



1. Preface

There are several arguments why websites must be archived.

- the content of intra/internet sites of public administrations can determine the acts and decisions of citizens and other services. The department can be held accountable for this.
- the government is accountable for their websites.
- the intra/internetsites are policy documents.
- more and more information is only available on intra/internetsites, and cannot be found in any other form.
- the information on these sites is volatile and subject to constant changes.
- other records are created through these sites.

Websites can only be archived effectively when they're efficiently designed and managed. Websites who meet these conditions, offers two advantages:

- these websites are accessible to as many people as possible. A number of the guidelines concerning design and management are based on the basic rules of the *Web Content Accessibility*. You will not reach your full potential of visitors when you sinn against these rules.
- the website version on the webserver can be used -more or less immediatelly- for record-keeping. A minimal amount of changes have to be made for the website to be ready for record-keeping.

This guideline consists of two parts. The first part is mainly intended for websitebuilders. The websitebuilders have to consider the possibility of record-keeping as soon as the creation of the website starts. They have to provide the facilities for the persons responsible for the content of the website, to document the changes of their website.

The second part contains guidelines for the creation and management of an off-line website archive. This website archive can be kept locally (at the departments themselves) or centrally (at the archival services or IT-staff.) When writing these guidelines, we assume that a minimal IT-infrastructure exists. Only an off-line browser is required. This guideline is applicable for websites with a static as well as with a dynamic content. They prescribe at the same time how each version or change can be saved for record-keeping. This guideline is a concrete application of the overwhole strategy to preserve websites as described in the DAVID-report, nr.6.

2. Guidelines to designing and managing websites.

These guidelines are based on the *Web Content Accessibility*-rules formulated by the World Wide Web Consortium (1999, <http://www.W3C.org/WAI/>), the quality demands for the recording keeping of websites, and the preliminary experiences of the Antwerp city archive concerning the record-keeping of websites.

In general:

-  make clear-cut agreements concerning the management of the website:
 - who is responsible for the design and the content?
 - who modifies the content of the website? Who takes care of the updates?
 - how do you make sure that the information is up-to-date?
 - what happens to outdated information?
 - who keeps track of documentation and manages the website?

 Bear in mind the possible record-keeping value when designing a website. Do not just consider the website itself but also related e-mails, logfiles, databases or document management systems. Assign in advance the archival value of the related documents and prevent them from being destroyed or over-written.

Long term planning:

 You have to consider the long-term perspective when creating a website. Create a procedure that is explicit and controllable in order to keep the website up-to-date on the one hand and that will keep older information available on the other hand. Keep this in mind when designing a folderstructure and creating links.

Folderstructure:

 Design on the webserver a transparent folderstructure. Follow the guidelines as described in "*Electronic record-keeping: guideline and advice nr. 3 concerning directory structure and filenames in electronic documents.*"

 Place all the files and folders that make up the website in one common rootfile.

- 📄 Think ahead and design immediately the further development of your website. Keep this in mind when developing a folderstructure so it will not have to change with every new version. Be consistent in the paths. (functionality of the links!).

Files:

- 📄 Use as many standardised fileformats as possible.(Electronicl record-keeping: *Guideline & Advice nr. 4 Standards for fileformats*) and apply the (X)HTML-rules meticulously.
- 📄 Give files and versions as unique a name as possible.

Images

- 📄 Consider the fileseize and the available band width.
- 📄 Always specify HEIGHT and WIDTH attributs of images, to enable webbrowsers to calculate the dimensions quickly.
- 📄 Provide alternative textual information through the ALT-attribute.
- 📄 Where possible, use interlaced GIF-images or progressive JPEG-images.

External webpages

Avoid or limit as much as possible the use of webpages or images of other websites. You cannot control their availability, content or the functionality of these webpages.

Links:

- 📄 Internal links: use document relative pathindications
- 📄 External links: use absolute pathindications.
- 📄 Document the link by clearly marking its goal.

Webpages:

- 📄 Use markup and stylesheets for their designed purpose. Keep structure and design clearly seperated. Use (X)HTML-tags for the markup of text and stylesheets for the lay-out.
- 📄 Construct surveyable webpages. Assign a clear structure to webpages. Design three parts in the webpage:

- ❑ first line: clarification of the used (X)HTML-version (and eventually specification of DTD):

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
```

or:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0
Strict//EN" "DTD/xhtml11-strict.dtd">
```

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

- 📄 Assign a title to each webpage:

<TITLE>Guideline webdesign</TITLE>



Make sure that every webpage contains the necessary metadata and contextual data:

- ❑ creation date:

```
<META NAME="date_creation" CONTENT="1 March 2002">
```

- ❑ changement date:

```
<META NAME="date_changement" CONTENT="1 April 2002">
```

- ❑ date of posting:

```
<META NAME="datum_posting" CONTENT="1 April 2002">
```

- ❑ name webdesigner:

```
<META NAME="webdesign" CONTENT="Telepolis">
```

- ❑ name author

```
<META NAME="author" CONTENT="John Johnson">
```

For instance:

```
index.htm - Kladblok
Bestand  Bewerken  Opmaak  Help
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<HTML>
<HEAD>
<TITLE>Richtlijn webdesign</TITLE>
<META NAME="datum_creatie" CONTENT="1 maart 2002">
<META NAME="datum_wijziging" CONTENT="1 april 2002">
<META NAME="datum_posting" CONTENT="1 april 2002">
<META NAME="webdesign" CONTENT="Telepolis">
<META NAME="auteur" CONTENT="Jan Janssens">
</HEAD>
<BODY></BODY>
</HTML>
```

(X)HTML-tags and attributes



Only use standardised (X)HTML-tags. Avoid the use of deprecated tags and attributes. The standards are: HTML 4.01 (ISO/IEC-15445:2000) and XHTML 1.0. View Appendix 1 for an overview of deprecated tags and attributes.



Respect the sequence of the (X)HTML-elements.



Place the elements at their proper positions: place the (closing) tags in the correct sequence.



Always close the elements

Logfiles



Do not leave the management of logfiles with archival value to a third party. (f.i an external provider)



Identify and document the interaction you want to register.



Assign clearly which elements are being documented in the logfiles. Avoid that logfiles are destroyed after an amount of time. File them

away on a magnetic carrier or an optical disk when the capacity of the harddisks is limited.

 Document the tabels in logfiles.

When using

 Non-textual info:

a textual alternative is available through the attributes "ALT" or "LONGDESC". Examples of non-textual elements are images, animations, image maps, applets, grafical buttons.

 Links:

The purpose of the link is obvious.

 Stylesheets:

The HTML-pages are also readable without stylesheets.

 Image maps

As many client image maps as possible are being used, avoid the usage of server image maps

 Tabels:

The headers of row and columns are clearly identified.

 Frames:

Each frame is being identified on the basis of a meaningfull name. Clarify the relationship between the frames when it's not possible to ascertain this from the name.

 Dynamical content:

An alternative is provided for in case of disfunctionment of the interaction.

 Scripts / applets:

The webpage still functions when the scripts/applets are turned off or aren't supported. If this is impossible, refer to a webpage with similar information.

3. Guidelines for record-keeping

In general

Make clear-cut agreements concerning the record-keeping of the website. :

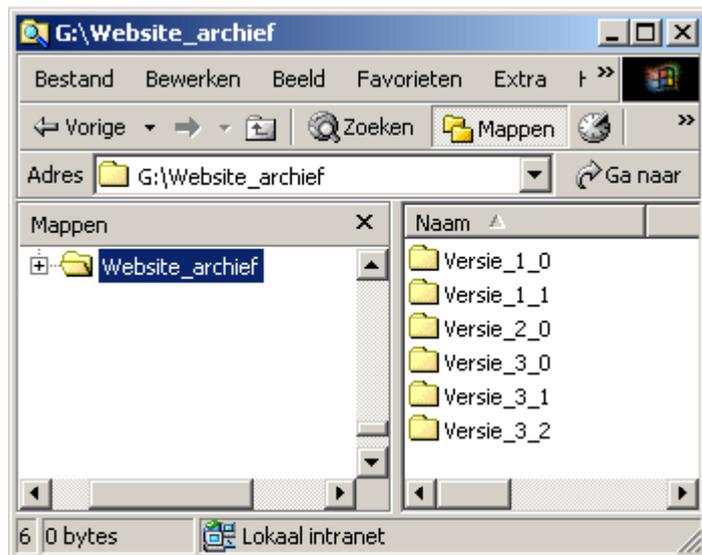
- assign who has to archive the website and the changes.
- who will signal the changes?
- arrange who will document the metadata of the website.
- make sure that the record-keeping becomes a part of the routine when making changes. The record-keeping moment should follow immediately after the website update.
- do not put your trust in back-up tapes when archiving websites. Electronic record-keeping implies more than the making of back-ups.
- delete files on the webserver only when you're sure that they're already archived.
- determine in advance the archival value of the website and the related records.

Website-archives

Keep an off-line website archive on the shared disk space of the administrative, archival- or IT-service:

- ❑ create a separate folder named 'Website_archive' for the archived websites and webpages of the service(s).
- ❑ keep this folder clearly separated from the workfolders in which the webpages are created.
- ❑ limit the access rights to this folder.
- ❑ create new subfolders for each website and for each new version. Name this folder: 'Version_X_Y' (F.I.. Version_2_1).
- ❑ create regular safteycopies of the folder 'Website_record'.

For instance:



Document the website

Keeping record of version?



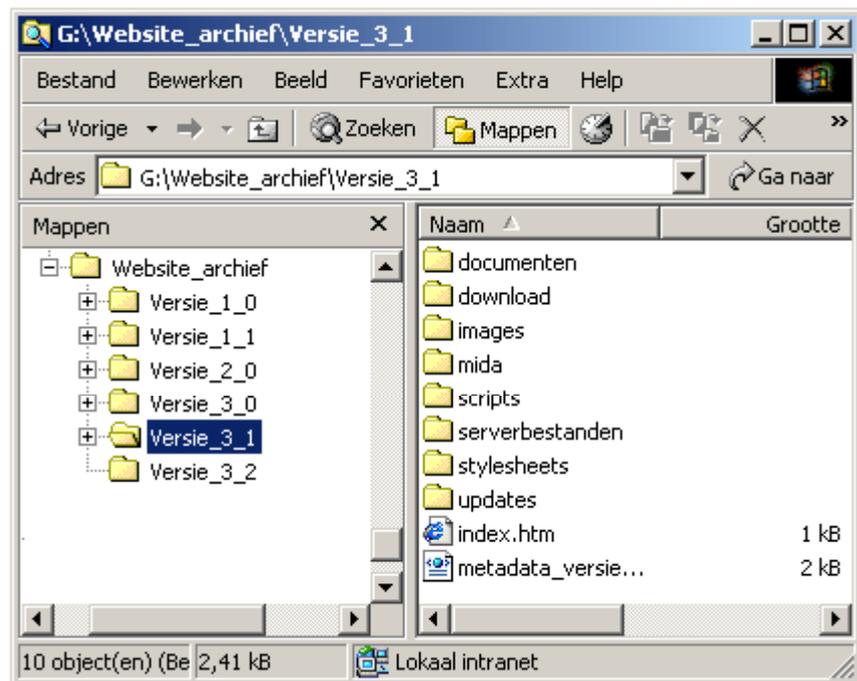
The website:

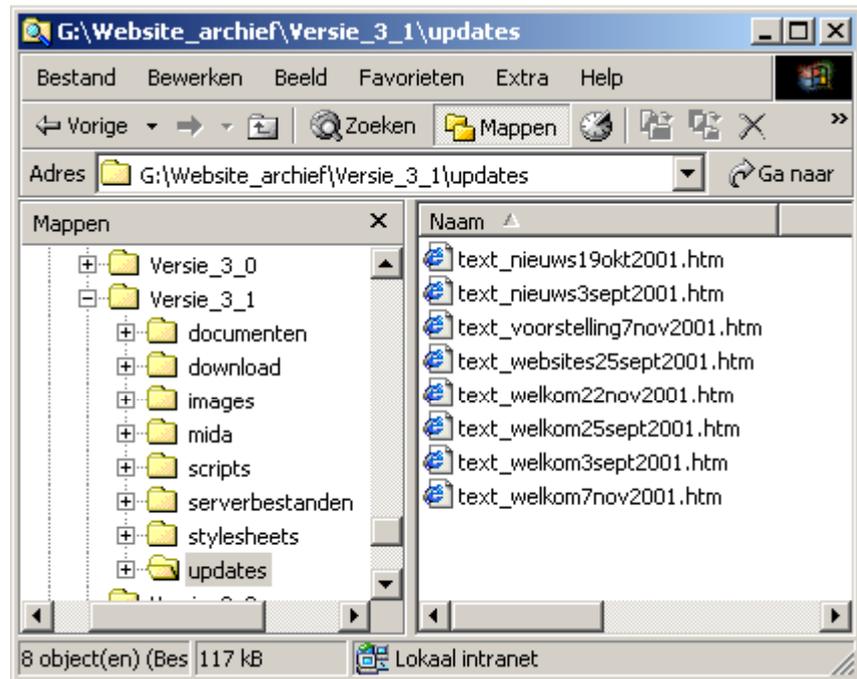
- at every major changement: file away a *full* version of the website. Place this version into a new subfolder of the folder "website_archive".
- at each minor changement: file away the *changed* webpage and linked files (f.i. changed images). Place the webpage into the subfolder "updates" in the versionfolders. Add the date of posting to the filename.



the scripts: copy the serverscripts after every change into the subfolder named "scripts". Add the date of posting to each filename or keep these records in a seperate file.

For instance:





Keeping records: how?

-  Website with fixed content (no link to databases, Flash): copy the files from the webserver into the corresponding versionfolder.
-  Dynamic website (with a link to databases): make a snapshot of each website/page with an off-line browser. It's for this reason that every webpage is filed away as an HTML-file, even though the original files on the webserver are ASP-, PHP- or JSP-files. Copy the ASP-, PHP- or JSP-files from the webserver to the subfolder "serverfolders". Now you dispose of an HTML-version which can be consulted off line and you can also preserve the original serverfiles with embedded serverscripts.

Pointers?

-  Bear in mind the maximum filename length of 31 characters. Adjust the off line browser in such a way that the number of characters is limited.
-  Adjust the off line browser in such a way that only the necessary files are placed off line. For instance, indicate that PDF-documents aren't saved when they remain present in the document management system.
-  Use the same structure as used in the directory structure.
-  Check whether the archived website contains all of the necessary files. Check especially the presence of
 - ▶ the second layer of the roll- over images
 - ▶ DTD's
 - ▶ stylesheets

If necessary, copy the missing files from the webserver to the folder with the website-archives. Delete the files from the webserver only when you're sure that the website-archives contains all the files.

-  Check the filenames: when taking a snapshot with an off line browser "%20" can replace the interspaces so that the links don't work anymore.
-  Check and rectify when necessary the links before admitting the website into the electronic archives. Do not forget to alter the links in the stylesheets.
-  Switch the visitor counter and the automatic date off: define the scripts as (X)HTML-commentary.
-  Websites in Flash: record the passwords of secured *.swf-bestanden.

Databases? Document management systems?

Dynamic websites are linked to databases and/or document management systems. The user calls up information from databases or adds information to them. It's not necessary to keep record of these databases and/or the content of document management systems each time when keeping record of the websites as interface themselves. This is only on condition that the information stays available in the databases or the document management systems and is being archived itself. Archive the databases or documents with archival value as separate records. Record the link with the website in the metadata.

E-mails

Websites with forms can give cause to the creation of e-mails. For archiving e-mails: DAVID-report on e-mailarchiving (<http://www.antwerpen.be/david>)

Downloads on the webserver

Downloads which aren't made available through a database or a document management system and which aren't saved at another place, are best archived with mirrors or snapshots of the website. Copy them directly from the webserver or with the aid of an off line browser.

Logfiles

The logfiles of the webserver can have archival value:

- archive the logfiles as databases or as flat textfiles
- keep documentation about the data fields.
- mark clearly the storage time of the logfiles.

Archiving: when?

-  Place the files that have to be archived immediatly after every change in the website archive.



Determine the procedure and the moment for deposition at the archival service.

Procedure: copy through the network or CD

Moment: after every major version change.

4. ?Questions? Suggestions ?



Turn to DAVID for all your questions and suggestions:

e-mail: david@stad.antwerpen.be

telephonenumber: 00 32 (0)3 206 9415



Appendix 1: Deprecated tags and attributs

The HTML 4.01 specification is available on the website of *World Wide Web Consortium*:
<http://www.w3.org/TR/html401/>

The XHTML 1.0 specification can be downloaded through <http://www.w3.org/TR/xhtml1/>

DEPRECATED TAGS	INSTEAD USE
<APPLET>	<OBJECT>
<BASEFONT>	stylesheet
<BLINK>	/
<BLOCKQUOTE>	stylesheet
<CENTER>	stylesheet
<DIR>	
	/
	stylesheet
<ISINDEX>	<INPUT>
<MARQUEE>	/
<MENU>	
<S> / <STRIKE>	stylesheet
	stylesheet
<U>	stylesheet
<ALIGN>	stylesheet

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