



# WEBSITESBEHEER EN CONTENT MANAGEMENT VOOR DIGITALE ARCHIVERING

Versie 2.0 (mei 2005)

## 1. Woord vooraf

De archivering van websites kan door verschillende redenen worden gemotiveerd:

- de inter/intranetsite van een openbare dienst kan de handelingen en beslissingen van de burger of andere diensten bepalen. De dienst moet hierover verantwoording kunnen afleggen.
- de overheid is aansprakelijk voor de inhoud van zijn websites
- de inter/intranetsites van openbare diensten zijn bestuursdocumenten
- alsmaar meer informatie is enkel op Inter/intranet aanwezig, en is niet meer in een andere vorm bij de diensten aanwezig
- informatie op websites is vluchtig en kan snel verloren gaan
- websites behoren tot ons digitaal erfgoed

Websites zijn maar doeltreffend archiveerbaar wanneer ze op een efficiënte wijze werden ontworpen en beheerd. Een website die hieraan voldoet, biedt twee voordelen:

- de website is toegankelijk voor zoveel mogelijk bezoekers. Een aantal richtlijnen voor het ontwerp en beheer zijn immers gebaseerd op de basisregels van *Web Content Accessibility*. Zondigen tegen deze regels betekent dat je niet het volle potentieel aan bezoekers bereikt.
- de website kan (nagenoeg) rechtstreeks van de webserver worden gearchiveerd. De aanpassingen die de website voor archivering nog moet ondergaan zijn minimaal.

Deze richtlijn bestaat uit twee delen. Het eerste deel is in de eerste plaats voor websitebouwers bedoeld. Van bij de creatie van een website dient immers al rekening worden gehouden met archivering. Zij moeten ook zorgen voor de nodige voorzieningen zodat inhoudsverantwoordelijken hun webpagina's en aanpassingen kunnen documenteren. In dit deel worden vereisten opgesomd voor de webpagina's en voor content management systemen.

Het tweede deel bevat richtlijnen voor het aanleggen en bijhouden van een website-archief. Dit website-archief kan lokaal (bij de diensten) of centraal (bij de archief- of de IT-dienst) worden bijgehouden. Voor de archiveringsrichtlijn wordt uitgegaan van een minimale IT-infrastructuur. Men dient enkel over een off line browser te beschikken. De archiveringsrichtlijn geldt voor zowel websites met een vaste als een dynamische inhoud. Tevens wordt aangegeven hoe elke versie of elke wijziging kan worden gearchiveerd. Deze richtlijn is een concrete toepassing van de algemene websitesarchiveringsstrategie die in [DAVID-rapport nr. 6](#) wordt beschreven.

## 2. WAT en HOE archiveren?

### 2.1 WAT archiveren?

Bij een website komen volgende onderdelen voor archivering in aanmerking:

- de webpagina's
- het deep web
- de neerslag van transacties en handelingen
- de metadata

Op basis van de identificatie van de archiefdocumenten en archiefwaardering worden de componenten van een website voor (middel)langetermijnarchivering geselecteerd. Meer informatie is beschikbaar in:

- F. BOUDREZ, *Archiveren van websites: een kwestie van waardering en 'capture'*, Antwerpen, 2005, p. 6 e.v.

### 2.2 Webpagina's: HOE archiveren?

Er zijn vier methoden om de webpagina's van een website te archiveren:

- archiveren van de broncode
- archiveren van een snapshot
- archiveren van de unieke responsen
- archiveren van een surfsessie.

Meer informatie is beschikbaar in:

- F. BOUDREZ, *Archiveren van websites: een kwestie van waardering en 'capture'*, Antwerpen, 2005, p. 10 e.v.

## 3. Richtlijnen voor het ontwerpen en het beheren van een website

Onderstaande richtlijnen zijn gebaseerd op het DAVID-onderzoek naar de kwaliteitsvereisten voor archiveerbare en gearchiveerde websites en op de eerste ervaringen van het stadsarchief Antwerpen inzake websitesarchivering. Deze richtlijnen stemmen in grote mate overeen met de *Web Content Accessibility*-regels van het World Wide Web Consortium (1999, <http://www.W3C.org/WAI/>), zodat naleving ook een meer toegankelijke website oplevert.


### 3.1 Richtlijnen voor de webpagina's

#### 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.


### Lange termijn planning:

 denk op lange termijn bij het starten en lanceren 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 op de webserver een duidelijke mappenstructuur. Volg de richtlijnen in [Digitaal Archiveren: richtlijn & advies nr. 3. over Mappenstructuur en bestandsnamen voor digitale documenten](#).


 plaats alle bestanden en mappen 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!).

### Webpagina's:

 respecteer de officiële standaarden voor mark-up-talen: vermijd het gebruik van de producenteigen of browserafhankelijke uitbreidingen op deze standaarden. Maak geen gebruik van DHTML of van tags die 'deprecated' of afgekeurd zijn verklaard. De officiële mark-up standaarden zijn:

- HTML 4.01: gebruik de strict variant
- XHTML 1.0: gebruik de strict variant

 gebruik markup en stylesheets waarvoor ze ontworpen zijn. 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/xhtml1-strict.dtd">
```

- header: <HEAD></HEAD>
- body: <BODY></BODY>

 ken aan elke webpagina een titel toe:

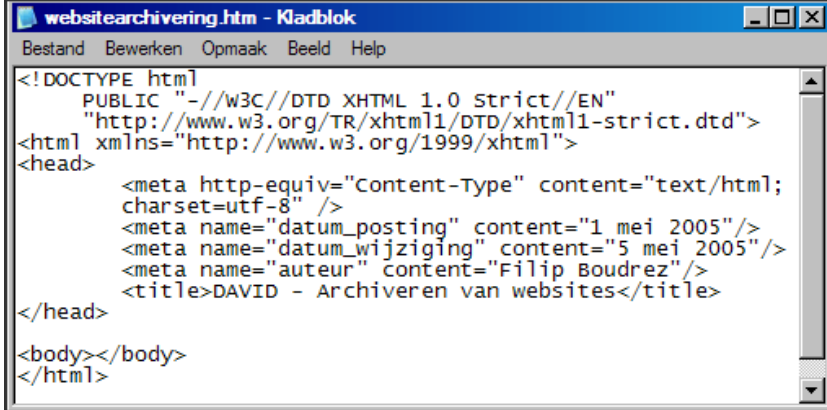
```
<TITLE>Richtlijn webdesign</TITLE>
```

 definieer de gebruikte karakterset voor elke webpagina. Gebruik dezelfde karakterset voor de volledige website

 zorg ervoor dat elke webpagina de nodige metadata en contextuele gegevens bevat:





- datum posting:  
`<META NAME="datum_posting" CONTENT="1 mei 2005">`
- datum wijziging:  
`<META NAME="datum_wijziging" CONTENT="5 mei 2005">`
- naam webdesigner/auteur:  
`<META NAME="auteur" CONTENT="Filip Boudrez">`

#### Voorbeeld:



```
websitearchivering.htm - Kladblok
Bestand Bewerken Opmaak Beeld Help
<!DOCTYPE html
PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html;
charset=utf-8" />
<meta name="datum_posting" content="1 mei 2005"/>
<meta name="datum_wijziging" content="5 mei 2005"/>
<meta name="auteur" content="Filip Boudrez"/>
<title>DAVID - Archiveren van websites</title>
</head>
<body></body>
</html>
```





#### (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. Zie *Bijlage 1* voor een overzicht van de afgekeurde tags en attributen.
-  schrijf structurele mark-up, geen presentatie mark-up
-  schrijf grammaticaal correcte mark-up:
  - respecteer de volgorde van de (X)HTML-elementen
  - plaats de elementen op hun juiste plaats: plaats de (afsluitende) tags in de correcte volgorde
  - sluit de elementen altijd af
-  valideer de webpagina's alvorens ze on line worden gepubliceerd





#### 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.


#### URL's

-  maak vriendelijke, menselijk begrijpbare URL's
-  zorg ervoor dat dynamisch samengestelde URL's nog steeds naar dezelfde informatie verwijzen, ook al is de informatie gewijzigd
-  vermeld geen sessies of querystrings in de URL's
-  vermijd het gebruik van spaties in map- en bestandsnamen


#### Hyperlinks:


-  interne links: gebruik document-relatieve pathaanduidingen
-  externe links: gebruik absolute pathaanduidingen.
-  documenteer de link door duidelijk zijn doel aan te geven.
-  bed geen javascriptfuncties in hyperlinks in (bijv. a href="javascript:alert('...')")


href="javascript:")


 bed geen hyperlinks in Flash ActionScript in

## Logbestanden

 laat het beheer van logbestanden met archiefwaarde niet aan derden over (bijv. een externe provider)





 identificeer en documenteer de interactie die je wil registreren

 leg duidelijk vast welke gegevens in de logbestanden worden bijgehouden. Voorkom dat de logbestanden na verloop van tijd worden vernietigd. Archiveer ze op een magnetische drager of een optische schijf wanneer de opslagcapaciteit van de harde schijven beperkt is.

 documenteer de tabellen in logbestanden

## Bij gebruik van

### Zorg er voor dat:

-  niet-tekstuele info: er een tekstueel alternatief is via de attributen "ALT" of "LONGDESC". Voorbeelden van niet tekstuele elementen zijn afbeeldingen, animaties, image maps, applets, grafische knoppen.
-  links: het doel van de link duidelijk is
-  stylesheets: de HTML-pagina's 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 verduidelijk 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

## 3.2 Richtlijnen voor een content management systeem

Veel websites worden aangestuurd door content management systemen. Met een content management systeem kunnen webpagina's heel snel aangepast worden, zonder dat hiervoor uitgebreide technische kennis nodig is. Idealiter is een content management systeem gekoppeld aan de records management applicatie van de organisatie. In de praktijk zijn er echter nog maar weinig organisaties die over een heuse applicatie voor records management beschikken. En als er al een records management applicatie is, dan zal het tot stand brengen van die koppeling in veel gevallen maatwerk met bijkomende hoge ontwikkelingskosten vergen. Voor de meeste organisaties zal de rechtstreekse archivering vanuit het content management systeem een meer realistische en aangewezen oplossing zijn. Hiertoe formuleert men best een aantal vereisten en functionaliteiten om de archivering van de content-items met archiefwaarde en hun metadata mogelijk te maken.

Een content management systeem bestaat uit verschillende componenten. De belangrijkste componenten zijn een beheer- en presentatiemodule. Voor het beheren van de content-items wordt meestal gebruik gemaakt van een databank. Deze content-items worden via een presentatiemodule op het web gepubliceerd. Hiervoor worden onder meer templates gebruikt. Deze templates halen hun inhoud op uit de gekoppelde databank en voegen die vervolgens in welbepaalde placeholders

in. Een aantal vereisten zijn dan ook van toepassing op de databank en de templates zodat archivering efficiënt kan verlopen:

- mbt de databank:
  - de gegevens worden beheerd door een open databank management systeem
  - de gegevens zijn bevroegbaar via SQL-statements en toegankelijk voor andere applicaties via ODBC- of JDBC-koppelingen
  - heeft een gedocumenteerd, overzichtelijk en uitbreidbaar datamodel
- mbt de templates voor webpagina's: op basis van de templates worden webpagina's samengesteld die voldoen aan de vereisten vermeld onder 1. vereisten voor webpagina's.

Daarnaast zijn er vanuit archiveringsstandpunt ook een aantal gewenste functionaliteiten voor het content management systeem zelf:

- inzake metadata: metadatavelden kunnen door de organisatie vrij worden gedefinieerd. Volgende metadata voor content-items zijn essentieel:
  - titel / naam
  - uniek webadres
  - versienummer
  - datum redactie van het content-item
  - datum on line beschikbaar (datum on line, datum off line)
  - publicatieplaats op de website
- inzake versiebeheer van de content-items en hun metadata:
  - het CMS kan de verschillende versies van content-items en hun metadata bijhouden
  - het CMS kan de mogelijkheid voorzien dat gepubliceerde content-items pas na versieverandering worden gewijzigd (bijv. check-in enkel mogelijk na versieverandering)
  - het CMS kan de verschillende versies van on line content op een statische wijze en als afzonderlijke objecten, en in combinatie met hun metadata, bewaren
- inzake archivering van de content-items en hun metadata:
  - het CMS kan nieuwe en gewijzigde webpagina's automatisch als een statische (X)HTML-pagina bewaren
  - het CMS laat toe dat de sitemaps en geselecteerde content samen met hun versieveranderingen automatisch kan worden gearchiveerd (bijv. via sitemaps voor de sitemaps of via statische (X)HTML-pagina's voor de webpagina's)
  - het CMS kan geselecteerde content-items en hun metadata in bulk exporteren
  - het CMS kan ondersteuning bieden om geselecteerde content-items en hun metadata bij export te migreren naar archiveringsformaten
- inzake logging: het CMS kan een gestructureerde audit-trail bijhouden waarin geselecteerde acties worden geregistreerd.

## 4. Archiveringsrichtlijnen

### Algemeen

maak duidelijke afspraken over het archiveren van de website:

- leg vast wie de website en de wijzigingen archiveert.
- wie meldt wijzigingen?
- spreek af wie de metadata van de website bijhoudt.
- zorg ervoor dat de archivering een vast onderdeel van de routine bij het aanbrengen van wijzigingen wordt. Laat het archiveringsmoment onmiddellijk aansluiten bij het updaten van de website
- vertrouw je niet op backuptapes bij het archiveren van websites. Digitaal archiveren is meer dan het maken van een backup.
- verwijder bestanden op de webserver alleen wanneer je zeker bent dat ze al zijn gearchiveerd.
- verwijder de on line versie pas na een succesvolle archivering.
- leg op voorhand de archiefwaarde van de website en de gerelateerde archiefdocumenten vast

### Website-archief

houd op de gedeelde schijfruimte van de administratieve, archief- of IT-dienst een off line website-archief bij:

- maak een aparte map met de naam 'Website\_archief' voor de gearchiveerde websites en webpagina's van de dienst(en).
- houd deze map duidelijk gescheiden van de werkmappen waarbinnen de webpagina's worden ontworpen.
- beperk de toegangsrechten tot deze map.
- maak voor elke website en elke versie nieuwe submappen. Geef deze map de naam 'Versie\_X\_Y' (Bijv. Versie\_2\_1).
- maak regelmatig een veiligheidskopie van de map 'Website\_archief'


### Documenteer de website

houd binnen elke versiemap een XML-bestand met de historiek van de website bij (metadata\_versie\_X\_Y.xml). Vul de metadata aan bij elke wijziging aan de website (niet de gekoppelde toepassingen zoals databanken). Met behulp van dit XML-bestand documenteer je de historiek van de website en kun je achteraf achterhalen wanneer welke webpagina on line was.


Op de DAVID-website is een sjabloon voor een dergelijk XML-document beschikbaar. Kopieer het bestand naar de versiemap en vul de gegevens aan. Je kan het bestand openen en invullen met een gewone teksteditor of tekstverwerker. Zorg ervoor dat je het bestand bewaart met de extensie \*.xml.

Vorbereidende documenten kunnen eventueel als aanvullende documentatie worden bijgehouden (o.a. structuur website, installatierichtlijn, systeembeschrijving, richtlijn voor administrator, enz.).

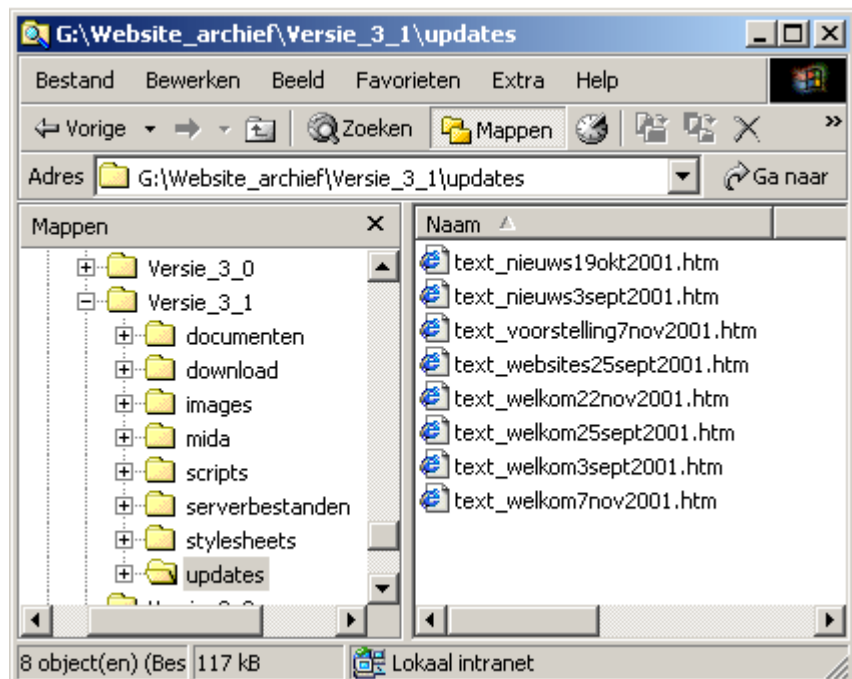
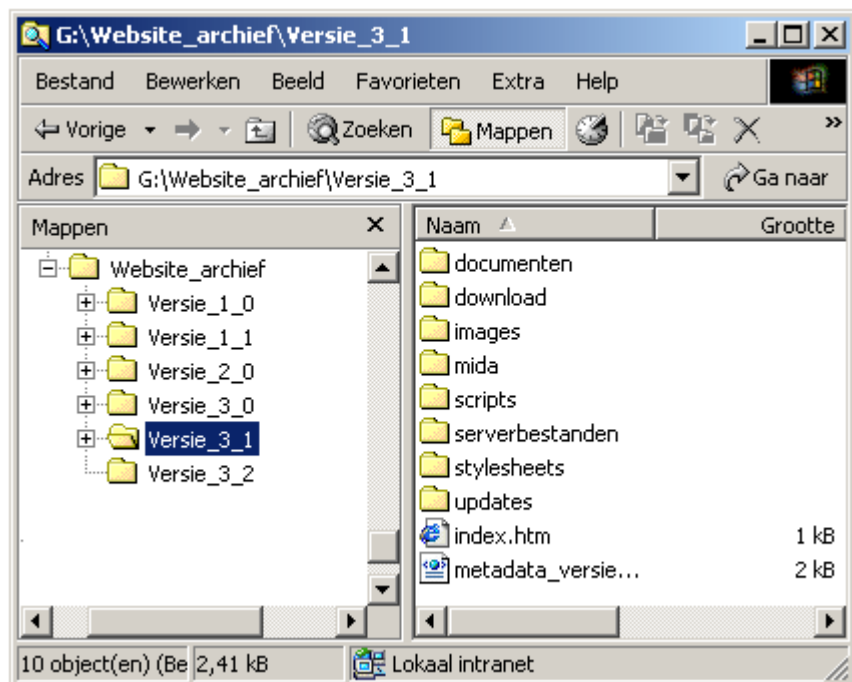
## Versies archiveren?

 de website:

- bij elke grote wijziging: archiveer een *volledige* versie van de website. Plaats de versie in een nieuwe submap van de map 'Website\_archief'.
- bij elke kleine wijziging: archiveer de *gewijzigde webpagina* en bijhorende bestanden (bijv. gewijzigde afbeeldingen). Plaats de webpagina in de submap 'updates' onder de versiemap. Voeg aan de bestandsnaam de datum van de versieverandering toe.

 de scripts: kopieer de serverscripts na elke wijziging naar de submap met als naam 'scripts'. Voeg aan elk bestandsnaam de datum van de posting toe of houd deze gegevens in een afzonderlijk bestand bij.

### Voorbeelden:



## Hoe archiveren?



momentopname:

- statische website (geen koppeling met databanken, Flash): kopieer de bestanden van de webserver naar de overeenstemmende versiemap (= archiveren van de broncode) of maak een snapshot wanneer de interne URL's dmv absolute pathaanduidingen worden aangegeven.
- dynamische of interactieve website (wel een koppeling met databanken): maak een snapshot van de website/pagina met een off line browser. Hierdoor wordt elke webpagina als een HTML-bestand opgeslagen, ook al zijn de originele bestanden op de webserver ASP-, PHP- of JSP-bestanden. Kopieer de ASP-, PHP- of JSP-bestanden van de webserver naar de submap 'serverbestanden'. Zo beschik je over een HTML-versie die off line kan geraadpleegd worden en bewaar je ook de originele serverbestanden met ingebede serverscripts (= archiveren van een snapshot).



surfsessie: registreer met een screencorder de website terwijl iemand de webpagina's bezoekt. Bewaar de surfsessie als een videobestand.

## Tips?



houd rekening met de maximale bestandsnaamlengte van 31 karakters. Stel de off line browser in zodat het aantal karakters in de bestandsnaam wordt beperkt (8+3).



stel de off line browser in zodat hij enkel de nodige bestanden off line plaatst. Geef bijvoorbeeld op dat PDF-documenten niet worden bewaard wanneer deze in het documentbeheerssysteem aanwezig blijven.



neem de mappenstructuur van de webserver over.



ga na of de gearchiveerde website alle nodige bestanden bevat. Controleer vooral de aanwezigheid van:

- de tweede laag van de roll-overafbeeldingen
- DTD's
- stylesheets

Kopieer indien nodig de ontbrekende bestanden van de webserver naar de map met het website-archief. Verwijder de bestanden pas op de webserver wanneer je zeker bent dat het website-archief alle bestanden bevat.



controleer de mapnamen: bij het nemen van een snapshot met een off line browser kan "%20" de spaties vervangen waardoor de links niet meer werken.



controleer en verbeter indien nodig de interne links alvorens de website in het digitaal archief op te nemen. Vergeet de links in de stylesheets niet aan te passen.



schakel de bezoekersteller, automatische datumweergave, e-maillinks, enz. uit: definieer de scripts of HTML-code als (X)HTML-commentaar.



websites in Flash: verwijder de paswoorden van beveiligde \*.swf-bestanden.

## Databanken? Documentbeheers- systemen?

Dynamische websites zijn gekoppeld aan databanken en/of documentbeheerssystemen. De gebruiker vraagt informatie uit databanken op of voegt er informatie aan toe. Bij het archiveren van de websites zelf, is het niet nodig om deze databanken en/of de inhoud van de documentbeheerssystemen iedere keer mee te archiveren. Voorwaarde is wel dat deze informatie in de databank of het documentbeheerssysteem aanwezig blijft en wordt gearchiveerd. Archiveer de databanken of documenten met archiefwaarden als afzonderlijke archiefdocumenten en leg in de metadata de band met de website vast.

## E-mails

Websites met formulieren kunnen aanleiding geven tot het vormen van e-mails. Voor het archiveren van e-mails: DAVID-rapport over e-mailarchivering (<http://www.antwerpen.be/david>)

## Downloads op webserver

Downloads die niet via een databank of documentbeheerssysteem ter beschikking worden gesteld en die niet op een andere plaats worden bewaard, archiveer je best bij de mirrors of snapshots van de websites. Kopieer ze rechtstreeks van de webserver of met behulp van een off line browser.

## Logbestanden

de logbestanden van de webserver kunnen archiefwaarde hebben:

- archiveer de logbestanden als databanken of als platte tekstbestanden
- houd documentatie over de gegevensvelden bij
- leg duidelijk de bewaartermijn van de logbestanden vast

## Wanneer archiveren?



plaats de te archiveren bestanden onmiddellijk na elke wijziging in het website-archief.



leg een procedure en tijdstip voor neerlegging bij de archiefdienst vast.

procedure: kopiëren via het netwerk of CD

tijdstip: na elke grote versieverandering

## 5. ? Vragen? Suggesties ?



Met al uw vragen en suggesties kan u terecht bij eDAVID: <mailto:info@edavid.be>



## Bijlage 1: Afgekeurde tags en attributen

De specificatie van HTML 4.01 is beschikbaar op de website van het *World Wide Web Consortium*: <http://www.w3.org/TR/html401/>. De XHTML 1.0-specificatie kan u downloaden via <http://www.w3.org/TR/xhtml1/>.

AFGEKEURDE TAGS	GEBRUIK IN DE PLAATS
<APPLET>	<OBJECT>
<BASEFONT>	stylesheet
<BLINK>	/
<BLOCKQUOTE>	stylesheet
<CENTER>	stylesheet
<DIR>	<UL>
<EM>	/
<FONT>	stylesheet
<ISINDEX>	<INPUT>
<MARQUEE>	/
<MENU>	<UL>
<S> / <STRIKE>	stylesheet
<STRONG>	stylesheet
<U>	stylesheet
<ALIGN>	stylesheet

AFGEKEURDE ATTRIBUTEN	HOORT BIJ TAG:
"align"	<CAPTION>, <APPLET>, <IFRAME>, IMG>, <INPUT>, <OBJECT>, <HR>, <LEGEND>, <TABLE>, <H1>-<H7>, <P>
"alink"	<BODY>
"background"	<BODY>
"bgcolor"	<TABLE>, <TR>, <TD>, <TH>, <BODY>
"border"	<IMG>, <OBJECT>
"clear"	 
"compact"	<DL>, <OL>, <UL>
"height"	<TD>, <TH>
"hspace"	<IMG>, <OBJECT>
"language"	<CDATA>
"link"	<BODY>
"noshade"	<HR>
"nowrap"	<TD>, <TH>
"size"	<HR>
"start"	<OL>
"text"	<BODY>
"type"	<LI>, <OL>, <UL>
"value"	<LI>
"version"	<HTML>
"vlink"	<BODY>
"vspace"	<IMG>, <OBJECT>
"width"	<HR>, <TD>, <TH>, <PRE>