duidelijk. Het is een feitenkwestie waarover de bibliothecaris moet oordelen op eigen verantwoordelijkheid.

<sup>144</sup> DECKER, U., 'Electronic Archives and the Press: Copyright Problems of Mass media in the Digital Age', *E.I.P.R. issue 7*, Sweet & Maxwell Limited, 1998, (256), 257.

<sup>145</sup> LANGE, J., 'De wet en elektronische publikaties',  
<http://nieuws.surfnet.nl/nieuws/snn-archief/achtergrond/jg97-98/kb-symposium2.html>  
 PRINS, J.E.J., 'Digitale duurzaamheid: een verloren geschiedenis', 2-3,  
<http://infolab.kub.nl/till/data/topic/digiduur.html>

<sup>146</sup> Cf. supra.

Ook voor deze bewerkingen moet de toestemming van de auteur gevraagd worden conform het exclusieve wijzigingsrecht, voorzover de website als een originele creatie kan beschouwd worden. In het auteursrecht is zelfs de miniemste wijziging van het werk onderworpen aan de voorafgaande toestemming van de auteur. Hier moet men wel opletten: het wijzigingsrecht kan als moreel recht niet overgedragen worden aan de werkgever of de opdrachtgever.<sup>147</sup> De archivaris zal voor deze bewerkingen dus de toestemming moeten vragen aan de eigenlijke auteur. Men zou kunnen stellen dat het resultaat van een volledige omzetting van de website geen wijziging is van het werk, maar een adaptatie, gecreëerd door een computerprogramma. Er is sprake van een adaptatie als er een nieuwe originele website is tot stand gebracht die weliswaar gebaseerd is op de oorspronkelijke. Het is dan zelfs mogelijk dat deze adaptatie zelf een auteursrechtelijk beschermde creatie is, op voorwaarde dat de oorspronkelijke vorm van de website herkenbaar blijft en dat de nieuwe broncode voldoende origineel is. Bij archivering is het uiteraard zo dat men de oorspronkelijke vorm zoveel mogelijk zal willen behouden. De nieuw totstandgekomen website is dan wat men noemt een ‘computerwerk’.

Een ‘computerwerk’ (*computer-generated work*) is een werk dat het resultaat is van de interactie tussen mens en computer. In de mate waarbij de inbreng van de computer groter is, en de creativiteit van de mens minder is, werpen computerwerken problemen op in het licht van het auteursrecht. Auteursrecht wordt immers slechts verleend in de impliciete veronderstelling dat een werk de uitdrukking is van menselijke creativiteit.<sup>148</sup> De Anglo-Amerikaanse rechtstraditie heeft veel minder moeite met het fenomeen ‘computerwerk’ dan het continentale recht. Hoewel beide tradities de laatste decennia dichter naar elkaar zijn toegegroeid onder invloed van de digitale evolutie en het toetreden van de Verenigde Staten tot de Conventie van Bern, neemt men in de Anglo-Amerikaanse rechtstraditie toch vooral het werk zelf en het uniek-zijn ervan onder de loep. Het Verenigd Koninkrijk is dan ook het enige land in Europa dat in zijn auteurswet een specifieke bepaling heeft opgenomen voor werken die worden voortgebracht door een computer en waarvoor er geen menselijke auteur kan worden aangeduid. Rechthebbende is dan de (rechts)persoon die de investering heeft gedaan in de computer of in het computerprogramma waarmee het werk werd voortgebracht. In België is de bescherming van een autonome creatie door de computer nog ondenkbaar. Meestal zal er echter, zij het een minimale inbreng van de programmeur kunnen aangemerkt worden, waardoor de getransformeerde websites toch onder de paraplu van het auteursrecht vallen.

### C.3. Digitale duurzaamheid

We gaan hier uit van de veronderstelling dat de website destijds in statische vorm gearchiveerd werd.<sup>149</sup> Emulatie van de off line browsers wordt hierboven als de beste oplossing aangereikt. Ook vanuit auteursrechtelijk oogpunt is dit de beste keuze. Dankzij het gebruik van emulatoren hoeven de HTML bestanden op termijn geen wijzigingen te ondergaan.

---

<sup>147</sup> Cf. supra.

<sup>148</sup> DE COCK BUNING, M., *Auteursrecht en Informatietechnologie – Over de beperkte houdbaarheid van technologiespecifieke regelgeving*, Amsterdam, Otto Cramwinkel, 1998, 175.

<sup>149</sup> Cf. supra.

#### C.4. Ter beschikking stelling

Het exclusieve mededelingsrecht omvat zoals gezegd iedere vorm van openbaarmaking. In de praktijk zijn er twee manieren om de gearchiveerde websites toegankelijk te maken voor het publiek. Een mogelijkheid is dat de websites in de leeszaal vanaf een harde schijf of vanaf een externe drager worden bekeken. Een andere mogelijkheid is het ter beschikking stellen door het archief van een portaal site (bijv. een subpagina van de eigen site) van waaruit elke gearchiveerde website raadpleegbaar is. Uit vrees voor misbruik zouden auteurs minder gemakkelijk toestemming geven voor het toegankelijk houden van elektronische publicaties via het Internet. Daarom zijn elektronische publicaties in de Nederlandse elektronische bibliotheek enkel in het gebouw te raadplegen.

Men zou kunnen argumenteren dat dit probleem zich niet stelt voor de archivering van websites omdat de toestemming om de site, inclusief de inhoud, op Internet te publiceren in principe al gevraagd is geweest aan auteur(s) van de website en van de inhoud (of deze persoon zelf heeft beslist zijn site op Internet te plaatsen). Toch moet de archivaris hiervoor opnieuw de toestemming vragen. Het is immers niet omdat een auteur zijn toestemming geeft om zijn werk mee te delen aan het publiek via een bepaald medium, dat eender wie op eigen initiatief mag beslissen om dat werk via een ander medium te verspreiden, louter op grond van het feit dat het medium gekozen door de auteur nl. het Internet, potentieel iedereen ter wereld kan bereiken.<sup>150</sup> De archivaris kan deze toestemming laten opnemen in de overeenkomst die het reproductierecht van de site regelt.

De AW voorziet een uitzondering op het reproductierecht naar aanleiding van de raadpleging van werken voor wetenschappelijke doeleinden. Artikel 22 §1 4<sup>o</sup>ter AW bepaalt dat de reproductie van werken toegelaten is zonder toestemming van de auteur, wanneer die reproductie wordt verricht voor wetenschappelijk onderzoek, verantwoord is door de nagestreefde niet-winstgevende doelstelling en geen afbreuk doet aan de normale exploitatie van het werk. De wetgever viseert hier de situatie waarbij een onderzoeker zich in de archiefinstelling aanmeldt en voor het wetenschappelijk onderzoek dat hij verricht een kopie wil verkrijgen van een gearchiveerd werk. Deze kopie kan natuurlijk ook een diskette zijn waar het werk op gekopieerd is. Wanneer de kopie in hoofde van de aanvrager bestemd is voor wetenschappelijk onderzoek, kan de archiefinstelling gebruik maken van deze dwanglicentie voor het maken van reproducties.

#### C.5. Bijzondere regimes

Twee soorten van creaties hebben een eigen auteursrechtelijke regeling gekregen. Ze worden geregeld door een ‘bijzondere wet’. Het gaat om computerprogramma’s en databanken. Beiden kunnen in bepaalde gevallen deel uitmaken van een website.

##### a) Computerprogramma’s

In het technische deel werd er betoogd dat de meest recente generatie websites het best met een computerprogramma worden vergeleken. Naast de data, die als vaste tekst op het scherm verschijnen

---

<sup>150</sup> Voor databanken werd een uitzondering voorzien: cf. infra.

bij het inladen van de website en meestal in HTML zijn opgemaakt, komen in dit soort websites scripts voor. Scripts zijn microprogramma's die op de site zijn ontworpen en opgeslagen, en die naar de computer (hetzij van de client, hetzij van de server) worden gezonden om er uitgevoerd te worden. Het resultaat van deze bewerking hangt af van de opdracht die de client heeft gegeven, bijv. het oproepen van een tekst op het scherm.

De vraag is nu of dit soort websites ook als een computerprogramma in de zin van de softwarewet kunnen gekwalificeerd worden. Deze vraag is van praktisch belang voor de archivaris, omdat de softwarewet ervan uitgaat dat de auteursrechten op computerprogramma's bij de werkgever (dit kan ook een administratie zijn) liggen indien de site door een werknemer werd ontworpen, en niet bij de auteur zelf zoals voor andere auteursrechtelijk beschermde creaties het geval is volgens het gemene auteursrecht.<sup>151</sup> De archivaris zal zich dan ook tot de werkgever moeten richten om toestemming te vragen voor archivering. Voor publiekrechtelijke archiefinstellingen van de overheid is de vraag minder relevant omdat het statuut de werkgever als auteursrechthebbende zal aanduiden, ook voor de andere creaties die geen computerprogramma's zijn.

De bescherming van computerprogramma's zag er begin jaren tachtig verschillend uit in de Europese lidstaten. Slechts een aantal landen hadden op dat moment de bescherming van computerprogramma's expliciet in hun auteurswetgeving opgenomen.<sup>152</sup> De andere landen kenden geen specifieke bescherming voor computerprogrammatuur. Verontrust door de verschillen in de (toekomstige) wetgeving van de lidstaten met betrekking tot computerprogramma's, die de werking van de gemeenschappelijke markt voor computerprogramma's negatief beïnvloedden, publiceerde de Europese Commissie in 1988 een 'Groenboek' waarin de indiening van een richtlijn over de rechtsbescherming van computerprogramma's via het auteursrecht werd aangekondigd. Deze softwarerichtlijn, die er kwam in 1991, werd omgezet in Belgisch recht door de wet van 30 juni 1994 houdende omzetting in Belgisch recht van de Europese richtlijn van 14 mei 1991 betreffende de rechtsbescherming van computerprogramma's (= de softwarewet, B.S. 27 juli 1994).

De richtlijn, noch de softwarewet geven een definitie van het begrip 'computerprogramma'. Men is hierover bewust vaag gebleven, omdat een technologiespecifieke definitie zeer snel zou verouderen.<sup>153</sup> In andere juridische bronnen vinden we wel een definitie terug:

WIPO: 'een verzameling instructies die bestemd zijn om een informatieverwerkende machine een bepaalde functie te laten uitvoeren.'

Amerikaanse Copyright Act 1976: 'a computerprogram is a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result.'

Dat de scripts die deel uitmaken van de website computerprogramma's zijn in de zin van de softwarewet, is duidelijk gemaakt in de rechtsleer. De website op zich, is echter niet te kwalificeren als een computerprogramma, ook al bestaat de broncode van de website voor het grootste deel uit scripts, en ook al lijkt de website een computerprogramma te zijn dat "on the fly" webpagina's genereert. Een juridisch argument vinden we in de auteursrechtelijke richtlijn van 22 mei 2001. Overweging 23 van de

---

<sup>151</sup> Art. 3 Softwarewet.

<sup>152</sup> NL. Duitsland, Frankrijk en Groot-Brittannië (1985) en Spanje (1987).

<sup>153</sup> COM (88) 816 – SYN 183, *J.O.C.E.*, 12 april 1989, n° C 91/9.

richtlijn suggereert impliciet dat websites onder het tijdelijke reproductierecht vallen dat geregeld wordt door art. 2 van de richtlijn. Websites vallen bijgevolg onder het algemene regime van de auteursrechtlijn en niet onder de meer specifieke softwarerichtlijn.<sup>154</sup> De scripts zelf zijn uiteraard wel onderworpen aan het regime voor software.

#### b) Databanken

Veel dynamische websites zijn gekoppeld aan een bestandensysteem. Dit bestandensysteem bevat vaak een databank, waaruit de surfer allerlei gegevens kan opvragen. In het geval van de overheid kan het gaan om formulieren, officiële akten en in de toekomst misschien zelfs de eigen persoonsgegevens uit de bevolkingsregisters na de implementatie van de elektronische identiteitskaart in het kader van het e-loket.<sup>155</sup> Gaat het om het intranet van een bedrijf, dan kunnen er via de databank allerlei bedrijfsdocumenten beschikbaar gesteld worden.

Databanken zijn onderworpen aan een bijzondere auteursrechtelijke reglementering. Deze regeling is er gekomen naar aanleiding van een aantal rechtszaken die begin jaren negentig werden beslist. In de zaak *Van Dale/Domme* besliste de Nederlandse Hoge Raad dat een databank slechts auteursrechtelijk beschermd is op voorwaarde dat ze het resultaat is van een selectie die de persoonlijke visie van de maker tot uitdrukking brengt.<sup>156</sup> In dezelfde lijn van het klassieke auteursrecht besliste het *Supreme Court* in de *Feist-case* dat een verzameling van louter feitelijke gegevens (in casu een telefoongids) niet in aanmerking komt voor bescherming onder de copyrightwetgeving bij gebrek aan enige creativiteit.<sup>157</sup> Beide arresten toonden aan dat het klassieke auteursrecht niet voldoende rechtsbescherming bood aan de samenstellers van databanken. Eender wie kon immers databanken bestaande uit feitelijke gegevens op zijn eigen computer kopiëren, de gegevens herschikken en dan als eigen product op de markt brengen. Bijgevolg was niemand nog geneigd om investeringen te doen in databanken waarvan de inhoud zelf niet auteursrechtelijk beschermd is. Daarom heeft de Europese databankenrichtlijn<sup>158</sup> een specifieke bescherming *sui generis* ingevoerd, dit wil zeggen ongeacht of de databank zelf auteursrechtelijke bescherming geniet. De toestemming van de samensteller van een databank is sindsdien vereist voor het kopiëren en/of verspreiden van de inhoud van deze databank, ook al gaat het om een verzameling van feiten.

De toepasselijkheid van dit bijzondere regime heeft opnieuw praktische gevolgen voor de archivaris omdat de auteursrechten op de databank, net zoals voor computerprogramma's, in beginsel berusten bij de werkgever, tenzij anders is overeengekomen.

Welke bescherming biedt dit regime nu voor databanken en wat zijn daarvan de gevolgen voor de archivaris die een aan een databank gekoppelde website wil gaan archiveren?

<sup>154</sup> Hoewel het regime van de auteursrechtlijn ook van toepassing is op computerprogramma's voorzover er geen bijzondere bepalingen voorkomen in de softwarerichtlijn.

<sup>155</sup> Cf. infra.

<sup>156</sup> Hoge Raad, 4 januari 1991, *NJ* 1991, 608.

<sup>157</sup> U.S. Supreme Court, 27 maart 1991, *Informatierecht/AMI* 1991, 179.

<sup>158</sup> Deze richtlijn van 11 maart 1996 betreffende de rechtsbescherming van databanken werd omgezet in Belgisch recht door de wet van 31 augustus 1998 betreffende de rechtsbescherming van databanken (*B.S.* 14 november 1998), die in de auteurswet van 1994 en afdeling 4bis heeft ingevoerd met bijzondere bepalingen betreffende databanken. Afdeling 4 bis van de AW regelt de bescherming van databanken die zelf auteursrechtelijk beschermd zijn. Het recht *sui generis* van producenten van databanken is blijven staan in de wet van 31 augustus 1998.

Er moet een onderscheid gemaakt worden tussen twee beschermingsniveaus:

Wanneer de archivaris een kopie wil maken van de databank voor archiveringsdoeleinden, rijst vooreerst de vraag naar de auteursrechtelijke bescherming van het materiaal (de inhoud) waaruit de databank is samengesteld. ‘Werken’ die door het auteursrecht worden beschermd<sup>159</sup>, mogen enkel mits toestemming van de auteur in een databank worden opgenomen.<sup>160</sup> De archivaris moet dus bij de maker van de databank verifiëren of de inhoud terecht in de databank werd opgenomen. Zoniet, dan maakt de archivaris zichzelf schuldig aan een schending van het auteursrecht door de reproductie ervan.

Vervolgens moet hij zich de vraag stellen naar de auteursrechtelijke bescherming van de databank zelf. De databank zelf wordt beschermd voorzover de selectie van het materiaal (bijv. de volledigheid ervan) of de ordening van dat materiaal blijk geeft van creativiteit. Databanken die aan websites gekoppeld zijn om te voldoen aan een individuele vraag om informatie bijv. in het kader van e-government, zullen meestal niet uitblinken door de originaliteit van de selectie van het materiaal of door hun ordening. Deze toets zal de archivaris moeten doorvoeren. Wanneer de databank niet origineel blijkt te zijn, dan zal hij toch nog rekening moeten houden met het sui generis-recht van de producent van zodra de verkrijging, de controle of de presentatie van de inhoud getuigt van een substantiële investering. Naast de auteursrechtelijke bescherming van de structuur van de databank, is er immers ook voorzien in een extractierecht ( ‘*right to prevent extraction*’ ) dat een vorm is van investeringsbescherming.<sup>161</sup>

De auteur van een databank die de originaliteitstoets doorstaat, heeft overeenkomstig art.5 van de databankenrichtlijn het exclusieve recht op het verlenen van toestemming voor elke vorm van gehele of gedeeltelijke reproductie en voor iedere openbare verspreiding. De archivaris heeft bijgevolg de toestemming nodig van de auteur van de databank om de inhoud ervan tijdelijk of permanent over te brengen op een andere drager, ongeacht op welke wijze of in welke vorm dit gebeurt, evenals om de databank aan het publiek ter beschikking te stellen.

De richtlijn voorziet enkele uitzonderingen<sup>162</sup>, maar die laten niet toe om te ontsnappen aan de verplichting voor de archivaris om met de databankproducent een overeenkomst af te sluiten om van de via het Internet gepubliceerde databank een kopie te mogen maken voor opname in het archief.

Voor databanken die aan overheidssites zijn gekoppeld, kunnen we zeggen dat ze vaak niet auteursrechtelijk beschermd zullen zijn, waardoor er zich ook geen problemen voor de reproductie stellen. Hun inhoud bestaat vaak slechts uit interne informatie, waardoor bepaalde leden van de overheid zelf als auteurs van de inhoud kunnen doorgaan. De databank zelf zal in de meeste gevallen niet uitblinken door originaliteit of volledigheid, zeker gezien het feit dat het materiaal waar de databank uit bestaat, eigendom is van de overheid. Van een bescherming van de investeringen en het harde werk van de auteur van de databank kan er bijgevolg ook geen sprake zijn.

Databanken hebben een soepeler regime wat de terbeschikkingstelling door archiefinstellingen betreft. Tijdens de voorbereiding van de omzetting in Belgisch recht van de databankenrichtlijn werd

<sup>159</sup> Cf. supra

<sup>160</sup> Zie art. 20 bis AW: ‘... en laat de bestaande rechten op de werken, gegevens of andere elementen vervat in de databank onverlet.’

<sup>161</sup> Er wordt soms ook wel gezegd dat het extractierecht het zweet van de databankproducent beschermt: HUGENHOLTZ, P., ‘De databankrichtlijn eindelijk aanvaard: een zeer kritisch commentaar’, *Computerrecht*, 1996/4, 132.

<sup>162</sup> Art. 22 bis AW.

de problematiek van instellingen die informatie aanbieden die van belang kan zijn voor wetenschappelijk onderzoek, zoals bibliotheken en archieven, aan de kaak gesteld. Er werd voorgesteld om een uitzondering op te nemen in de databankenwet met betrekking tot de reproductie en mededeling van databanken voor wetenschappelijk doeleinden door instellingen die databanken kosteloos of onder bezwarende titel ter beschikking stellen van anderen. Met deze opmerkingen vanuit de sector werd er in de wet rekening gehouden en art. 22bis §1 2°-4° AW bepaalt nu dat de toestemming van de auteur niet nodig is om een databank mee te delen aan het publiek (bijv. terbeschikkingstelling via het Internet), noch om een databank te reproduceren, voorzover deze handelingen gebeuren ten behoeve van wetenschappelijk onderzoek. Gaat het om een databank waarvan de inhoud op zich niet auteursrechtelijk beschermd is, maar wel aan het extractierecht is onderworpen, dan moet er voor de terbeschikkingstelling toch de toestemming gevraagd worden, maar niet voor een gedeeltelijke reproductie.<sup>163</sup> Ook hier is er voorzien dat de databank zonder toestemming van de auteur mag gereproduceerd worden wanneer er iemand voor wetenschappelijke doeleinden een kopie vraagt.

## C.6. Besluit

Hoe moet de archivaris nu te werk gaan wanneer hij een website wil gaan archiveren. Er moeten verschillende stappen doorlopen worden:

**STAP 1:** De archivaris moet bij zichzelf nagaan of de website auteursrechtelijk beschermd is. Dit is een feitenkwestie die soms moeilijk te beoordelen is. Bij twijfel gaat hij of zij er best van uit dat het werk getuigt van enige originaliteit. Men moet ook uitmaken of de inhoud van de website eventueel auteursrechtelijk beschermd is. Is het antwoord op beide vragen negatief, dan kan de archivaris zonder problemen de bewerkingen die nodig zijn voor archivering doorlopen. Is het antwoord op minstens één van beide vragen positief, dan moet de archivaris overgaan naar stap 2.

**STAP 2:** Vervolgens moet de archivaris bepalen welke bewerkingen er allemaal moeten uitgevoerd worden op de website (reproductie, wijzigingen...) en deze kwalificeren als vallende onder hetzij de vermogensrechten, hetzij de morele rechten.

**STAP 3:** Tenslotte moet de archivaris bepalen bij wie de vermogensrechten en/of de morele rechten berusten. De morele rechten berusten in principe bij diegene die het werk daadwerkelijk tot stand heeft gebracht. De vermogensrechten kunnen bij overeenkomst afgestaan zijn. Soms is er bij wet bepaald dat de vermogensrechten principieel bij de werkgever berusten (computerprogramma's en databanken). Aan deze perso(o)n(en) moet de toestemming gevraagd worden om de nodige bewerkingen uit te voeren.

---

<sup>163</sup> Art. 7 2° databankenwet.

## D. OPLOSSING: UITZONDERING VOOR PRESERVERINGSDOELEINDEN

De huidige wetgeving inzake auteursrecht biedt onvoldoende soelaas voor preserveringsdoeleinden, zowel van papieren werken als van digitale werken. In principe is voor al de handelingen die een archivaris regelmatig moet stellen, de toestemming van de auteurs nodig, zowel van de ontwerper van de website als van de auteur van de inhoud.

Soms is echter niet duidelijk bij wie de auteursrechten berusten of zijn er verschillende auteurs. Vooral voor websites op het vaak anonieme Web, is vaak niet te achterhalen bij wie de auteursrechten liggen. Voor archiefinstellingen is het onbegonnen werk om de toestemming te vragen van iedere auteur. Dit is niet alleen erg omslachtig, maar ook bijzonder tijdrovend. Bovendien loopt men het risico om deze toestemming niet te krijgen. Voor websites van overheidsinstellingen hebben we de problematiek hierboven op bepaalde punten enigszins gerelativeerd, maar dit neemt niet weg dat zij vroeg of laat ook te maken krijgen met de obstakels van het auteursrecht.

Bestaande wetgeving moet uiteraard ook zoveel mogelijk gelden voor de digitale wereld. Alleen als er sprake is van belemmering voor deze digitale wereld, moet die wetgeving worden aangepast.<sup>164</sup> Na bovenstaande analyse denken we te kunnen zeggen dat het auteursrecht een belemmering vormt voor archiefinstellingen, niet alleen met betrekking tot de digitale archivering van websites, maar met betrekking tot de archivering van alle auteursrechtelijk beschermde werken. Een wetswijziging is dus niet alleen nodig door de toenemende digitalisering, maar het zal tevens een oud zeer van de archiefsector uit de wereld helpen.

Art. 46 7° AW bepaalt dat de exclusieve rechten van uitvoerende kunstenaars worden beperkt in het geval van de bewaring door het Koninklijk Belgisch Filmarchief van het cinematografisch patrimonium door middel van kopieën, dubbels, restauraties en overzettingen. De AW voorziet dus wel een uitzondering voor de preservatie van auteursrechtelijk beschermd materiaal, maar helaas is deze uitzondering beperkt tot één Belgische archiefinstelling en heeft ze enkel betrekking op de rechten van de uitvoerende kunstenaar. Er is wetgevend optreden nodig om deze uitzondering te veralgemenen.

In het voorjaar van 1994 werd er van start gegaan met de harmonisatie van het auteursrecht. De werkgroep die in mei 1994 het zogenaamde *Bangemann-rapport* afleverde, was van mening dat *‘in order to stimulate the development of new multimedia products and services, existing legal regimes - both national and Union - will have to be re-examined to see whether they are appropriate to the new information society’* en dat *‘where necessary, adjustments will have to be made’*.<sup>165</sup> Pas op 9 april 2001 nam de Raad van Ministers de nieuwe auteursrechtlijn aan.<sup>166</sup> Ter gelegenheid van de aanpassing van het auteursrecht aan de opkomst van allerlei digitale reproductie- en verspreidingstechnieken, werd er voor de lidstaten de mogelijkheid gecreëerd om in uitzonderingen te voorzien op het auteursrecht. Van harmonisatie zal er op dit punt echter niet veel terecht komen vermits de meeste uitzonderingen facultatief zijn. Bovendien is de limitatieve lijst van op het auteursrecht toegelaten beperkingen, opgenomen in artikel 5 van de richtlijn, een samenraapsel van de reeds bestaande

<sup>164</sup> LANGE, J., ‘De wet en elektronische publikaties’,

<http://nieuws.surfnet.nl/nieuws/snn-archief/achtergrond/jg97-98/kb-symposium2.html>

<sup>165</sup> Europe and the global information society, Recommendations to the European Council <http://www.bookmarks.de/lib/politics/bangemann/report.html#section14>

<sup>166</sup> Richtlijn 2001/19/EG van het Europees Parlement en de Raad betreffende de harmonisatie van bepaalde aspecten van het auteursrecht en de naburige rechten in de informatiemaatschappij, Pb. EG 2001 L 167/10).

uitzonderingen in de lidstaten. Het auteursrecht zal in de meeste lidstaten dus vermoedelijk blijven zoals het is, inclusief de uitzonderingen.

- Deze aanpak is voor de archiefsector een gemiste kans met zware gevolgen. Artikel 5 lid 2 c) van de richtlijn voorziet dat de lidstaten in hun auteursrecht restricties kunnen invoeren op het exclusieve reproductierecht van de auteur ten behoeve van archieven die niet direct of indirect het behalen van een economisch of commercieel voordeel nastreven. Het Europees Parlement, evenals Nederland en een aantal andere lidstaten hadden aangedrongen op deze bepaling, om zo digitaal materiaal in een archief te kunnen opslaan. Enkel niet-commerciële instellingen, zoals overheidsarchieven toegankelijk voor het publiek, worden geëxempteerd met deze uitzondering. Voor alle duidelijkheid herhalen we hier dat de vrijwillig in te voeren uitzondering niet de reproductie van reeds gearchiveerde werken betreft, zoals bijvoorbeeld ten behoeve van onderzoekers (hiervoor bestaat er reeds een uitzondering: zie hierboven), maar wel reproducties die nodig zijn om de werken te kunnen archiveren bijv. bij het downloaden van websites, of om ze te kunnen blijven bewaren in het geval van technologische veroudering.

De lidstaten mogen ook een uitzondering aanbrengen in hun nationale wetgeving op het exclusieve recht van mededeling aan het publiek door bibliotheken, musea en archieven van hun collecties. Deze uitzondering moet wel aan enkele strikte voorwaarden voldoen <sup>167</sup>:

- De terbeschikkingstelling mag enkel gebeuren in de gebouwen van de instellingen via speciale terminals. Voor de ter beschikking stelling via het Internet van gearchiveerd digitaal materiaal zal dus nog steeds de toestemming nodig zijn van de rechthebbende.
- Deze uitzondering mag ook enkel gelden voor materiaal dat niet elders te koop wordt aangeboden (bijv. software of documenten die op een website te koop worden aangeboden) of aan licentievoorwaarden is onderworpen. Dit is logisch: het uitgangspunt van het auteursrecht is immers dat aan de auteurs een loon dient te worden gewaarborgd als vergoeding voor hun artistieke creatie, hun bijdrage aan de culturele productie en hun intellectuele inspanning. Dit is mogelijk gemaakt door het reproductierecht waardoor de auteur zijn werk exclusief voor zichzelf kan commercialiseren. Door het regime van de dwanglicenties verkrijgt de auteur een uitgesteld loon in de gevallen waarbij een exclusief reproductierecht niet te verantwoorden is. De lidstaten mogen niet toelaten dat materiaal dat reeds gecommercialiseerd werd, door een beroep te doen op archiefinstellingen toch gratis door derden kan geconsulteerd worden.
- De consultatie mag enkel toegelaten worden voor onderzoek of voor privé-studie. Net zoals de bibliothecaris bij het reproductierecht voor didactische doeleinden <sup>168</sup>, zal de archivaris het niet gemakkelijk hebben om de bedoeling na te gaan in hoofde van de persoon die zich in de

<sup>167</sup> Art. 5.3 n) van de richtlijn: *‘De lidstaten kunnen beperkingen stellen op het mededelingsrecht ten aanzien van het gebruik van niet te koop aangeboden of aan licentievoorwaarden onderworpen werken dat onderdeel uitmaakt van de verzamelingen van de in lid 2 onder c) bedoelde instellingen (niet-commercieel ingestelde musea, archieven en bibliotheken), hierin bestaande dat het werk of materiaal via speciale terminals in de gebouwen van die instellingen voor onderzoek of privéstudie meegedeeld wordt aan of beschikbaar gesteld wordt voor individuele leden van het publiek.’*

<sup>168</sup> Zie voetnoot 143.

leeszaal, maar is het toch zijn eigen verantwoordelijkheid om dit vooraf te controleren. Eventueel zal hij de voorleggen van de nodige bewijsstukken kunnen vragen.

Voor België kunnen we gerust zeggen dat de wetgever hier (tot op heden) de bal heeft misgeslagen. Momenteel ligt er in de Senaat een wetsvoorstel voor tot wijziging van de wet van 30 juni 1994 betreffende het auteursrecht en de naburige rechten in de context van de ontwikkeling van de informatiemaatschappij.<sup>169</sup> Op 15 februari 2002 werd er tevens een voorontwerp van wet goedgekeurd door de Ministerraad. In beide wetgevende initiatieven worden de vrijwillig in te voeren uitzonderingen op het reproductierecht, inclusief de uitzondering voor preserving, van tafel geveegd. De uitzonderingen die reeds bestonden, worden wel behouden.

In de voorbereidende werken kunnen we lezen dat door de aanvaarding van de overige voorgestelde uitzonderingen de regeling van het auteursrecht volkomen onderuit zou halen.<sup>170</sup> De idee is dat het aannemen van al deze uitzondering de essentie van het auteursrecht zou teniet doen, en men zou evolueren naar een systeem van overheidssubsidiëring van auteurs. Nu is het zo dat reproductie, mededeling enz. slechts in een beperkt aantal gevallen zonder toestemming van de auteur toegelaten is. Dan krijgt men een systeem waarbij reproductie, mededeling enz. principieel zijn toegelaten omwille van de vele uitzonderingen, en de auteur ‘in ruil’ een billijke vergoeding ontvangt van de overheid.

De casus van archivering heeft echter niets te maken met het al dan niet geven van een vergoeding aan de auteur. Het gaat hier niet om een afweging van de belangen van de éne particulier of organisatie ten opzichte van de auteur. Het gaat bij archivering door publiekrechtelijke archiefinstellingen om een afweging van de belangen van auteurs en de belangen van het publiek. Men voelt zo aan dat de inbreukmakende handelingen die nodig zijn voor de archivering van digitale informatie ‘moeten kunnen’. Wanneer het wetsvoorstel blijft zoals het nu is voor wat betreft de omzetting van de facultatieve uitzonderingen uit de richtlijn, zullen archivariissen echter geen aansluiting vinden bij een bestaande wettelijke beperking.

Men kan het betreuren dat de uitzondering voor preservingdoeleinden niet als een verplichte uitzondering is opgenomen in de richtlijn. We hopen dat de Belgische wetgever de uitzondering alsnog zal opnemen in de auteurswet. De uitzondering voor preservingdoeleinden zou dan best beperkt blijven tot reproducties die noodzakelijk zijn om de bewaartaken te volbrengen. Deze uitzondering zou dan als volgt kunnen luiden:

*“Als een inbreuk op het auteursrecht op werken van letterkunde of kunst wordt niet beschouwd, de reproductie door instellingen belast met bewaartaken die niet een direct of indirect economisch of commercieel voordeel nastreven, en enkel in de mate dat de reproductie noodzakelijk is om die bewaartaken te kunnen uitoefenen.*

---

<sup>169</sup> Wetsvoorstel van 23 maart 2001 tot wijziging van de wet van 30 juni 1994 betreffende het auteursrecht en de naburige rechten in de context van de ontwikkeling van de informatiemaatschappij, *Parl. St.*, Senaat, 2000-2001, 2-704/1.

<sup>170</sup> Wetsvoorstel van 23 maart 2001 tot wijziging van de wet van 30 juni 1994 betreffende het auteursrecht en de naburige rechten in de context van de ontwikkeling van de informatiemaatschappij, *Parl. St.*, Senaat, 2000-2001, 2-704/1, 4.

*Als een inbreuk op het auteursrecht op werken van letterkunde of kunst wordt niet beschouwd, de mededeling aan het publiek door de instellingen genoemd in het vorige lid, van materiaal uit hun collectie voor zover:*

- *dit niet elders te koop wordt aangeboden of aan licentievoorwaarden is onderworpen, en*
- *de mededeling gebeurt via terminals in de gebouwen van de genoemde instellingen, en*
- *de consultatie uitsluitend gebeurt voor onderzoek of privé-studie”*

Totdat een dergelijke bepaling ooit in de AW zal zijn opgenomen, zijn archivariissen verplicht om steeds weer de toestemming te vragen van de rechthebbenden. In de niet-ondenkbare situatie waarbij men deze toestemming ooit eens niet zou kunnen bekomen, is het maar te hopen dat de rechters met betrekking tot archivering de theorie van het Nederlandse Dior/Evora-arrest volgen bij gebrek aan een wettelijke uitzondering. De feiten die ten grondslag lagen aan dit arrest waren dat Dior op basis van het auteursrecht op de verpakkingen van haar parfums trachtte te voorkomen dat het Kruidvat de verpakkingen in haar reclamefolder zou afdrukken. Zo probeerde Dior de parallelimport van haar parfums af te wenden, nu dit op basis van haar uitgeputte merkenrecht niet langer mogelijk was. De Nederlandse auteurswet kent echter geen uitzondering met betrekking tot deze situatie. Door het gesloten stelsel van de uitzonderingen op het auteursrecht was Dior er dus bijna in geslaagd om de verkoop door het Kruidvat te beletten. De Hoge Raad heeft hier echter ruimte gecreëerd om de grenzen van het auteursrecht verder te beperken aan de hand van een afweging van de belangen van de auteur en van de maatschappelijke of economische belangen van anderen.<sup>171</sup>

Misschien zal de rol van het auteursrecht inderdaad gaan verminderen in de toekomst, zoals heel wat auteurs voorspellen. Auteursrecht is altijd het belangrijkste juridische middel geweest, en is dit nog steeds, om de massa-verspreiding van informatie te beperken.<sup>172</sup> Het karakteristieke van het Internet is nu precies dat informatie zeer ruim kan verspreid worden en in principe toegankelijk is voor iedereen die een Internetaansluiting heeft.

## VIII. ARCHIVEREN VAN PERSOONSGEGEVENS

### A. PERSOONSGEGEVENS EN INTERNET

Naast het probleem van het auteursrecht moeten we ons bij webarchivering de vraag stellen naar aspecten van privacybescherming. Dat de privacyargumenten ook op Internet spelen, blijkt duidelijk uit het stopzetten van het webarchiveringsproject van de Zweedse nationale bibliotheek door de Zweedse *Data Inspection Board*. In november 2001 was de Zweedse nationale bibliotheek gestart met het automatisch verzamelen en bewaren van Zweedse websites met als bedoeling om ook dit deel van het culturele erfgoed voor de toekomst te bewaren. De *Data Inspection Board* was van mening dat het

<sup>171</sup> HR 20 oktober 1995, NJ 1996, 682; ALBERDINGK THIJM, C., ‘Fair use: het auteursrechtelijk evenwicht hersteld’, [http://www.ivir.nl/publicaties/overig/alberdingk\\_thijm/fair-use.doc](http://www.ivir.nl/publicaties/overig/alberdingk_thijm/fair-use.doc)

<sup>172</sup> HUGENHOLTZ, B., *Intellectual Property rights on the Information Superhighway*, Report to the Commission of the European Communities (DG XV), August 1994.

houden van zo'n collectie neerkomt op een verwerking van persoonsgegevens en daarom moet stroken met de *Personal Data Act*. Hoewel de *Board* erkende dat er nood is aan het bewaren van het culturele erfgoed voor historische doeleinden en onderzoeksdoeleinden, werd er beslist dat het verzamelen van en het toegankelijk maken van persoonsgegevens voor dergelijke doeleinden in een specifieke wet moet geregeld worden. In afwachting van een dergelijke wetgeving moest de Zweedse nationale bibliotheek haar webarchiveringsactiviteiten stopzetten.

Iedere verwerking van persoonsgegevens, ook wanneer die door archivariissen wordt uitgevoerd, is in België onderworpen aan het regime van de wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens (verder: de WVP, wet verwerking persoonsgegevens). “Verwerking” betekent “iedere bewerking met betrekking tot persoonsgegevens, zoals het verzamelen, opvragen (downloaden), bewaren, verstrekken door middel van doorzending enz. Een “persoonsgegeven” is iedere informatie betreffende een geïdentificeerde of identificeerbare natuurlijke persoon. Op websites komen vaak heel wat persoonsgegevens voor: namen, adressen, telefoonnummers, enz. zijn allemaal persoonsgegevens in de zin van de wet.

De HTML-pagina's van dynamische websites worden pas samengesteld op de server na het ontvangen van een HTTP-verzoek. Wat de bezoeker uiteindelijk te zien krijgt, is afhankelijk van o.a. het verzoek dat hij formuleerde of van zijn profiel. De reconstructie van de inhoud van deze websites in het archief is enkel mogelijk als ook deze interactie wordt bewaard. Om deze interactie vast te leggen, moeten soms logbestanden, IP-adressen en/of gepersonaliseerde informatiestromen gebaseerd op cookies mee worden gearhiveerd zoals beschreven werd in het technische deel van dit rapport. Om uit te maken of de verwerking van deze gegevens door de archiefdienst onder de reikwijdte van de WVP valt, moeten we eerst iets zeggen over de organisatie en het beheer van het Internet.

#### A.1. Organisatie en beheer van het Internet

Het Internet is een wereldomspannend computernetwerk dat gebruikt wordt voor informatieuitwisseling en communicatie. Op het Internet wordt elke computer geïdentificeerd door een uniek, numeriek IP-adres met als vorm A.B.C.D., waarbij A, B, C en D getallen zijn tussen 0 en 255 (bijv. 67.154.85.243). Hoewel het Internet door velen als een internationale gemeenschap van personen wordt beschouwd, is een IP-adres dus niet in eerste instantie gekoppeld aan een persoon, maar aan een computer. Een domeinnaam is een andere, gemakkelijkere manier om een locatie op het Internet aan te duiden. Het is een makkelijk te onthouden lettercombinatie bijv. een bedrijfsnaam, die correspondeert met een bepaald IP-adres. De domeinnaam wordt door de mens begrepen, het IP-adres door de computer.

Omdat de informatie die via het Internet wordt aangeboden verspreid is over miljoenen computers, wordt er gebruik gemaakt van *Uniform Resource Locators* (URL's). Een URL (ook wel “internetadres” genoemd) is niet hetzelfde als een domeinnaam. Een URL is de standaardmanier om de plaats van de verschillende informatiebronnen aan te geven. Het eerste gedeelte van de URL geeft aan welk protocol er moet gebruikt worden om de informatie op te roepen (bijv. http:// voor bestanden op een webserver, ftp:// voor bestanden op een FTP server, news:// voor bestanden van een Usenet nieuwsgroep enz.). Daarnaast bevat een URL het IP-adres van de computer waarop de gevraagde informatie is opgeslagen. Een algemene URL bestaat alleen uit de protocolaanduiding en het IP-adres van de computer, en levert meestal de homepage op van deze computer. Op sommige computers staan er echter meerdere homepagina's. Ze krijgen dan elk een eigen *directory* of *folder* op de

computer die dan ook mee in de URL wordt opgenomen.<sup>173</sup> Tenslotte vermeldt de URL ook vaak de naam en het formaat van het opgevraagde document (bijv. .html, archiveren-van-websites.pdf)

Al heel snel bleek dat het onmogelijk was om miljoenen domeinnamen wereldwijd te coördineren en uniek te houden. Daarom werd het *Domain Name System* (DNS) bedacht. Dit is een systeem voor het toekennen van domeinnamen aan door een IP-adres geïdentificeerde computers. Het vertaalt een IP-adres in een domeinnaam (of omgekeerd) en omvat een wereldwijde, gedecentraliseerde databank van domeinnamen die hiërarchisch opgebouwd is. Domeinnamen hebben het formaat <namen>.topniveaudomein, waarin het topniveaudomein een generiek domein kan zijn, zoals .com voor commerciële websites, of een geografisch domein, zoals .be. Het topniveaudomein kan je vergelijken met het zonenummer van een telefoonnummer. De domeinen die de server moet doorzoeken om de gevraagde informatie te vinden, worden op die manier beperkt. Wanneer een surfer via zijn browser de opdracht geeft om een bepaalde website op te vragen, zal de server van de ISP het door de surfer opgegeven internetadres (de URL) vertalen in het IP-adres van de geadresseerde. Servers die zich bevinden op internetknooppunten, nl. de zgn. routers, gaan dan op zoek naar de beheerder van de betreffende site.

Iedere persoon of organisatie die toegang wil krijgen tot het Internet, dient zijn systemen dus te voorzien van een IP-adres. De *Internet Corporation for Assigned Names and Numbers* <sup>174</sup> is verantwoordelijk voor de toewijzing van IP-adressen aan internetaanbieders, die ze op hun beurt aan hun klanten verstrekken. Om een verbinding te krijgen, moet deze persoon of organisatie een contract sluiten met een internetaanbieder <sup>175</sup>, waarbij hij zijn/haar naam, adres en andere persoonsgegevens moet opgeven. De abonnee ontvangt dan een gebruikersnaam (*user ID*) en een paswoord om te vermijden dat anderen op Internet zouden inloggen aan de hand van dit abonnement. De internetaanbieder is dus in principe steeds in staat om aan de hand van het IP-adres de individuele netwerkgebruiker te identificeren.

Dit is ook het geval voor zogenaamde “dynamische” IP-adressen. Dynamische IP-adressen zijn het gevolg van het feit dat het aantal IP-adressen beperkt is. Internetaanbieders beschikken dus ook over een beperkt aantal IP-adressen die ze kunnen verdelen over hun abonnees. Daarom krijgen de abonnees die geen permanente verbinding hebben met het Internet, steeds een ander IP-adres toegewezen wanneer ze inloggen. Als een abonnee zijn internet sessie beëindigt, komt het IP-adres vrij voor een andere gebruiker. Een internetaanbieder kan de individuele gebruiker identificeren aan de hand van een dynamisch IP-adres, op voorwaarde dat de internetaanbieder de verkeersgegevens bewaard die aangeven op welk moment een bepaald IP-adres aan welke loginnaam werd toegekend. Internetaanbieders houden uit veiligheidsoverwegingen systematisch registraties bij van de data, de tijdstippen en de duur van de sessies van hun abonnees en van de aan hen toegewezen (dynamische) IP-adressen.

Aan de webserverkant worden er door de ISPs ook logbestanden bijgehouden van het internetverkeer. Eenmaal de verbinding met de website tot stand is gekomen, begint deze informatie te

---

<sup>173</sup> De website van het ICRI wordt beheerd op dezelfde computer als de website van de faculteit Rechtsgeleerdheid van de K.U.Leuven: <http://www.law.kuleuven.ac.be/icri>

<sup>174</sup> <http://www.icann.org>

<sup>175</sup> Een internetaanbieder is niet hetzelfde als een Internet Service Provider (ISP). ISPs doen aan webhosting nl. het plaatsen van webpagina's op hun webserver. Een internetaanbieder maakt dus voor zijn abonnees de toegang mogelijk tot de informatie die op de computers van de verschillende ISPs staat. Een internetaanbieder heeft als toegangspoort naar het Internet kennis van alle verkeer van de abonnee, terwijl een ISP alleen op de hoogte is van wat er op de eigen servers gebeurt.

verzamelen over de bezoeker. Al deze informatie wordt geregistreerd in de serverlogbestanden waardoor er grote hoeveelheden gegevens over het communicatieproces verzameld worden: wie heeft de website bezocht, op welke datum en tijdstip, duur van het bezoek, de uitgevoerde acties, *clickstreams*, de bezochte pagina's, welke bestanden gedownload werden enz. De laatste tijd is er nieuwe bewakingssoftware beschikbaar voor ISPs die in real time rapporteert aan de ISP over de soorten inhoud die gebruikers bekijken en binnenhalen.<sup>176</sup>

Een andere mogelijkheid om “gepersonaliseerde” informatie over bezoekers te bekomen en te registreren, zijn *cookies*. Cookies worden vaak door direct-marketing bedrijven gebruikt om je interesses en je consumptiegedrag te leren kennen. De bezochte website zal dan een cookie naar de webbrowser sturen, die het vervolgens opslaat op de harde schijf van de computer van de surfer. Wanneer deze persoon vervolgens de website opnieuw bezoekt, zal de webbrowser systematisch zijn voorkeuren en zijn gebruikersprofiel meedelen. De communicatiestromen die op basis daarvan tot stand komen, moeten dan ook mee gearchiveerd worden.

## A.2. IP-adressen

Met betrekking tot IP-adressen moeten we ons afvragen of de verwerking ervan door een archiefdienst onder de toepassing van de privacyreglementering valt. Om deze vraag te kunnen beantwoorden, moeten we uitmaken of een IP-adres als een persoonsgegevens moeten beschouwd worden. Deze kwestie is helemaal niet eenduidig te beantwoorden vanuit het standpunt van het Belgisch recht. Toch is ze van belang omdat het antwoord bepaalt welke rechten de betrokkenen kunnen doen gelden ten aanzien van de archiefdienst (recht op mededeling en correctie). Bovendien kan de niet-naleving van de wet aanleiding geven tot zware strafsancities. Deze onduidelijkheid heeft te maken met de interpretatie van de term “identificeerbaar”. In de Europese richtlijn 95/46 van 24 oktober 1995 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrij verkeer van die gegevens wordt een “identificeerbare” persoon gedefinieerd als “een persoon die direct of indirect kan worden geïdentificeerd, met name aan de hand van een identificatienummer of van één of meer specifieke elementen die kenmerkend zijn voor zijn of haar fysieke, fysiologische, psychische, economische, culturele en sociale identiteit.” Uitgaande van overweging 26 bij de richtlijn<sup>177</sup> verduidelijkt de Memorie van Toelichting bij de Belgische WVP dat informatie betreffende een persoon als een persoonsgegeven moet beschouwd worden, zolang iemand nog in staat is om, met welk redelijkerwijs inzetbaar middel ook, te achterhalen op welk individu de informatie betrekking heeft.”<sup>178</sup> Om een gegeven als een persoonsgegeven te bestempelen, is het bijgevolg reeds voldoende dat een andere persoon dan de verantwoordelijke voor de verwerking in staat is om met een redelijkerwijs inzetbaar middel het

---

<sup>176</sup> *Privacy op Internet – Een geïntegreerde EU-aanpak van on-linegegevensbescherming*, Groep Gegevensverwerking artikel 29, goedgekeurd op 21 november 2000, Europese Unie, 50.

<sup>177</sup> “Om te bepalen of een persoon identificeerbaar is, moet worden gekeken naar alle middelen waarvan mag worden aangenomen dat zij redelijkerwijs door degene die voor de verwerking verantwoordelijk is dan wel door enig ander persoon in te zetten zijn om genoemde persoon te identificeren.”

<sup>178</sup> Memorie van Toelichting bij het wetsontwerp tot omzetting van de Richtlijn 95/46/EG van 24 oktober 1995 van het Europees Parlement en de Raad betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrij verkeer van die gegevens, *Parl. St.*, Kamer, 1997-1998, 1566/1, 12.

individue te identificeren. Of met andere woorden: bij anonimisering verliest informatie over personen pas het karakter van persoonsgegevens indien de anonimisering absoluut is en er geen terugweg meer mogelijk is uit die anonimiteit.<sup>179</sup>

Toegepast op IP-adressen betekent dit dat een IP-adres een persoonsgegeven is zolang iemand nog in staat is om met welk redelijkerwijs inzetbaar middel ook, te achterhalen op welk individu het IP-adres betrekking heeft. Een archiefdienst die de IP-adressen van de bezoekers mee wenst te archiveren, verwerkt bijgevolg meestal persoonsgegevens. Immers, de internetaanbieder is in staat om aan de hand van de IP-adressen de individuele surfer te identificeren, tenzij wanneer, in het geval van dynamische IP-adressen, de verkeersgegevens betreffende het toekennen van IP-adressen niet bewaard werden of na verloop van tijd niet meer voorhanden zijn.

In de andere lidstaten van de Europese Unie is deze strikte interpretatie niet gevolgd. In Nederland bijvoorbeeld heerst er een gans andere opvatting. Dit komt o.a. tot uiting in een advies van de Nederlandse Registratiekamer (de tegenhanger van onze Commissie voor de bescherming van de persoonlijke levenssfeer) van 19 maart 2000<sup>180</sup>, waarin zij stelt dat een IP-adres niet altijd als een persoonsgegeven moet beschouwd worden. Ten aanzien van het begrip “identificeerbaar” speelt volgens de Registratiekamer vooral de vraag of de identiteit van de persoon, redelijkerwijs, zonder onevenredige inspanning, vastgesteld kan worden door de verantwoordelijke voor de verwerking (en dus niet door enig ander persoon)<sup>181</sup>. Dit hangt mede af van de mogelijkheden waarover de houder beschikt en de bekendheid of de beschikbaarheid van aanvullende informatie. Hierbij moet uitgegaan worden van een redelijk toegeruste houder. In concrete gevallen moet ook rekening worden gehouden met de bijzondere expertise en de technische faciliteiten van de verantwoordelijke voor de verwerking, aldus de Registratiekamer.

Dit standpunt, dat allicht veel pragmatischer is en beter bijdraagt tot de bescherming van de persoonlijke levenssfeer<sup>182</sup>, heeft voor gevolg dat IP-adressen slechts als persoonsgegevens kunnen worden beschouwd indien de verantwoordelijke voor de verwerking over de mogelijkheid beschikt om aan de hand van IP-adressen individuele internetgebruikers te identificeren. Het is niet zo dat bij dit standpunt de informatie waarover derden beschikken niet mee in rekening mag gebracht worden. Wanneer een derde (bijv. een internetaanbieder) over gegevens beschikt die identificatie mogelijk maken, dan moet de verantwoordelijke voor de verwerking (bijv. de archivaris) aantonen dat hij niet beschikt of niet kan beschikken over de bijkomende informatie die nodig is om de betrokkene te identificeren.

Men kan zich afvragen of deze interpretatie van het begrip “identificeerbaar” voortaan niet in het Belgisch recht kan worden gehanteerd. De Privacycommissie heeft immers in haar advies over het voorontwerp van de WVP deze interpretatie gesteund. Ze stelde dat de vraag of de terugweg uit de anonimiteit redelijkerwijs nog mogelijk is, moet beoordeeld worden in hoofde van de verantwoordelijke voor de verwerking.<sup>183</sup> Het wettelijk kader laat echter niet toe om dit standpunt te verdedigen. Het uitvoeringsbesluit bevestigt immers de zeer strikte uitlegging die destijds in de

<sup>179</sup> *Ibid.*

<sup>180</sup> Advies z2000-0340, 19 maart 2000, <http://www.registratiekamer.nl/bis/content-1-1-9-3-7-2.html>

<sup>181</sup> In de Nederlandse Privacywet wordt de term “houder” gebruikt in plaats van “verantwoordelijke voor de verwerking).

<sup>182</sup> De verantwoordelijke voor de verwerking die zelf niet over de middelen beschikt om de persoon te identificeren, zal immers aarzelen om te trachten toch tot identificatie over te gaan, omdat hij dan onder het toepassingsgebied van de wet zou komen.

<sup>183</sup> Advies 30/96.

Memorie van Toelichting werd gegeven. In de bepalingen betreffende het historisch, statistisch en wetenschappelijk onderzoek is er sprake van *gecodeerde persoonsgegevens*. Hieruit moet men besluiten dat de codering van gegevens, en dus de anonimiteit in hoofdte van de verwerker, niets afdoet aan de kwalificatie als persoonsgegevens.

Toch bestaat er een verdedigbaar alternatief voor het standpunt van de wetgever dat niet in strijd is met de tekst van de wet, en waar de archiefdienst een beroep op kan doen om aan de toepassing van de WVP te ontsnappen. Uit de formulering van de definitie van het begrip “identificeerbaar” in art. 1 §1<sup>184</sup> van de wet kan men een weerlegbaar vermoeden van identificeerbaarheid afleiden. Het uitgangspunt blijft dat een persoon als identificeerbaar wordt beschouwd zodra er technisch gezien ergens een middel bestaat om die persoon te identificeren. De verantwoordelijke kan echter het tegenbewijs leveren dat hij niet beschikt of niet kan beschikken over dat middel om de identificatie door te voeren.<sup>185</sup> Hoewel het gaat om een negatief feit, zal de archiefdienst dit bewijs in veel gevallen vrij gemakkelijk kunnen leveren bij de archivering van IP-nummers, aangezien er geen enkele relatie bestaat tussen de archiefdienst en de internetaanbieders van de abonnees die de te archiveren website hebben bezocht.

### A.3. Logbestanden

De gegevens in de logbestanden van ISPs, afkomstig van het registreren van het surfgedrag, zijn meestal niet onmiddellijk te herleiden tot een natuurlijke persoon, maar wel tot een IP-adres. Indien het mogelijk is om aan de hand van dat IP-adres de gebruiker te identificeren, dan zijn de logbestanden aan te merken als persoonsgegevens. Zolang het echter is uitgesloten dat de logbestanden kunnen gebruikt worden om het bezoek aan bepaalde websites door individuele gebruikers te analyseren, is er geen sprake van persoonsgegevens.

Ten tweede zijn de logbestanden ook “verkeersgegevens” in de zin van het gemeenschappelijk standpunt over het ontwerp van richtlijn die de huidige richtlijn 97/66/EG betreffende de verwerking van persoonsgegevens en de bescherming van de persoonlijke levenssfeer in de telecomunicatiesector moet vervangen. Verkeersgegevens in de zin van de richtlijn zijn “gegevens die worden verwerkt voor het overbrengen van communicatie over een elektronisch communicatienetwerk of voor de facturering ervan.” Het gemeenschappelijk standpunt bepaalt dat verkeersgegevens principieel vertrouwelijk zijn en dat het opslaan of anderszins registreren door anderen dan de netwerkgebruikers verboden is. Enkel de technische opslag die nodig is voor het overbrengen van informatie is toegelaten.<sup>186</sup> Maar zelfs in dat geval moeten de verkeersgegevens principieel gewist worden of anoniem worden gemaakt wanneer ze niet langer nodig zijn voor het instandhouden van de communicatie, met andere woorden zodra de internetgebruiker toegang heeft tot de website.<sup>187</sup>

Toch stellen we vast dat ISP’s systematisch en veel langer dan nodig voor het instandhouden van de communicatie, logbestanden bijhouden. De nieuwe wet inzake informaticacriminaliteit verplicht

<sup>184</sup> “Als identificeerbaar wordt beschouwd...”

<sup>185</sup> Deze mening werd ontleend aan: DE BOT, D., *Verwerking van persoonsgegevens*, Antwerpen, Kluwer, 2001, 32 e.v.

<sup>186</sup> Art. 5 van het gemeenschappelijk standpunt.

<sup>187</sup> Art. 6 van het gemeenschappelijk standpunt.

ISPs zelfs om hun logbestanden gedurende minstens 12 maanden bij te houden met het oog op de opsporing en de vervolging van strafbare feiten gepleegd via een communicatienetwerk.<sup>188</sup> Deze uitzondering op het principe van de vertrouwelijkheid van telecommunicatie kan echter enkel ter waarborging van de nationale veiligheid ingeroepen worden. De bewaring van logbestanden door een archiefdienst kan deze toets niet doorstaan. Tenzij men dus als instelling zelf de eigen website host en dus optreedt als ISP, zal men als archiefdienst zeer moeilijk aan de logbestanden kunnen geraken van de te archiveren website.

## **B. VERWERKING VAN PERSOONSGEGEVENS VOOR HISTORISCHE DOELEINDEN: IN OVEREENSTEMMING MET HET OORSPRONKELIJKE DOEL?**

De vorige paragraaf toont aan dat een archiefdienst in een aantal gevallen op persoonsgegevens zal stoten bij webarchivering. De vraag is dan hoe deze persoonsgegevens moeten verwerkt worden. Het antwoord is gelegen in de betekenis van het begrip “historische doeleinden” in de WVP.

Het uitgangspunt van de Belgische privacyreglementering is dat de verwerking van persoonsgegevens principieel niet toegelaten is, tenzij in welbepaalde, duidelijk omschreven gevallen. De gevallen waarin een verwerking toelaatbaar is, worden omschreven in art. 5 van de WVP. In de gevallen waarin één van de toelaatbaarheidscriteria is vervuld (bijv. toestemming van de betrokkene), mag de verwerker bovendien enkel persoonsgegevens verzamelen voor welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doeleinden.

Bewaring van persoonsgegevens is wat men noemt een nieuwe, *verdere* verwerking van persoonsgegevens. Wat wordt er daarmee nu bedoeld? Het verzamelen en registreren van persoonsgegevens is op zich al een vorm van verwerking. Het is de bedoeling dat er daarna iets gebeurt met die gegevens. Een verdere verwerking van persoonsgegevens is iedere bewerking die men na de verkrijgen met betrekking tot de gegevens verricht zoals de bewaring ervan, maar ook de raadpleging, de verspreiding enz.<sup>189</sup> Een verdere verwerking moet volgens de wet verenigbaar zijn met de doeleinden waarvoor de gegevens werden verkregen, rekening houdend met alle relevante factoren met name de redelijke verwachtingen van de betrokkene.<sup>190</sup> Wanneer men bijvoorbeeld persoonsgegevens verkregen heeft voor direct marketing doeleinden, dan mag men de gegevens voor deze doeleinden ook bewaren.

De richtlijn, en in navolging daarvan de Belgische WVP, bepalen dat een verdere verwerking van persoonsgegevens *voor historische doeleinden* niet onverenigbaar is met het oorspronkelijke primaire doel (en dus toegelaten is), voorzover er passende garanties worden geboden.<sup>191</sup> De vraag is dus of de verwerking van persoonsgegevens door een archivaris een verwerking is voor historische doeleinden. Het antwoord op deze vraag is van groot belang, aangezien België als passende garanties voor de verwerking van persoonsgegevens voor historische doeleinden (en voor statistische en wetenschappelijke doeleinden) een gradueel driestapssysteem heeft ingevoerd dat is terug te vinden in hoofdstuk II van het uitvoeringsbesluit van de WVP. Is het antwoord op deze vraag positief, dan moet de archivaris het driestapssysteem toepassen.

<sup>188</sup> Eerst moet er evenwel een K.B. uitgevaardigd worden dat bepaalt welke gegevens er precies moeten bewaard worden.

<sup>189</sup> DE BOT, D., *o.c.*, nr. 155, 119.

<sup>190</sup> Art. 4 §1 2° WVP.

<sup>191</sup> Art. 6.1 b) van de richtlijn en art. 4 §1 2° WVP.

In beginsel moet de historische verwerking aan de hand van anonieme gegevens gebeuren. Indien de historische doeleinden van de verwerking niet verwezenlijkt kunnen worden door het verwerken van anonieme gegevens, dan mag de verwerker gecodeerde persoonsgegevens verwerken. Indien ook dan de historische doeleinden nog niet gerealiseerd kunnen worden, dan mogen de persoonsgegevens verwerkt worden in hun originele vorm. Hoewel het anonimiseren van gegevens in een digitale context een stuk gemakkelijker is dan gegevens op papier, zou deze werkwijze toch heel wat extra handelingen meebrengt voor de archivaris, en bovendien zouden de websites niet in hun originele vorm kunnen gearchiveerd worden.

Noch de tekst, noch de consideransen van de richtlijn bevatten evenwel een definitie van het begrip “historisch”. Volgens de algemene taalkundige betekenis verwijst de term “historisch” naar verwerkingen van persoonsgegevens met het oog op de analyse van een vroegere gebeurtenis of om die analyse mogelijk te maken. Dit is in essentie wat een archivaris doet: het mogelijk maken om in de toekomst een vroegere gebeurtenis te analyseren. Het Verslag aan de Koning bepaalt echter uitdrukkelijk dat de loutere archivering door de verantwoordelijke voor de verwerking van zijn eigen bestanden niet beschouwd wordt als een bewaring met historische doeleinden en derhalve niet tot het toepassingsgebied van hoofdstuk II behoort.<sup>192</sup> Enkel de archivering door een andere persoon dan de verantwoordelijke lijkt een verwerking voor historische doeleinden te zijn. Het driestapsysteem vervat in hoofdstuk II van het uitvoeringsbesluit is bovendien enkel van toepassing op de latere verwerking voor historische doeleinden die niet verenigbaar zijn met de oorspronkelijke doeleinden waarvoor de gegevens werden verzameld of verkregen.<sup>193</sup> Toegepast op websites betekent dit dat de archivering van persoonsgegevens in het kader van webarchivering als volgt moet verlopen:

- **Bewaren van de eigen website (of naar analogie van andere persoonsgegevens) door de verantwoordelijke voor de verwerking zelf**  
= latere verwerking (niet voor historische doeleinden) maar verenigbaar met oorspronkelijk doel  
→ archivering van deze persoonsgegevens is toegelaten en is onderworpen aan de normale regeling. Alle verwerkingen in de dynamische en semi-statische fase behoren hiertoe + sommige verwerkingen in de statische fase nl. wanneer de archiefdienst waaraan de bestanden zijn overgedragen deel uitmaakt van de verantwoordelijke voor de verwerking bijv. de archiefdienst van een financiële instelling.
- **Bewaren van een website (of naar analogie van andere persoonsgegevens) door een andere persoon dan de verantwoordelijke voor de verwerking**  
= latere verwerking voor historische doeleinden  
In deze situatie moet men zich afvragen of de bewaring in overeenstemming is met het oorspronkelijke doel waarvoor de gegevens verzameld werden, rekening houdend met de redelijke verwachting van de betrokkene. Het al dan niet publieke karakter van de persoonsgegevens kan een leidraad vormen om uit te maken of de betrokkene kon verwachten dat zijn persoonsgegevens later zouden verwerkt worden voor historische doeleinden.

---

<sup>192</sup> Verslag aan de Koning bij het uitvoeringsbesluit, B.S. 13 maart 2001, 7846.

<sup>193</sup> *Ibid.*

- ❑ Zo ja, dan is de archivering van deze persoonsgegevens toegelaten en onderworpen aan de normale regeling (bijv. de archivering door het Antwerpse stadsarchief van de website van lokale verenigingen)
- ❑ Zo nee, dan is de archivering van deze persoonsgegevens toegelaten voorzover de regeling van hoofdstuk II van het uitvoeringsbesluit (bijv. de archivering door een genealoog van een website die hij nodig heeft voor historisch onderzoek) wordt nageleefd.

## IX. DE VLAAMSE OVERHEDEN OP HET INTERNET

### A. DE SITUATIE ANNO 2002

In december 1999 werd het e-Europa-initiatief door de Europese Commissie gelanceerd met als doelstelling Europa on line te brengen. Naar aanleiding van de positieve ontvangst van e-Europa door de lidstaten, het Europees Parlement en de belangrijkste betrokkenen uit de sector, hebben de staatshoofden en regeringsleiders tijdens de op 23 en 24 maart 2000 te Lissabon gehouden Europese Raad een aantal concrete maatregelen met bijbehorende agenda vastgelegd om e-Europa op gang te krijgen. Er werd o.a. op gewezen hoe belangrijk het is dat Europa gebruik gaat maken van de kansen van de nieuwe economie, in het bijzonder het Internet.

Reagerend op deze aanmoediging heeft de Europese Commissie op 24 mei 2000 een actieplan aangenomen, waarover tussen de lidstaten overeenstemming werd bereikt op de Europese Raad van Feira van 19 en 20 juni 2000. Het actieplan is gericht op oplossingen en concentreert zich op wat moet gedaan worden, door wie en wanneer. De actielijnen zijn gegroepeerd rond drie doelstellingen:

- Een goedkoper, sneller en veilig Internet
- Investeren in mensen en vaardigheden
- Stimuleren van Internetgebruik

In het kader van deze laatste doelstelling wordt er verzocht om op alle niveaus van de overheidsdiensten nieuwe technologieën te gebruiken teneinde overheidsinformatie zo toegankelijk mogelijk te maken en om een algemene elektronische toegang tot belangrijke overheidsdiensten te realiseren tegen 2003.<sup>194</sup>

Recent onderzoek doet het ergste vrezen. België en Vlaanderen worden in zeer weinig internationale studies en rapporten vermeld als een voorbeeld van een land met hoge e-government ambities en vergevorderde realisaties.<sup>195</sup> Het moet inderdaad gezegd, er is nood aan een inhaalbeweging ten opzichte van sommige andere lidstaten. De aanwezigheid van de federale zowel

<sup>194</sup> [http://europa.eu.int/information\\_society/eeurope/action\\_plan/pdf/actionplan\\_nl.pdf](http://europa.eu.int/information_society/eeurope/action_plan/pdf/actionplan_nl.pdf)

<sup>195</sup> GOESAERT, J., 'E-Government: Do New Technologies Build a Bridge between Government and Citizen?', in *A Decade of Research @ the Crossroads of Law and ICT*, DUMORTIER, J. (ed.), Gent, Larcier, 2001, 92.

als van de Vlaamse overheid op het Internet is nog verre van voldoende gerealiseerd. Dit heeft onder andere te maken met het uitblijven van een veilig organisatorisch en technisch kader voor digitale handtekening-toepassingen. Op dit moment wordt er op federaal niveau aan de implementatie van de elektronische identiteitskaart gewerkt.<sup>196</sup> Het gevolg van de afwezigheid van een technisch kader is dat dienstverlening op Internet nog beperkt is. De meeste Vlaamse overheidswebsites staan op dit moment nog nergens wat betreft interactieve on line dienstverlening en hun democratisch participatiegehalte.<sup>197</sup> Bijna alle sites gaan uit van eenrichtingsverkeer van de overheid naar de burger. Men kan ze vergelijken met folders waarin allerlei nuttige informatie voor de burger is opgenomen, zoals een cultuuragenda, een gemeentelijke dienstengids enz. Meer en meer komt ook de mogelijkheid voor om via e-mail aan de overheid vragen te stellen, maar enkel voor aangelegenheden waarvoor het niet van belang is om met zekerheid te kunnen vaststellen wie de surfende burger is.

Toch maakt de Vlaamse overheid de laatste jaren in toenemende mate gebruik van informatie- en communicatietechnologieën voor het verwerken, ontsluiten en verspreiden van informatie en worden er heel wat inspanningen geleverd om dit nog op te drijven. Er heerst een algemeen gevoel dat men de e-Europa-trein niet mag missen. Hoewel de Vlaamse regering het e-loket hoog op de politieke agenda heeft geplaatst, is het interactief gebruik van het Internet zoals gezegd nog niet voor morgen. Ondanks deze (voorlopige) beperking haasten de meeste Vlaamse overheden zich om aanwezig te zijn op het Internet.

De eerste gemeentelijke websites in Vlaanderen kwamen er 1995 (Antwerpen, Gent, Kortrijk, Knokke-Heist). In het najaar van 1997 had net iets meer dan 1 op 10 Vlaamse gemeenten een eigen website op het Internet en in september 1998 was dat reeds 3 op 10 geworden.<sup>198</sup> Volgens een laatste stand van zaken is net geen 60 % van de Vlaamse lokale overheden on line.<sup>199</sup> De Vlaamse overheid is op het Internet vertegenwoordigd via haar portaalsite: [www.vlaanderen.be](http://www.vlaanderen.be). Ook de vijf Vlaamse provincies hebben een eigen website.<sup>200</sup>

## B. CASE: HET ELEKTRONISCH LOKET BEVOLKING VAN CEVI

Cevi, het subregionale computercentrum van de provincies Oost- en West-Vlaanderen, heeft een e-loket opgezet ten behoeve van zijn klanten. Elk van de 93 gemeentebesturen die nu klant zijn van Cevi voor de toepassing bevolkingsadministratie, kunnen de diensten van dat elektronisch loket ter beschikking stellen van hun burgers.

De burger die de website van zijn gemeentebestuur raadpleegt, dient de knop van het elektronisch loket aan te klikken. In Gent kan de burger bijvoorbeeld een afschrift uit de eigen huwelijksakte via het Internet aanvragen. Hij kan ook de stand van een dossier nagaan. Wanneer een burger een document wenst te bestellen, dient hij zich aan te kondigen met zijn rijksregisternummer (= nummer

<sup>196</sup> Cf. infra

<sup>197</sup> In 1999 stelde JO STEYAERT nog vast dat bijna de helft van de gemeentelijke sites nauwelijks de naam 'website' waardig was. Ze bleken namelijk 'in opbouw', niet te functioneren of over slechts één pagina te beschikken. Zie STEYAERT, J., 'Het Internet in Vlaanderen: zijn steden, zijn gemeenten, zijn inwoners...', in *Digitale steden en gemeenten. Handleiding*, GOUBIN, E. (ed.), 1999, Politeia, Brussel, 36.

<sup>198</sup> STEYAERT, J., GOUBIN, E. en PLEES, Y., 'Digitale gemeentelijke communicatie in Vlaanderen: de cijfers', in *Digitale steden en gemeenten in Vlaanderen. Een stand van zaken*, STEYAERT, J. (ed.), Brussel, Politeia, 2000, 39.

<sup>199</sup> Toestand op 13 juni 2001: <http://belgium.fgov.be/links/1141.htm>

<sup>200</sup> <http://www2.cipal.be/cipal/flinks2.html>

rechts bovenaan op de SIS-kaart). De controle in het rijksregister of in het bevolkingsregister gebeurt dan automatisch. De documenten worden via de klassieke post naar het officieel adres van de burger gestuurd. Hoewel er nog niet gewerkt wordt met een unieke *identifier* maar met het rijksregisternummer (dat in principe ook op de identiteitskaart voorkomt en dus niet erg geheim is), is dit de meest geavanceerde vorm van e-loket in Vlaanderen op dit moment.

Het gemeentebestuur dat een e-loket wil aanbieden, dient over een eigen website te beschikken. Op deze website wordt er een verwijzing geplaatst naar de URL van het e-loket, die gevestigd is op een beveiligde webserver van Cevi. Op dit moment kan het gemeentebestuur in geen geval bevolkingsgegevens tonen via deze website omdat er nog geen sluitende elektronische identificatie van de burger bestaat.

## X. AANSPRAKELIJKHEID VAN DE OVERHEID VOOR HAAR EIGEN WEBSTEK

### A. KAN DE BURGER RECHTEN ONTLENEN AAN DE INHOUD VAN OVERHEIDSWEBSES?

Buiten het feit dat het technologisch kader voor een interactieve e-overheid op het Internet nog wat op zich laat wachten, kan men nog een andere reden bedenken waarom de Vlaamse overheid een voorzichtig informatiebeleid voert. Dit kan te maken hebben met de onduidelijkheid die er momenteel bestaat omtrent aansprakelijkheidvraagstukken.

Het gebruik van ICT bij de overheidsdienstverlening kan tot heel wat problemen aanleiding geven, waarbij de burger schade kan leiden. De volgende situaties zijn denkbaar:

- De transmissie van een bericht afkomstig van een overheidsinstelling loopt mis of bereikt de burger met vertraging
- Een onbevoegde neemt kennis van de inhoud van een elektronisch bericht
- Er is sprake van onbevoegd gebruik van een elektronische handtekening

Een burger kan ook schade leiden die het gevolg is van door de overheid op het Internet ter beschikking gestelde onjuiste, verouderde of onvolledige informatie. De vraag rijst dan of de burger er toch rechten kan aan ontnemen. Hier moeten we een onderscheid maken tussen twee soorten informatie:

- Informatie die strijdig is met de wet

Het probleem van foute of onvolledige informatie is eerder een zaak van aansprakelijkheid dan van het al dan niet kunnen ontnemen van rechten door de burger aan deze informatie. Immers, men heeft slechts een bepaald recht als er een beslissing van de overheid is voorafgegaan die aan de burger dat recht toekent. Deze beslissing kan de vorm aannemen van een wet, een K.B., een gemeentelijk

politiereglement enz. De burger kan nagaan welke rechten hij heeft aan de hand van het Belgisch Staatsblad. Hierin worden alle algemeen verbindende bepalingen die aan de burgers rechten toekennen bekendgemaakt.<sup>201</sup> Het is ook deze versie die als enige authentieke wordt aangemerkt.<sup>202</sup>

De overheid kan niet gebonden zijn door inlichtingen die zij verstrekt en die strijdig zijn met de wet. De burger kan er dus geen rechten aan ontleen. Met andere woorden: men kan van de overheid niet verlangen dat zij de wet zou moeten schenden, ook niet wegens bepaalde foutief verstrekte inlichtingen. Stel dat er op een website door de overheid informatie wordt verstrekt met betrekking tot het verschuldigde belastingbedrag. De belastingplichtige beschikt dan achteraf niet over een titel om het bedrag van de aanslag terug te brengen tot het aangekondigde niveau. Hij zal uiteindelijk immers slechts betalen wat de wet voorschrijft en lijdt dus geen schade.<sup>203</sup>

- Andere foutieve, verouderde of onvolledige informatie

Voor andere foutieve, verouderde of onvolledige informatie, die op zich niet in strijd is met één of andere wettelijke bepaling, geldt dat de gemeenrechtelijke aansprakelijkheidsregels van toepassing zijn.<sup>204</sup> Ook de overheid is aansprakelijk voor schade die ontstaat door haar optreden. Dit principe werd voor het eerst vastgelegd in het bekende Flandria-arrest van het Hof van Cassatie.<sup>205</sup> Geen enkele grondwettelijke of wettelijke bepaling of algemeen beginsel stelt de administratieve overheid vrij van de verplichting ex de artikelen 1382 en 1383 van het Burgerlijk Wetboek, inclusief de daar opgelegde zorgvuldigheidsplicht.<sup>206</sup>

De delictuele aansprakelijkheid van artikel 1382 B.W. draait rond drie begrippen. Dit betekent dat de burger die zich heeft laten leiden door foutieve, onvolledige of verouderde overheidsinformatie op een website slechts aanspraak kan maken op een schadevergoeding als de overheid een *fout* heeft begaan en de burger daardoor (*oorzakelijk verband*) schade heeft geleden.

Een *fout* houdt een onrechtmatig handelen in. Onrechtmatig handelen kan bestaan uit de schending van een rechtsregel, ofwel uit een inbreuk op wat men noemt de ‘algemene zorgvuldigheidsnorm’. Voor de overheid bestaat er nergens een echte verplichting opgelegd door een rechtsregel om via de eigen website correcte en volledige informatie te verstrekken.<sup>207</sup> Het niet zorgvuldig opstellen en bijwerken van de inhoud van de website zal voor de overheid daarom eerder als een schending van de algemene zorgvuldigheidsnorm aangemerkt worden. De algemene zorgvuldigheidsnorm is een ongeschreven regel volgens dewelke iedereen, ook de overheid, zich in het maatschappelijk verkeer

<sup>201</sup> Art. 190 Grondwet.

<sup>202</sup> Van Orshoven, P., *Bronnen en beginselen van het recht*, Leuven, Acco, 1994, 47.

<sup>203</sup> GIJSSELS, J., ‘De overheidsaansprakelijkheid in verband met informatie’, *Rechtskundig Weekblad*, 1979-1980, (1202), 1212.

<sup>204</sup> *Overheidsinformatie: een essentiële hulpbron voor Europa. Groenboek over overheidsinformatie in de informatiemaatschappij*, Europese Commissie, COM (1998) 585.

<sup>205</sup> Cass., 5 november 1920, *Pas.*, 1920, I, 223-239.

<sup>206</sup> Cass., 20 juni 1997, *Arr. Cass.*, 1997, II, 677.

<sup>207</sup> Er bestaat bovendien voor de overheid geen verplichting om allerlei informatie (zoals toeristische informatie, contactgegevens van overheidsdiensten...) via een eigen website of op een andere manier op eigen initiatief ter beschikking te stellen, de regels van de actieve openbaarheid van bestuur buiten beschouwing gelaten. GIJSSELS kwalificeert deze inspanningen als ‘te waarderen blijken van vrijwillige hulpvaardigheid die zowel de burgers als de administratie ten goede komen.’

moet gedragen als een normaal zorgvuldig en vooruitziend persoon die zich in dezelfde omstandigheden bevindt.

De fout moet bovendien aan de overheid toerekenbaar zijn. Een onrechtmatige handeling zal niet toerekenbaar zijn als er sprake is van een verschonings- of rechtvaardigingsgrond, zoals onoverwinnelijke dwaling of overmacht. Door de voorbeeldfunctie van de overheid kunnen we aannemen dat een verschoningsgrond niet snel aanwezig zal zijn. Websites moeten samengesteld worden met dezelfde zorg als andere informatiebronnen van de overheid, zoals folders of tv-spots.

De burger zal, buiten de fout van de overheid vervolgens ook moeten aantonen dat hij *door deze fout schade* heeft geleden. Het begrip schade wordt zeer ruim verstaan. Het kan gaan om een geleden verlies, maar ook om een gemiste kans (bijv. bij sollicitaties). Of er werkelijk schade is geleden, zal o.m. afhangen van de inhoud van de verstrekte informatie. Gaat het om vrij vage informatie, losse commentaren of vrijblijvende raadgevingen, dan zal de betrokkene het bestuur bezwaarlijk gerechtelijk ter verantwoording kunnen roepen en schadevergoeding vorderen. De schade zal daarentegen veel sneller bewezen worden geacht bij werkelijk foutieve of onvolledige informatie, die de betrokkene in dwaling brengt over zijn recht, of zijn rechtstoestand schade toebrengt. Dit is uiteraard een feitenkwestie waarover de rechter zich in laatste instantie zal uitspreken.<sup>208</sup>

Om de fout van de overheid aan te tonen, moet de burger het bewijs leveren dat er op een bepaald moment foutieve, onvolledige of verouderde informatie op de website stond. Het bewijsrecht is bijgevolg een belangrijk argument waarom de websites van de overheid best op geregelde tijdstippen gearchiveerd worden.

## B. ZIN OF ONZIN VAN DISCLAIMERS

Op heel wat websites kom je exonerationeclausules (*disclaimers*) tegen. Dit zijn clausules waarin de titularis van de website verklaart niet aansprakelijk te zullen zijn voor de inhoud ervan.<sup>209</sup> Ook op sommige overheidswebsites komen dergelijke exonerationeclausules voor.<sup>210</sup> Mag de overheid haar aansprakelijkheid zomaar afwentelen waardoor de burger in de kou blijft staan? Het antwoord hierop is niet eenduidig.

De principiële geldigheid van exonerationeclausules wordt in België erkend, en dit zowel wat de contractuele als wat de buitencontractuele aansprakelijkheid betreft.<sup>211</sup> Dit wil zeggen dat niet alleen een debiteur de wettelijke aansprakelijkheid wegens niet-nakoming van zijn verbintenissen kan beperken, maar ook dat een schadeverwekker zich ten aanzien eventuele schadelijders kan ontlasten van zijn delictuele aansprakelijkheid.

De vraag is nu of ook overheden zich op voorhand kunnen exoneren van aansprakelijkheid uit onrechtmatige daad, zoals door het plaatsen van disclaimers op hun website. In het algemeen worden exonerationeclausules ten voordele van de overheid streng beoordeeld.<sup>212</sup> Sommige auteurs vinden een

<sup>208</sup> GJJSSELS, J., *l.c.*, 1214.

<sup>209</sup> Bijv. <http://www.envida.net/nl/about/terms>

<sup>210</sup> Bijv. [http://www.europa.eu.int/geninfo/disclaimer\\_nl.htm](http://www.europa.eu.int/geninfo/disclaimer_nl.htm)

<sup>211</sup> Cass., 21 februari 1907, *Pas.*, 1907, I, 135; Cass., 29 september 1972, *Arr. Cass.*, 1973, 121.

<sup>212</sup> VANDENBERGHE, H., 'Exoneratie- en vrijwaringsbeding bij onrechtmatige daad. Samenloop en coëxistentie', in *Exoneratiebedingen*, HERBOTS, J. (ed.), Brugge, Die Keure, 1993, 81.

exoneratieclausule geoorloofd als het niet de bedoeling heeft afbreuk te doen aan de wezenlijke taken die de wet aan de overheid toevertrouwt.<sup>213</sup> Het opzetten en onderhouden van een overheidssite behoort niet tot de essentiële overheidstaken en bijgevolg zou zo'n clausule geoorloofd zijn. Anderen situeren ongeoorloofde exoneratieclausules dan weer daar waar de burger verplicht gebruik moet maken van sommige diensten.<sup>214</sup> Het gebruik van overheidswebsites, zelfs van het e-loket, is en blijft vooralsnog facultatief.

Samen met PRINS zijn wij van oordeel dat een overheid die serieus werk maakt van een betrouwbaar elektronisch imago niet alle mogelijke schadeclaims mag afwijzen via een op het Internet geplaatste exoneratieclausule. De risico's zullen via andere wegen afgedekt moeten worden bijv. via een verzekering. Om deze risico's te verminderen dient er serieus aandacht te worden besteed aan de beschikbaarheid, de betrouwbaarheid en de integriteit van zowel de communicatie als van de infrastructuur.<sup>215</sup> Het uitbouwen van een efficiënte PKI-infrastructuur is hiervoor het middel bij uitstek. Anderzijds moet de inhoud van de website permanent opgevolgd worden en met grote zorg worden samengesteld. Er moet ook op toegezien worden dat inlichtingen die via het e-loket gegeven worden, correct zijn.

## XI. PORTAALSITES: DE TOEKOMST

### A. VAN INFORMATIE NAAR INTERACTIE EN INTEGRATIE

Overheidswebsites zijn zoals gezegd vandaag de dag zeer beperkt in hun opzet. De meeste overheidsdiensten bieden weliswaar afzonderlijk elektronische informatie aan. Deze informatie volgt echter steeds de logica van de overheid en niet die van de burger. Bovendien is de dienstverlening niet geïntegreerd. De burger moet in een wirwar van websites de relevante informatie zien te vinden, zonder voldoende garanties op volledigheid en juistheid. Tenslotte worden er door de overheidsdiensten nauwelijks elektronische transacties met de burger aangeboden, waarbij er een elektronische tweerichtingscommunicatie tot stand komt en er integratie is met de achterliggende workflow bij de verschillende overheidsdiensten.

In de toekomst moeten de bestaande overheidswebsites, met behulp van de elektronische identiteitskaart, uitgroeien tot gebruiksvriendelijke front office-sites<sup>216</sup>, waar de burger een logisch geordend en geïntegreerd aanbod van diensten kan aantreffen, gaande van informatieverstrekking tot transacties. Elektronische dienstverlening door de overheid via Internet is slechts één onderdeel van e-government. Het is voor de burger het meest zichtbare gedeelte, maar een goed georganiseerde

<sup>213</sup> MAUSSION, F., 'Reflexions sur la théorie de l'organe', in *La responsabilité des pouvoirs publics*, Brussel, 1991, 96 e.v.

<sup>214</sup> VAN HOOYDONK, E., 'De geldigheid van in havenreglementen opgenomen bevrijdingsbedingen', *Rechtskundig Weekblad*, 1990-1991, 1394.

<sup>215</sup> PRINS, J., E-overheid: Evolutie of revolutie? <http://rechten.kub.nl/prins/Publicatnl/eoverheprev.pdf>

<sup>216</sup> De term 'front office' wordt gebruikt om de elektronische dienstverlening door de overheid naar de burger toe aan te geven, in tegenstelling tot 'back office' waarmee men het elektronische verkeer binnen de overheidsdiensten en tussen de overheidsdiensten onderling bedoelt.

gegevensuitwisseling binnen en tussen de verschillende overheidsdiensten is minstens even essentieel om de portaalsites optimaal te laten functioneren. Informatie moet dan door de burger slechts eenmalig aan de overheid worden meegedeeld, waardoor het contact met de overheid kan worden verminderd. Op die manier hoeft de burger ook de deur niet meer uit om ergens tussen het kastje en de muur lange wachttijden te trotseren om een papieren attest te bemachtigen, dat dan nadien aan een andere overheidsdienst moet overhandigd worden.

## **B. ÉÉN VIRTUELE OVERHEID**

De uitwerking van dit ambitieuze e-government project gebeurt op dit moment vooral op federaal niveau. Een nieuwe federale overheidsdienst informatie- en communicatietechnologie (FEDICT) coördineert het hele gebeuren. Op dit moment wordt er gewerkt aan twee gebruiksvriendelijke portaalsites, die zijn opgebouwd volgens de logica van de gebruiker en waarmee er interactie mogelijk is met de interne informatiesystemen van de verschillende federale overheidsdiensten. De overheidsopdracht om de portalen te bouwen werd in september 2001 uitgeschreven.

Ondanks de verdeling van bevoegdheden en taken over verschillende overheidsniveaus (gemeenten, provincies, gemeenschappen, gewesten, federale overheid, Europa), ervaart de burger de overheid als één geheel. Daarom werd er een akkoord gesloten tussen de federale overheid en de gemeenschappen en de gewesten om in hun elektronische dienstverlening één virtuele overheid te realiseren. Praktisch gezien komt dit er op neer dat de burger via de portaalsite van de federale overheid wordt doorverwezen naar de overheid die bevoegd is voor de vraag of transactie waarmee de burger zich tot de overheid richt. Op de portaalsites van de andere overheidsniveaus zullen er anderzijds ook verwijzingen worden geïntegreerd naar de informatie en transacties beschikbaar op het federale niveau. In het kader van deze samenwerking verbinden de verschillende overheidsniveaus zich ertoe om dezelfde hulpmiddelen (basissoftware, elektronische identiteitskaart met elektronische handtekening, unieke identificatiesleutel) aan te bieden aan de burger.

## **C. HET WETTELIJKE KADER**

### **C.1. Front office: de elektronische identiteitskaart**

Wanneer de communicatie van de burger naar de overheid toe via het anonieme Internet zal verlopen, is het nodig dat deze communicatie beveiligd wordt: de overheid moet zeker weten welke burger zich aan het e-loket aanmeldt en de vertrouwelijkheid en de integriteit van de berichten moet gegarandeerd zijn. Om dit te realiseren, heeft de Ministerraad op 20 november 2000 een nota goedgekeurd over het opzetten van een federale PKI-infrastructuur en het gebruik van een elektronische identiteitskaart door alle burgers. Op 19 juli 2001 werd dan de beslissing genomen om de elektronische identiteitskaart voor natuurlijke personen in te voeren. Het wettelijk kader voor de elektronische identiteitskaart wordt geboden door de Europese richtlijn inzake elektronische handtekeningen, omgezet in Belgisch recht op 9 juli 2001 door de wet certificatedienstverleners. Momenteel ligt er een voorstel klaar om de nationale wetgeving op de identiteitskaart aan te passen. De elektronische identiteitskaart zal de volgende functies kunnen vervullen:

- Identificatie van de houder
- Authenticatie van de houder (nl. het bewijs dat de houder van de kaart wel degelijk de persoon is die door de kaart wordt geïdentificeerd)
- Middel om een elektronische handtekening met juridische waarde te plaatsen
- Elektronisch bewijs van een hoedanigheid, een mandaat of een kenmerk van de houder, op initiatief van de houder
- Drager van programma's die binnen de chip van de kaart kunnen worden uitgevoerd (bijv. voor het genereren van sleutelparen)

De elektronische identiteitskaart zal op visuele en elektronisch leesbare wijze volgende identificatiegegevens bevatten:

- Naam en voornaam
- Geslacht
- Geboorteplaats
- Geboortedatum
- Nationaliteit
- Rijksregisternummer
- Foto

De kaart heeft dus twee functies. Enerzijds kan de burger zich met zijn kaart in een elektronische omgeving op een controleerbare manier identificeren en anderzijds wordt het plaatsen van een elektronische handtekening mogelijk, bijv. op officiële overheidsdocumenten. Daartoe worden in de processorchip van de kaart twee private sleutels<sup>217</sup> met bijhorende identiteitscertificaten opgenomen, tenzij de houder zich hiertegen verzet. Het gebruik van de private sleutels wordt beveiligd met een PIN-code. De identiteitscertificaten zullen worden afgeleverd door een certificatedienstverlener gekozen door de overheid op basis van een overheidsopdrachtenprocedure, met de bedoeling om een lage prijs hiervoor te bekomen.

De elektronische identiteitskaart zal uitgereikt worden door de gemeenten ter vervanging van de huidige plastieke identiteitskaart en zal een geldigheidsduur hebben van vijf jaar. De lokale bevolkingsdiensten zullen de kaarten verkopen voor 9 EURO. De identificatiegegevens worden op een onveranderlijke wijze op de kaart opgeslagen. Indien één van deze gegevens wijzigt, dient er een nieuwe elektronische identiteitskaart te worden uitgereikt. In de loop van 2002 zal de elektronische identiteitskaart worden uitgereikt aan de inwoners van elf pilootgemeenten. Aan de burgers van deze gemeenten wordt de mogelijkheid geboden om hun huidige kaart al onmiddellijk in te ruilen voor een elektronische kaart. Kaarthouders wiens huidige kaart vervallen is, zullen in ieder geval een nieuwe elektronische kaart krijgen. Op die manier zouden binnen een periode van vijf jaar alle kaarten moeten zijn vernieuwd. Na een verdeelperiode van zes maanden zal de Ministerraad beslissen of het systeem klaar is om over heel het land te worden verspreid.

---

<sup>217</sup> Voor de beginselen van *Public Key Infrastructure*: zie het DAVID-rapport *Wat en hoe archiveren? Op zoek naar de rol van PKI voor digitale archieven*.

De elektronische identiteitskaart kan ook een belangrijke toegevoegde waarde betekenen voor de verzegeling door de archivaris van elektronische documenten. Het is immers niet alleen de bedoeling om e-government toepassingen mogelijk te maken, maar ook andere toepassingen die een gecertificeerde, elektronische identiteit vereisen. In de processorchip van de elektronische identiteitskaart zal de houder een beperkt aantal bijkomende sleutelparen kunnen bewaren waaraan kenmerkcertificaten zijn verbonden en waarmee hij dus elektronisch de hoedanigheid van bijvoorbeeld archivaris kan bewijzen.

## C.2. Back office: uniek identificatienummer

Het hebben van een website waar de burger zich aan de hand van zijn elektronische identiteitskaart kan kenbaar maken, is uiteraard niet voldoende. Een eerste essentiële voorwaarde voor een efficiënte e-overheid is een eenduidige identificatie van de entiteiten waarover informatie wordt uitgewisseld. Momenteel wordt de wettelijke basis gecreëerd voor het gebruik van een uniek identificatienummer voor natuurlijke personen, ondernemingen en organisaties. Dit nummer zal gebruikt worden in alle informatiesystemen van de overheid, zodat een efficiënte elektronische gegevensuitwisseling tussen overheidsdiensten en dus de eenmalige inzameling van gegevens bij burgers en ondernemingen mogelijk wordt. Een gebrek aan een eenduidige identificatie verhindert het optimaal hergebruik van de reeds beschikbare informatie.

Vandaag de dag worden voor de meeste entiteiten door overheidsdiensten één of meerdere nummers gebruikt, die meestal niet stabiel zijn doorheen de tijd (bijv. het handelsregisternummer, RSZ-nummer, BTW-nummer...). Voor de toekomst wordt ernaar gestreefd om slechts één nummer over te houden, dat wel stabiel is doorheen de tijd. Voor natuurlijke personen zal het rijksregisternummer als unieke identificatiesleutel worden gebruikt en zal “persoonsnummer” worden genoemd. Voor personen die nooit in het bevolkings-, vreemdelingen- of wachtregister zijn ingeschreven (en dus geen rijksregisternummer hebben) zal het identificatienummer van de sociale zekerheid worden gebruikt. Voor ondernemingen en organisaties zal het bestaande BTW-nummer worden gebruikt. Dit nummer wordt het “ondernemings- en organisatienummer” genoemd. Andere nummers, zoals het handelsregisternummer en RSZ-nummer, zullen dus in de toekomst niet meer worden gebruikt.

Bij het toekennen van een nummer moet de initiërende overheidsdienst steeds een voor de betrokken entiteit vastgestelde set van basisidentificatiegegevens aanleveren die worden bewaard in een gegevensbank beheerd door de instantie die het identificatienummer toekent. Voor natuurlijke personen vervult het rijksregister deze rol; voor ondernemingen en organisaties zal er een Kruispuntbank Ondernemingen worden opgericht. Er is reeds een voorontwerp van wet uitgewerkt tot oprichting van een Kruispuntbank van Ondernemingen, die vanaf begin 2003 operationeel moet worden. Naar aanleiding van het veralgemeend gebruik van het rijksregisternummer werkt men momenteel aan een herziening van de regelgeving op het gebruik van het rijksregisternummer.

De overheid moet er ook voor waken dat zij bij de burger geen gegevens meer opvraagt die reeds bij een andere overheidsdienst aanwezig is. Daarom zullen de informatiesystemen van de federale overheid onderling verbonden worden in een netwerk waardoor allerlei informatie snel en op een beveiligde manier kan uitgewisseld worden. De wettelijke basis hiervoor wordt geboden door art. 102 programmawet van 30 december 2001 dat bepaalt dat de federale openbare diensten verplicht kunnen worden om de gegevens waarover zij beschikken mee te delen, bij voorkeur langs elektronische weg,

aan andere federale openbare diensten die deze gegevens nodig hebben voor de uitvoering van hun opdracht.<sup>218</sup>

Omdat de informatie binnen de overheid verspreid zit over verschillende platformen, zal er op technisch vlak een *message engine* worden ontwikkeld die het mogelijk moet maken om op intelligente wijze gestructureerde berichten uit te wisselen tussen de informatiesystemen van de federale overheid in ruime zin, tussen de informatiesystemen van de federale overheid en de websites of portalen en tussen de informatiesystemen van de federale overheid en deze van andere overheidsniveaus gemeenschappen, gewesten, gemeenten en provincies. Tussen de overheidsdiensten in het netwerk zal er een taakverdeling worden vastgelegd voor de opslag van gegevens in authentieke vorm. De overheidsdienst die bepaalde gegevens bijhoudt, moet ze ook up-to-date houden. Wanneer een overheidsdienst gegevens nodig heeft, raadpleegt hij eerst het authentieke bronbestand. Zijn de gegevens daar niet beschikbaar, dan mag hij de gegevens inzamelen bij de betrokkene.<sup>219</sup>

## XII. BESLUIT & AANBEVELING

Dat het archiveren van websites een noodzaak is, daar twijfelt na het lezen van dit rapport hopelijk niemand meer aan. Dat dit een complexe aangelegenheid is, waarvoor de oplossingen die de techniek aanreikt nog niet helemaal op punt staan, is eveneens duidelijk. Het archiveren van statische websites levert weinig of geen problemen op. De nieuwste soorten websites zijn echter dynamisch van aard. De inhoud van dergelijke webpagina's kan afhankelijk zijn van factoren zoals het ontvangen verzoek, het gebruikersprofiel of de voorkeuren van de bezoeker of van de informatie die op dat ogenblik in het gekoppelde documentbeheerssysteem of in de achterliggende databank aanwezig is. Een geïntegreerde en volledig functionale archivering van deze websites is niet mogelijk. In de plaats daarvan worden de verschillende componenten van het dynamisch informatiesysteem afzonderlijk gearchiveerd. Een kwaliteitsvolle websitesarchivering is vrij arbeidsintensief en vraagt samenwerking met de archiefvormer. De emulatiestrategie lijkt op dit moment de meest aangewezen oplossing voor de lange termijn leesbaarheid van mirrors en snapshots. Websitesarchivering toont duidelijk de nood aan om van bij de creatie van de archiefdocumenten al aan archivering te denken.

Na de vaststelling (in vorige DAVID-rapporten) dat het telecommunicatiegeheim en de privacyreglementering in het algemeen, onvoldoende rekening houden met de behoeften van instellingen die de preservatie van ons culturele erfgoed als voornaamste taak hebben, ontsnapt ook het auteursrecht niet aan deze aanklacht. De archivering van om het even welk auteursrechtelijk beschermd werk brengt vroeg of laat praktische moeilijkheden door het feit dat elke reproductiehandeling op een rigoreuze wijze wordt voorbehouden aan de auteur. Het probleem stelt zich anno 2001 des te scherper omdat de archivering van digitale werken, en in het bijzonder van werken die via het Internet verspreid worden, reeds van bij het begin reproductie vergen. Waar ons auteursrecht er traditioneel prat op gaat ons culturele erfgoed te beschermen door de reproductie van auteursrechtelijk beschermde werken voor te behouden aan diegene die er de creatieve inspanning

<sup>218</sup> B.S. 31 december 2001.

<sup>219</sup> Zie art. 102 2° en 3° van de programmawet.

voor heeft geleverd, zal dit voorbehoud zoals het nu geformuleerd is er in de toekomst meer en meer toe leiden dat precies het auteursrecht verantwoordelijk is voor de teloorgang van het (digitale) culturele erfgoed. Dit rapport is dan ook een oproep aan de wetgever om in het wetsvoorstel betreffende de hervorming van het auteursrecht dat nu voorligt in de Senaat rekening te houden met de noden van archiefdiensten en -instellingen op dit vlak, en bij uitbreiding van musea en bibliotheken.

## XIII. BIBLIOGRAFIE

### A. DIGITAAL ARCHIVEREN

*A policy for keeping records of web-based activity in the Commonwealth Government*, januari 2001. (beschikbaar op: [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) )

C. AMMEN, *MINERVA: Mapping the INternet Electronic Resources Virtual Archive -Web Preservation at the Library of Congress*, Lezing gehouden te Darmstadt op 8 sept. 2001. (beschikbaar op <http://www.bnf.fr/pages/infopro/dli%5Fecd12001.htm> )

W.Y. ARMS, *Web preservation project. Interim report*, 2001.

A. ARVIDSON, *Harvesting the Swedish webspace*, Lezing gehouden te Darmstadt op 8 sept. 2001. (beschikbaar op <http://www.bnf.fr/pages/infopro/dli%5Fecd12001.htm>)

A. ASCHENBRENNER, *Long-term preservation of digital material. Building an archive to preserve digital cultural heritage from the Internet*, Verhandeling Technische Universiteit Wenen, 2001.

S. BORDWELL, *Objective-Archival preservation of websites*, Lezing gegeven te Londen, 25 april 2002.

M. BURNER en B. KAHLE, *WWW Archive File Format Specification*, (beschikbaar op <http://www.alex.com/company/arcformat.html>)

W. CATHRO, C. WEBB en J. WHITING, *Archiving the web: the pandora archive at the National Library of Australia*, Lezing gehouden tijdens de conferentie *Preserving the Present for the Future Web Archiving*, Kopenhagen, 18-19 juni 2001. (beschikbaar op <http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html>).

B. CHRISTENSEN-DALSGAARD, *Archive Experience, not Data*, Lezing gehouden in Darmstadt, 8 sept. 2001. (beschikbaar op <http://www.bnf.fr/pages/infopro/dli%5Fecd12001.htm>)

C. DOLLAR, *Archival preservation of Smithsonian web resources: strategies, principles, and best practices*, (beschikbaar op <http://www.si.edu/archives/archives/dollar%20report.html>)

I. ENGHOLM, *Digital design history and the registration of web development*, (beschikbaar op <http://www.deflink.dk/eng/arkiv/dokumenter2.asp?id=695>).

IM FORUM, *Government of Canada Internet Guide*, Ottawa, 1999. (beschikbaar op [http://www.cio-dpi.gc.ca/ig-gi/index\\_e.asp](http://www.cio-dpi.gc.ca/ig-gi/index_e.asp))

*Guidelines for keeping records of web-based activity in the Commonwealth Government*, maart, 2001. (beschikbaar op [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html))

J. HAKALA, *Collecting and Preserving the Web: Developing and Testing the NEDLIB Harvester*, in *RLG-DigiNews*, April 15, 2001, vol.5, nr. 2.

J. HAKALA, *Harvesting the Finnish Web space - practical experiences*, Lezing gehouden te Darmstadt op 8 sept. 2001. (beschikbaar op <http://www.bnf.fr/pages/infopro/dli%5Fecd12001.htm>)

J. HONEYCUTT (e.a.), *Het complete handboek Internet*, Schoonhoven, 1997.

B. KAHLE, *Archiving the Internet*, in *Scientific American*, 11 april 1996.

A.R. KENNEY en O.Y. RIEGER, *The National Library of Australia's Digital Preservation Agenda, an interview with C. Webb*, in *RLG-DigiNews*, 15 febr. 2001

D. LÉGER, *Legal Deposit and the Internet: Reconciling Two Worlds*, Lezing gegeven te Darmstadt, 8 sept. 2001. (beschikbaar op <http://www.bnf.fr/pages/infopro/dli%5Fecd12001.htm>)

*Managing Internet and Intranet Information for Long-term Access and Accountability. Implementation Guide*, 1999.

J. MASÉNAS, *The BnF-project for web-archiving*, Lezing gehouden te Darmstadt op 8 sept. 2001. (beschikbaar op <http://www.bnf.fr/pages/infopro/dli%5Fecd12001.htm>)

S. MCKEMMISH en G. ACLAND, *Accessing essential evidence on the web: towards an australian recordkeeping metadata standard*, (beschikbaar op: <http://ausweb.scu.edu.au/aw99/papers/mckemmish/paper.html>)

C.J.M. MOSCHOVITIS, *History of the Internet: a chronology, 1843 to the Present*, Santa-Barbara (California), 1999

K. PERSSON, *The Kulturarw3 Project - The Swedish Royal Web Archiw<sup>3</sup>*, Lezing gehouden in Svetlogorsk, aug. 2000. (beschikbaar op <http://kulturarw3.kb.se>) .

J. QUAST, *Het internetarchief van het IISG*, in *Nederlands Archievenblad*, september 2000, p. 16-17.

A. RAUBER e.a., *Austrian on-line archive. Current status and next steps*, Lezing gehouden te Darmstadt op 8 sept. 2001. (beschikbaar op <http://www.bnf.fr/pages/infopro/dli%5Fecd12001.htm>)

D. SHENK, *The world wide library*, in *Hotwired*, 2 sept. 1997 (beschikbaar op [http://hotwired.lycos.com/synapse/feature/97/35/shenk1a\\_text.html](http://hotwired.lycos.com/synapse/feature/97/35/shenk1a_text.html)).

G. VOERMAN (e.a.), *Het belang van het archiveren van websites*, in *Information Professional*, 2001, p. 16-19

WORLD WIDE WEB CONSORTIUM, *Web Content Accessibility Guidelines 1.0*, 1999. (beschikbaar op <http://www.w3.org/WAI/>)

## **B. WET- EN REGELGEVING**

### **Boeken**

BAINBRIDGE, D., *Introduction to Computer Law*, Gosport, Ashford Colour Press, 2000, 349 p.

BUYDENS, M., *Auteursrechten en Internet – problemen en oplossingen voor het creëren van een online databank met beelden en/of tekst*, Brussel, Federale diensten voor wetenschappelijk, technische en culturele aangelegenheden, 1998, 104 p.

CLAPPAERT, W., ‘Auteursrecht en Internet’, in *Telecom en Internet. Recht in beweging*, DE POORTER, B. (ed.), Gent, Mys en Breesch, 1999, 357-376.

CORBET, J., *Auteursrecht*, Algemene Praktische Rechtsverzameling, Antwerpen, Kluwer, 1997, 173 p.

DE COCK BUNING, M., *Auteursrecht en Informatietechnologie – Over de beperkte houdbaarheid van technologiespecifieke regelgeving*, Amsterdam, Otto Cramwinkel, 1998, 281 p.

GOTZEN, F., *Auteursrecht, tekeningen en modellen*, KUL Faculteit Rechtsgeleerdheid, 1998, 249 p.

HUGENHOLTZ, P. (ed.), *The future of copyright in a digital environment*, Antwerpen, Kluwer, 1996, 248 p.

SPOOR, J., *Scripta Manent: de reproductie in het auteursrecht*, Groningen, Tjeenk Willink, 1976, IV, 140 p.

VANHEES, H., *Auteursrecht in een notendop*, Leuven, Garant, 1998, 137 p.

VOORHOOF, D., ‘Multimedia en auteursrecht. Afschermen en beschermen van informatie. Juridische problemen rond de beschikbaarheid en de reproductie van informatie op één drager’, in *Multimedia, Interactiviteit, Kennisspreiding*, verslagboek, Symposium van de Stichting Boek, 14 en 15 november 1995, LUC Diepenbeek.

WESTERBRINK, B.N., *Juridische aspecten van het Internet*, Amsterdam, Otto Cramwinkel, 1996, 195 p.

## Artikels

ALBERDINGK THIJM, C., 'Fair use: het auteursrechtelijk evenwicht hersteld',  
[http://www.ivir.nl/publicaties/overig/alberdingk\\_thijm/fair-use.doc](http://www.ivir.nl/publicaties/overig/alberdingk_thijm/fair-use.doc)

DECKER, U., 'Electronic Archives and the Press: Copyright Problems of Mass media in the Digital Age', *E.I.P.R. issue 7*, Sweet & Maxwell Limited, 1998, 256-265.

GIJSSELS, J., 'De overheidsaansprakelijkheid in verband met informatie', *Rechtskundig Weekblad*, 1979-1980, 1202-1222.

GOESAERT, J., 'E-Government: Do New Technologies Build a Bridge between Government and Citizen?', in *A Decade of Research @ the Crossroads of Law and ICT*, DUMORTIER, J. (ed.), Gent, Larcier, 2001, 87-100.

HUGENHOLTZ, P., *DIPPER Digital Intellectual Property Practice Economic Report – Copyright aspects of caching*, Institute for Information Law, University of Amsterdam, September 1999, 46 p.

HUGENHOLTZ, P., 'De databankrichtlijn eindelijk aanvaard: een zeer kritisch commentaar', *Computerrecht*, 1996/4,

HUGENHOLTZ, P., *Recht en Internet*, Nederlands juristenblad, Zwolle, Tjeenk Willink, 1998.

LANGE, J., 'De wet en elektronische publikaties',  
Beschikbaar op: <http://nieuws.surfnet.nl/nieuws/snn-archief/achtergrond/jg97-98/kbsymposium2.html>

PRINS, J.E.J., 'Digitale duurzaamheid : een verloren geschiedenis' Beschikbaar op :  
<http://infolab.kub.nl/till/data/topic/digiduur.html>

SPOOR, J., 'The copyright approach to copying on internet: (over)stretching the reproduction right?', in *The future of copyright in a digital environment*, HUGENHOLTZ, P. (ed.), Antwerpen, Kluwer, 1996, 248 p.

STEYAERT, J., 'Het Internet in Vlaanderen: zijn steden, zijn gemeenten, zijn inwoners...', in *Digitale steden en gemeenten. Handleiding*, GOUBIN, E. (ed.), 1999, Politeia, Brussel, 23-36.