

# Eerste Hulp bij Website Archivering

Bibliografie websitearchivering

Erika Hokke, februari 2006



Keizersgracht 630  
1017 ER Amsterdam  
T (020) 524 1166  
F (020) 595 1160  
info@archiefschool.nl  
www.archiefschool.nl

# INHOUDSOPGAVE

<b>1</b>	<b>Inleiding</b>	<b>3</b>
<b>2</b>	<b>Het archiveren van (een) website(s) van andere personen of organisaties</b>	<b>4</b>
2.1	Inleiding	4
2.2	Beleid voor het archiveren van website(s)	4
2.3	Ontwerpen en implementeren van een website archief	4
2.4	Snapshots: het archiveren van een kopie van de site	5
2.5	Event: het archiveren van de interactie	6
2.6	Acquisitie van websites	6
2.7	Beschikbaar stellen	6
2.8	Juridische aspecten	7
<b>3</b>	<b>Het archiveren van een eigen website</b>	<b>8</b>
3.1	Inleiding	8
3.2	Ontwerpen van een archiveerbare website	8
3.3	Beleid voor archivering van eigen website(s)	9
3.4	Implementatie van archivering binnen de eigen organisatie	10
3.5	Waardering en selectie van websites	10
3.6	Beschikbaar stellen	10
3.7	Risico's en bewaarbelangen, juridische aspecten	10
<b>4</b>	<b>Praktijkervaringen en informatie</b>	<b>11</b>
4.1	Inleiding	11
4.2	Projecten	11
4.2.1	Pilot projecten	11
4.2.2	Bibliotheek	12
4.2.3	Archieven	17
4.2.4	Overige initiatieven	20
4.3	Congressen	22
4.4	Portalsites, bibliografieën, discussielijst.	23

# 1 INLEIDING

Deze bibliografie bevat verwijzingen naar publicaties over het archiveren van websites. Deze literatuurstudie is uitgevoerd tijdens de twee onderzoeken die de Archiefschool de afgelopen jaren heeft uitgevoerd naar de archivering van websites. In 2002-2003 is in het project *Blijvend Beschikbaar* in opdracht van het ministerie van Verkeer en Waterstaat onderzoek gedaan naar de vraag of websites archief zijn. In 2005 is in opdracht van de ICTU het project *Regels voor websites* uitgevoerd. In dit project zijn vanuit archiveringsperspectief eisen geformuleerd voor het ontwerpen van archiveerbare websites. Deze eisen vullen de eerder door de ICTU geformuleerde webrichtlijnen aan.

Deze bibliografie heeft als basis:

1. het overzicht van projecten en publicaties uit H.A. Hokke *Blijvend beschikbaar* (Amsterdam, 2003) via: [http://www.elo.nl/elo/Images/blijbeschik\\_tcm70-20514.doc](http://www.elo.nl/elo/Images/blijbeschik_tcm70-20514.doc) 06-02-2006
2. Andreas Asschenbrenner *Web Archiving Bibliography* (april 2004) via <http://www.ifs.tuwien.ac.at/~aola/links/WebArchiving.html> 06-02-2006

Beide documenten zijn samengevoegd, geordend en aangevuld met recente, voornamelijk digitale, publicaties. De digitale publicaties waren op 6 februari 2006 via internet te vinden. Helaas zijn veel links naar publicaties kwetsbaar, het kan dus zijn dat een publicatie niet direct te vinden is via de gegeven URL. Soms is even zoeken noodzakelijk.

De bibliografie is bedoeld voor studenten Archivistiek, archivariissen en DIV-adviseurs en verder iedereen die zich bezig wil houden met het archiveren van websites. De achterliggende gedachte is dat organisaties / personen óf wel websites van anderen willen archiveren (zoals archiefdiensten, bibliotheken en andere collectiebeherende instellingen dat nu doen), danwel dat organisaties / personen hun eigen website willen archiveren (zoals overheidsorganisaties), danwel dat men meer algemene informatie wil over websitearchivering.

Dit heeft geleid tot de volgende rubrieken:

- Het archiveren van websites van andere personen of organisaties ( zie paragraaf 2)
- Het archiveren van een eigen website ( zie paragraaf 3)
- Praktijkervaringen en informatie ( zie paragraaf 4)

Binnen de rubrieken zijn de publicaties alfabetisch op achternaam van de auteur geordend.

De bibliografie is zeker niet volledig. Aan een volgende versie wordt een begrippenlijst en meer inhoudelijke informatie toegevoegd.

Aanvullingen zijn van harte welkom, via: [h.a.hokke@archiefschool.nl](mailto:h.a.hokke@archiefschool.nl)

## 2 HET ARCHIVEREN VAN (EEN) WEBSITE(S) VAN ANDERE PERSONEN OF ORGANISATIES

### 2.1 Inleiding

Bibliotheken, archiefdiensten en andere collectiebeherende instellingen, zoals het Documentatiecentrum voor Nederlandse Politieke Partijen (DNPP), houden zich bezig met het archiveren van websites.

Dit hoofdstuk geeft een overzicht van publicaties over het archiveren van “andermans” websites. Het zijn publicaties over beleidsmatige aspecten (§ 2.2); over het ontwerpen en implementeren van een website archief (§ 2.3), waaronder een publicatie over kosten van archiveren van websites. Verder bevat dit hoofdstuk publicaties over archiveringsstrategieën (§ 2.4 en 2.5); over de acquisitie van websites (§ 2.6); over het opnieuw beschikbaar stellen van gearchiveerde websites (§ 2.7). Een aantal publicaties over juridische aspecten (§ 2.8) sluit dit hoofdstuk af.

### 2.2 Beleid voor het archiveren van website(s)

Aschenbrenner, Andreas; Rauber, Andreas *Die Bewahrung unserer Online-Kultur, Vorschläge zu Strategien der Webarchivierung* (2003) via: <http://www.onb.ac.at/sichtungen/beitraege/aschenbrenner-a-1a.html> 06-02-2006

Boudrez F., Eynde S. van den *DAVID, archivering van websites* (Antwerpen - Leuven, mei 2002) via <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006

Brügger, Niels *Archiving websites. General Considerations and strategies* (Aarhus, The Centre for Internet Research, 2005) via: <http://cfi.imv.au.dk/eng/pub/webarc> 06-02-2006

Commission on Preservation and Access and the Research Library Group *Preserving Digital Information* (1996) via: <http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf> 06-02-2006

Masanès, Julien “Towards Continuous Web Archiving. First Results and an Agenda for the Future” in *D-Lib magazine* (2002) via: <http://www.dlib.org/dlib/december02/masanes/12masanes.html> 06-02-2006

National Library of Australia *Archiving Web Resources. Issues for Cultural Heritage organisations, Conference Report* (2004) via: <http://www.nla.gov.au/webarchiving/ConferenceReport.rtf> 06-02-2006

Teske, Frank *Archivierung des Internets, eine Aufgabe für Archive?* (Hauptstaatsarchiv Stuttgart und Archivschule Marburg, 2003) via: [http://www.landesarchiv-bw.de/sixcms/media.php/25/transf\\_teske\\_internet.pdf](http://www.landesarchiv-bw.de/sixcms/media.php/25/transf_teske_internet.pdf) 06-02-2006

### 2.3 Ontwerpen en implementeren van een website archief

Arms, W.Y. *Web preservation Project, final report* (2001) via <http://lcweb2.loc.gov/cocoon/minerva/html/minerva-home.html> 06-02-2006

Boudrez, F. *Archivering van websites: een kwestie van waardering en 'capture'* (Antwerpen, 2005) via <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006

Boudrez, F. *Van back-up tot gearchiveerde website* (Antwerpen, 2002). Via <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006  
Christensen-Dalsgaard e.a. *Experiences and conclusions from a pilot study: Web archiving of the District and County elections 2001. Final Report for The Pilot Project "netarkivet.dk"* (February 2003). <http://netarchive.dk/rap/index-en.php> 06-02-2006

Oss, Jeroen van; Bermont, Ivar; Wijngaarden, Peter van; Veldhuizen, Guus *Handboek archivering websites. Concept 1.1* (Gemeentearchief Rotterdam, 2005)

Schmitz, Rudolf "Aufbau und Struktur eines Internet-Archivs. Die Archivierung von Internet-Auftritten der SPD und ihrer Fraktionen in den Parlamenten" in *Der Archivar* (2004-4) p. 318-320 en via: [http://www.archive.nrw.de/archivar/2004-04/Archivar\\_2004-4.pdf](http://www.archive.nrw.de/archivar/2004-04/Archivar_2004-4.pdf) 06-02-2006

## architectuur

Masanès, Julien *Web Archive's Architectural design* (IWA5, 2005) via: <http://www.iwaw.net/05/masanes1.pdf> 06-02-2006

Julien Masanès: Towards Continuous Web Archiving: First Results and an Agenda for the Future. *D-Lib Magazine*, v.9, n.12 (December 2002) via: <http://www.dlib.org/dlib/december02/masanes/12masanes.html> 06-02-2006

Milinic, Miroslav; Topolscak, Nebojsa "The architecture of DAMP: a System for Harvesting and Archiving Web Publications" in: *Widwisawn vol 3.3* via: [http://widwisawn.cdlr.strath.ac.uk/Issues/Vol3/issue3\\_3\\_1.html](http://widwisawn.cdlr.strath.ac.uk/Issues/Vol3/issue3_3_1.html) 06-02-2006

## kosten

Philips, Margaret E. *Selective Archiving of Web Resources: A Study of Acquisition Costs at the National Library of Australia* (2005) via: [http://www.rlg.org/en/page.php?Page\\_ID=20666#article0](http://www.rlg.org/en/page.php?Page_ID=20666#article0) 06-02-2006

## 2.4 Snapshots: het archiveren van een kopie van de site

Boyko, Andrew *Test Bed Taxonomy for Crawler* (International Internet Preservation Consortium – Metrics and Testbed Working Group, 2004) via: <http://www.netpreserve.org/publications/iipc-r-002.pdf> 06-02-2006

Koerbin, Paul *Report on the Crawl and Harvest of the Whole Australian Web Domain, Undertaken during June and July 2005* (10 oktober 2005) via: [http://pandora.nla.gov.au/documents/domain\\_harvest\\_report\\_public.pdf](http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf) 06-02-2006

Marill, Jennifer; Boyko, Andrew; Ashenfelder, Michael *Web Harvesting Survey* (International Internet Preservation Consortium – Metrics and Testbed Working Group, 2004) via: <http://www.netpreserve.org/publications/iipc-r-001.pdf> 06-02-2006

Sigurdsson, Kristinn *Incremental crawling with Heritrix* (2005) via: <http://www.iwaw.net/05/papers/iwaw05-sigurdsson.pdf> 06-02-2006 zie ook <http://crawler.archive.org/> 06-02-2006

Thomassen, Bo Hovgaard *Tests of software and strategies for micro-archiving websites* (Department of Information and Media Studies, University of Aarhus, 2004) via: [http://cfi.imv.au.dk/eng/pub/webarc/test\\_premises\\_thomassen.pdf](http://cfi.imv.au.dk/eng/pub/webarc/test_premises_thomassen.pdf) 06-02-2006

Verhoeven, Hans *Archiving web publications* (IBM/Koninklijke bibliotheek, Den Haag, 2002) via: [http://www.kb.nl/hrd/dd/dd Onderzoek/summary\\_ltpstudy6.html](http://www.kb.nl/hrd/dd/dd Onderzoek/summary_ltpstudy6.html) 06-02-2006

## 2.5 Event: het archiveren van de interactie

Fitch, Ken *Web site archiving – an approach to recording every materially different response produced by a website* via: <http://ausweb.scu.edu.au/aw03/papers/fitch> (AUSWEB03, 2003) 06-02-2006 of via: <http://www.projectcomputing.com/products/pageVault/index.html> 06-02-2006

Foote, Adrian *White Paper Vignette Webcapture* (2005) via: [http://a2032.g.akamai.net/7/2032/812/71d143db78caa7/www.vignette.com/Downloads/WP\\_WebCapture.pdf](http://a2032.g.akamai.net/7/2032/812/71d143db78caa7/www.vignette.com/Downloads/WP_WebCapture.pdf) 06-02-2006 of via: <http://www.vignette.com/contentmanagement/0,2097,1-1-1928-4149-1968-4535,00.html> 06-02-2006

Marsili, Diana “Web Content meets Records Management” in *AIIM E-DOC magazine* (2002) via: [http://www.edocmagazine.com/vault\\_articles.asp?ID=25004](http://www.edocmagazine.com/vault_articles.asp?ID=25004) 06-02-2006

National Archives Australia *Archiving Web Resources: a policy for Keeping Records of Web-bases Activity in the Commonwealth Government* (NAA, 2001) via: [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) 06-02-2006

National Archives Australia *Archiving Web Resources: Guidelines for Keeping Records of Web-bases Activity in the Commonwealth Government* (NAA, 2001) via: [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) 06-02-2006

## 2.6 Acquisitie van websites

Boudrez, F. *Archivering van websites: een kwestie van waardering en ‘capture’* (2005) via: <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006

National Library of Australia: *Online Australian Publications: Selection Guidelines for Archiving and Preservation*. (2005) via: <http://pandora.nla.gov.au/selectionguidelines.html> 06-02-2006

## 2.7 Beschikbaar stellen

Via The Wayback Machine, WERA en NutchWax kan gezocht worden in collecties websites. Met Xinq kan gezocht worden in een XML database.

*NutchWax* (Nutch with Web Archive eXtensions) via: <http://archive-access.sourceforge.net/projects/nutch/index.html> 06-02-2006

*Wayback Machine* via: <http://www.archive.org/web/web.php> 06-02-2006

*Web aRchive Access* (WERA) (voorheen NWA toolset) via: <http://archive-access.sourceforge.net/projects/wera/index.html> 06-02-2006

*Xinq* via: <http://www.nla.gov.au/xinq> 06-02-2006 en <http://www.nla.gov.au/xinq/presentations/ausweb-paper.html> 06-02-2006

## 2.8 Juridische aspecten

Charlesworth, Andres *Legal issues relating to the archiving of Internet resources in the UK, EU, US and Australia* (version 1.0, february 2003) via: [http://library.wellcome.ac.uk/doc\\_WTL039231.html](http://library.wellcome.ac.uk/doc_WTL039231.html)

Mossink, Wilma 'Juridische aspecten van het archiveren van websites' in: Frank den Hollander en Gerrit Voerman *Het web gevangen. Het archiveren van websites van politieke partijen* (Groningen, Archipol, z.j.) p. 35 – 44

Philips, Margaret E.; Cunningham, Adrian *Keeping online information accessible for E-government and E-democracy* (2005) via: <http://www.nla.gov.au/nla/staffpaper/2004/phillips1.html> 06-02-2006

## 3 HET ARCHIVEREN VAN EEN EIGEN WEBSITE

### 3.1 Inleiding

Het archiveren van een eigen website begint met het ontwerpen van goede websites. De ICTU heeft richtlijnen ontworpen waar goede websites aan dienen te voldoen. Wereldwijd hebben diverse organisaties (een voorstel voor) beleid of richtlijnen voor het ontwerpen en beheren van websites ontwikkeld. Er zijn onder andere richtlijnen ontwikkeld door de nationale archieven van de VS, Australië, Groot-Brittannië en richtlijnen van organisaties die hun eigen website archiveren of willen archiveren. Deze documenten bevatten advies/voorschriften over het ontwikkelen van beleid en over bevoegdheden rondom websitearchivering.

Dit hoofdstuk verwijst naar publicaties over het ontwerpen (§ 3.2) en archiveren van “eigen” websites. Het betreft publicaties over archiveringsbeleid (§ 3.3); het implementeren van archivering van websites (§ 3.4); waardering en selectie van websites (§ 3.5) en risico's en juridische aspecten (§ 3.6). Voor publicaties over het opnieuw beschikbaar stellen van eigen websites, zie hoofdstuk 2.

### 3.2 Ontwerpen van een archiveerbare website

Boudrez, F. *Archivering van websites: een kwestie van waardering en 'capture'* (2005) via: <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006

Boudrez, F. *Websitesbeheer en contentmanagement voor digitale archivering* versie 2.0 (2005) via: <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006

Dack, Diana *Persistence is a virtue Paper presented to the Online Information Conference*, (2001) via: <http://www.nla.gov.au/nla/staffpaper/2001/dack.html> 06-02-2006

Hokke, H.A. *Naar archivering van websites. Implementatieadvies* (Amsterdam, Archiefschool, ministerie van Verkeer en Waterstaat, 2003) zie: [http://www.elo.nl/elo/Images/implarchwebs\\_tcm70-20515.doc](http://www.elo.nl/elo/Images/implarchwebs_tcm70-20515.doc) 06-02-2006

ICTU *Webrichtlijnen voor de overheid* (versie 1.1 maart 2005) via: <http://webrichtlijnen.overheid.nl>

Sorensen, Jan Dalsten *'Standards for metadata and documentation: practical experiences'*. – (Dansk Riksarchivet, z.j.) via: <http://www.sa.dk/sa/itogarkiv/teknologi/samarbejde/jdsmetadata.htm> 06-02-2006

Voorburg, René J.J. *Webontwerp: richtlijnen voor archivering* (Amsterdam, Uselab & ministerie van Verkeer en Waterstaat, 2004) via: [http://digidiv.amsterdam.nl/handleidingen/digitale\\_duurzaamheid/downloads/webont.pdf](http://digidiv.amsterdam.nl/handleidingen/digitale_duurzaamheid/downloads/webont.pdf) 06-02-2006

Voorburg, R. J.J.; Goutier, J.L.E. *Webarchivering bij het Ministerie van Verkeer en Waterstaat. Verslag van een onderzoek* (2005) via: [http://www.capsis.nl/files/verslag\\_onderzoek\\_webarchivering\\_minvenw\\_capsis\\_publicatie\\_20050822.pdf](http://www.capsis.nl/files/verslag_onderzoek_webarchivering_minvenw_capsis_publicatie_20050822.pdf) 06-02-2006

### 3.3 Beleid voor archivering van eigen website(s)

Barry, Rick "Web Sites as Recordkeeping and 'Recordmaking Systems'" in: *The Information Management Journal, the professional journal of ARMA* (Vol. 38, No. 6, Nov/Dec 2004) via: <http://www.mybestdocs.com> 06-02-2006

Boudrez, F.; Eynde, S. van den *Archivering van websites* (2002) via: <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006

Dollar, Ch. *Archival preservation of Smithsonian Web Resources: strategies, principles and best practices* (2001) via: <http://www.si.edu/archives/archives/dollar%20report.html> 06-02-2006

Gun, José van der *Gevangen van het Net. Onderzoek naar de archivering van intranetsites* (Afstudeerscriptie Hogeschool van Amsterdam, 2003) via: <http://www.archiefschool.nl/docs/intranetarchivering.pdf> 06-02-2006

Hokke, H.A. *Blijvend beschikbaar. Onderzoek naar de Archivering van websites* (Amsterdam, Archiefschool, ministerie van Verkeer en Waterstaat, 2003) via: [http://www.elo.nl/elo/Images/blijbeschik\\_tcm70-20514.doc](http://www.elo.nl/elo/Images/blijbeschik_tcm70-20514.doc) 06-02-2006

LeFurgy, William G. *Records and Archival Management of World Wide Web Sites* (2001) via: <http://www.rbarry.com/lefurgy-w-grn0104.htm> 06-02-2006

McClure, Charles R.; Timothy Sprehe, J. *Analysis and Development of Model Quality Guidelines for Electronic Records Management on State and Federal websites* (januari 1998) via: <http://www.ii.fsu.edu/~cmclure/#1998> 06-02-2006

McClure, Charles R.; Timothy Sprehe, *Guidelines for electronic records management on state and federal agency websites* (1998) via: <http://www.ii.fsu.edu/~cmclure/#1998> 06-02-2006

National Archives Australia *Archiving Web Resources: a policy for Keeping Records of Web-bases Activity in the Commonwealth Government* (2001) via: [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) 06-02-2006

National Archives Australia *Archiving Web Resources: Guidelines for Keeping Records of Web-bases Activity in the Commonwealth Government* (2001) via: [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) 06-02-2006

The National Archives UK, voorheen Public Records Office *Managing Web resources. Management of electronic records on websites and intranets: an ERM toolkit* (version 1.0, Public Records Office London, 2001) via: <http://www.nationalarchives.gov.uk/electronicrecords/advice> 06-02-2006

NARA *guidance on managing web records* (January 2005) via: <http://www.archives.gov/records-mgmt/policy/managing-web-records-index.html> 06-02-2006 of: <http://www.archives.gov/records-mgmt/pdf/managing-web-records-index.pdf> 06-02-2006

Ullman, Angela *Thesen zur archivering von Netzressourcen* (Parlamentsarchiv des Deutschen Bundestages, 2005) via: <http://www.bundestag.de/bic/archiv/oeffent/ThesenArchivierungNetzressourcen.pdf> 06-02-2006

Wolters, Diana *Het geheime web. Onderzoek naar webarchivering bij de Militaire Inlichtingen- en Veiligheidsdienst* (afstudeerscriptie Opleiding IDM, Saxion Hogeschool IJsselland, 2004)

### 3.4 Implementatie van archivering binnen de eigen organisatie

Boudrez, F. *Websitesbeheer en contentmanagement voor digitale archivering* versie 2.0 (2005) via: <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006

Dollar, Charles *Archiving Smithsonian Websites: an evaluation and recommendation for a Smithsonian Institution Archives Pilot Project* (2003) via: <http://siarchives.si.edu/pdf/websitepilot.pdf> 06-02-2006

Hokke, H.A. *Naar archivering van websites. Implementatieadvies* (Amsterdam, Archiefschool, ministerie van Verkeer en Waterstaat, 2003) via: [http://www.elo.nl/elo/Images/implarchwebs\\_tcm70-20515.doc](http://www.elo.nl/elo/Images/implarchwebs_tcm70-20515.doc) 06-02-2006

### 3.5 Waardering en selectie van websites

*Advies Raad voor Cultuur over Ontwerp-selectielijst overheidsinformatievoorziening over de periode 1945-1999*, arc-2002.4316/2. (13 juni 2002) p. 6 via: <http://www.cultuur.nl/files/pdf/advies/arc-2002.4316.2.pdf> 06-02-2006

### 3.6 Beschikbaar stellen

Zie hoofdstuk 2 paragraaf beschikbaar stellen.

### 3.7 Risico's en bewaarbelangen, juridische aspecten

Kenney, Anne R.; McGovern, Nancy 'Preservation Risk Management for Web Resources' in: *The information Management Journal* (september/oktober 2002) p. 52-61 of via: <http://www.dlib.org/dlib/january02/kenney/01kenney.html> 06-02-2006

McClure, Charles R. *Recent Key Policies Affecting Federal Website Management and Evaluation* (2005) [http://www.ii.fsu.edu/~cmclure/federal\\_website\\_policies.pdf](http://www.ii.fsu.edu/~cmclure/federal_website_policies.pdf)

Meijer, Albert "Accountability in an Information age: Opportunities and Risks for Records Management" in *Archival Science* 1(2002) p. 361 – 372

NARA *Guidance on managing Web Resources, Risk and risk assessment* (2005) via: <http://www.archives.gov/records-mgmt/policy/managing-web-records.html> 06-02-2006

National Archives Australia *Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government* (2001) via: [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) 06-02-2006

Philips, Margaret E. Cunningham, Adrian *Keeping online information accessible for E-government and E-democracy* (2005) via: <http://www.nla.gov.au/nla/staffpaper/2004/phillips1.html> 06-02-2006

## 4 PRAKTIJKERVARINGEN EN INFORMATIE

### 4.1 Inleiding

Dit hoofdstuk geeft een overzicht van verschillende initiatieven met betrekking tot de archivering van websites (niet uitputtend). Het bevat verwijzingen naar projecten waarin ervaring is opgedaan met websitearchivering (§ 4.2), naar congressen over websitearchivering (§ 4.3) en naar portalsites en andere bibliografieën (§ 4.4).

De projecten lopen uiteen van het opzetten van archieven met websites, het ontwikkelen van gebruikersinterface tot het geven van richtlijnen voor archivering. De projecten zijn gegroepeerd in projecten die opgezet zijn door bibliotheken (zie § 4.2.2), projecten opgezet door of betrekking hebbend op archieven (inclusief recordsmanagement) (zie § 4.2.3) en projecten opgezet door andere instellingen (kenniscentra, musea, bedrijven etc.) (zie § 4.2.4).

Per project wordt achtereenvolgens genoemd:

- Naam het project (indien aanwezig) en/of naam van de organisatie
- Korte omschrijving van het project of onderzoek
- Links naar websites en literatuur over het project

Paragraaf 4.2.1. benoemt een aantal vroege projecten, de zgn. pilot-projecten.

### 4.2 Projecten

#### 4.2.1 Pilot projecten

Wereldwijd zijn sinds 1996 initiatieven ontplooid om websites te archiveren. Een aantal projecten is te beschouwen als pionierprojecten, zoals The Internet Archive, het Kulturarw3 project van de Nationale Bibliotheek in Zweden, het project Minerva van de Amerikaanse Library of Congress, en het Pandora project van de Nationale Bibliotheek in Australië. In Nederland is Archipol, het archief met websites van politieke partijen, het eerste website-archief. De initiatieven komen voort uit de bibliotheekwereld en zijn vooral gericht op het archiveren van websites als “publicaties”. Inmiddels hebben ook archiefdiensten en recordsmanagers zich gericht op het archiveren van websites, vooral in het ontwerpen van richtlijnen voor het ontwerpen en beheren van websites door archiefvormers en het formuleren van acquisitiebeleid voor websites.

Het in 1996 opgerichte Internet Archive, [www.archive.org](http://www.archive.org), maakt sinds 1996 periodiek kopieën (snapshots) van alle beschikbare websites in het world wide web met als doel deze permanent toegankelijk te hebben voor onderzoekers. Dit kopiëren gebeurt volledig automatisch, via web crawlers. Het is niet mogelijk om alle websites te kopiëren. Websites die beschermd zijn tegen kopiëren worden niet opgenomen in The Internet Archive. Het archief bestond in januari 2002 uit zo'n 10 miljard webpagina's, ongeveer 100 miljoen terabytes. Aan The Internet Archive is een eigen zoekmachine toegevoegd, de Wayback Machine. De Wayback Machine geeft aan de hand van een opgegeven URL een overzicht van de aanwezige versies van de websites. Van de oudste websites (ca. 1996 – 1998) zijn vaak slechts enkele pagina's aanwezig, de nieuwere websites zijn vollediger. Sinds 2005 is het ook mogelijk zelf websites toe te voegen aan The Internet Archive via Archive-IT.

Een andere pionier is de Nationale Bibliotheek in Zweden. Het Kulturarw3 programma <http://www.kb.se/kw3/ENG/Default.htm>, ook gestart in 1996, heeft als doel het kopiëren, beheren en beschikbaar stellen van de websites van het Zweedse domein van het World Wide Web. Dat wil zeggen dat alle websites waarvan de URL eindigt op .se en alle Zweedse websites die eindigen op .com, .org, .net en .nu worden gekopieerd. Ook Kulturarw3 hanteert de bulk-aanpak; van alle beschikbare websites worden periodiek een snapshot genomen. Dit is inmiddels acht keer volledig gedaan. Het archief Kulturarw3 is momenteel niet toegankelijk via Internet.

De Nationale Bibliotheek van Australië hanteert in het programma Pandora <http://pandora.nla.gov.au/index.html> een andere aanpak. Binnen Pandora is gekozen voor selectie van websites en online publicaties. Dus geen bulkarchivering, maar archivering van een beperkt aantal websites uit het Australische domein. De keuze welke websites opgenomen worden in het webarchief volgt uit de algemene selectiecriteria van de Nationale Bibliotheek. Daarbij geldt ook nog dat publicaties die zowel in papieren vorm als digitaal beschikbaar zijn alleen in een papieren vorm worden bewaard. Ook Minerva <http://www.loc.gov/minerva>, het project van de Amerikaanse Library of Congress hanteert deze selectieve archiveringsstrategie. Minerva heeft bijvoorbeeld websites over de aanslagen op 11 september gearchiveerd.

Ook in Nederland loopt een aantal projecten. Het archief Archipol <http://www.archipol.nl> van het Documentatiecentrum voor Nederlandse Politieke Partijen (DNPP) en de Universiteitsbibliotheek van de Rijksuniversiteit Groningen bestaat al enige jaren. Dit project, gestart in 2000, heeft twee doelstellingen. Allereerst de inrichting van een digitaal archief van websites van politieke partijen dat gebruikt kan worden voor onderzoek en onderwijs. Ten tweede het ontwikkelen van een model van digitale archivering van websites, waarmee andere instellingen ook tot de archivering van websites kunnen overgaan. Inmiddels is de ontwikkelfase van het project afgerond. Het Archipol-archief bevat websites van Nederlandse politieke partijen, van politici en van nevenorganisaties. Archipol bevat ook het campagne-archief, met de websites van de politieke partijen gedurende de Tweede Kamerverkiezingen in 2002 en 2003 en de Europese verkiezingen van 2004.

Een ander archiveringsproject is Occasio, <http://www.iisg.nl/occasio>. Occasio is een samenwerkingsproject tussen de stichting Antenna, een onafhankelijke internetprovider uit Nijmegen, en het Internationaal Instituut voor Sociale Geschiedenis (IISG). In 1995 zijn IISG en Antenna gestart met het archiveren van webdocumenten over de oorlog in Joegoslavië (de Yugo-collectie) en de nieuwsgroepen van de Association for Progressive Communications (APC). Verzamelen, opslaan, ontsluiten en ter beschikking stellen van deze berichten verloopt volledig automatisch.

Het ministerie van Verkeer en Waterstaat ontplooit activiteiten voor het archiveren van haar eigen websites. Het Ministerie heeft hiervoor onderzoeken laten uitvoeren door de Archiefschool en Capsis. Tenslotte heeft het Gemeentearchief Amsterdam een onderzoek laten uitvoeren naar het archiveren van Amsterdamse websites, aansluitend bij het eigen acquisitiebeleid.

## 4.2.2 Bibliotheek

### Australië, Nieuw-Zeeland en Noord-Amerika

#### *Pandora: Preserving and Accessing Networked Documentary Resources of Australia (National Library of Australia)*

Australië

Pandora is een project van de National Library van Australië, gestart in 1996. Het is een klein archief, in juni 2001 bevatte het 1250 websites, waaronder de website van de Olympische Spelen in Sydney. Pandora verzamelt webpublicaties die alleen on-line beschikbaar zijn en die binnen de selectierichtlijnen van de NLA vallen. Pandora is gestart in 1996. Het project geeft informatie over de keuze voor selectieve versus bulk aanpak.

Zie:

Pandora : <http://pandora.nla.gov.au> 06-02-2006

- Pandora *Guidelines for the selection of Online Australian Publications Intended Preservation y the National Library of Australia* via <http://pandora.nla.gov.au/selectionguidelines.html> 06-02-2006 of: <http://pandora.nla.gov.au/bsc49.doc> 06-02-2006
- Pandora “The national Library of Australia’s Digital Preservation Agenda, an interview with Colin Webb” in: *RLG DigiNews* via: <http://www.rlg.org/preserv/diginews/diginews5-1.html> 06-02-2006
- Cathro, W.; Webb, C.; Whiting J. *Archiving the web: the Pandora Archive at the National Library of Australia*. (juni 2001) via: <http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html> 06-02-2006
- Daniels, Anne PANDORA - Archiving electronic publications. *M/C Reviews* (September 15, 1999). <http://www.uq.edu.au/mc/reviews/features/ejournal/pandora.html>
- Gatenby, Pam *Digital Archiving - Developing Policy And Best Practice Guidelines At The National Library Of Australia*. ICSTI Forum No.33 (March 2000). <http://www.icsti.org/forum/33/#Gatenby> 06-02-2006
- Meacham, Steve *Editing our future* artikel over Pandora in: The Sydney Morning Herald 4 mei 2005 <http://www.smh.com.au/articles/2005/05/03/1115092503076.html> 06-02-2006
- Phillips, Margaret E. (2003). "PANDORA, Australia's Web archive, and the digital archiving system that supports it." *DigiCULT.info*, 6, 24-28. December 2003. [http://www.digicult.info/downloads/dc\\_info\\_issue6\\_december\\_20031.pdf](http://www.digicult.info/downloads/dc_info_issue6_december_20031.pdf) 06-02-2006
- Philips, Margaret E. Cunningham, Adrian *Keeping online information accessible for E-government and E-democracy* (2005) <http://www.nla.gov.au/nla/staffpaper/2004/phillips1.html> 06-02-2006
- Philips, Margaret E. *Selective Archiving of Web Resources: A Study of Acquisition Costs at the National Library of Australia* (2005) [http://www.rlg.org/en/page.php?Page\\_ID=20666#article0](http://www.rlg.org/en/page.php?Page_ID=20666#article0) 06-02-2006
- Webb, Colin; Preiss, Lydia “Who will save the Olympics? The Pandora archive and other digital preservation case studies at the National Library of Australia” in: *Digital Past, Digital Future - An Introduction To Digital Preservation*. OCLC/Preservation Resources Symposium (2001). <http://www.oclc.org/education/conferences/presentations/2001/preservation/preisswebb.htm> 06-02-2006

### ***Electronic Collection. A virtual collection of Monographs and Periodicals***

Canada

Deze collectie bestaat uit Canadese websites en digitale boeken en tijdschriften. Na een onderzoek in 1993 voerde de National Library of Canada in 1994-1995 een project uit waarin Canadese digitale tijdschriften werden verzameld.

Zie: <http://www.collectionscanada.ca/electroniccollection/003008-200-e.html> 06-02-2006

### ***Minerva: Mapping the Internet: the Electronic Resources Virtual Archive (Library of Congress)***

Verenigde Staten

Minerva is het website archief van de Amerikaanse Library of Congress. Doel van het project was in eerste instantie ervaring opdoen met het verzamelen en archiveren van (voor iedereen toegankelijke) websites. Minerva heeft ervoor gekozen om van een geselecteerd aantal websites snapshots te archiveren. Het

archief is gestart in 2000 en heeft vooral informatie opgeleverd over de problemen bij het downloaden van websites; kosten van selectieve versus bulk selectie; en verschillende soorten onderzoek dat gedaan kan worden met gearchiveerde websites.

Zie:

Minerva website <http://lcweb2.loc.gov/cocoon/minerva/html/minerva-home.html> 06-02-2006

- Arms, W.Y. *Web preservation Project, interim report* (januari 2001) via <http://lcweb2.loc.gov/cocoon/minerva/html/minerva-home.html> 06-02-2006
- Arms, W.Y. *Web preservation Project, final report* (september 2001) via <http://lcweb2.loc.gov/cocoon/minerva/html/minerva-home.html> 06-02-2006
- Arms, W.Y.; Adkins, R.; Ammen, C.; Hayes, A. "Collecting and Preserving the Web: The Minerva Prototype" in *RLG DigiNews* (april 2001) via: <http://www.rlg.org/preserv/diginews/diginews5-2.html> 06-02-2006

## Europa

### *Bibliothèque Nationale de France*

Frankrijk

- Masanès, Julien; Lupovici, Catherine (eds.): *Workshop Notes of the ECDL 2001 workshop - What's next for Digital Deposit Libraries?*, Darmstadt, Germany, 8th September 2001. ECDL 2001. <http://bibnum.bnf.fr/ecdl/2001/index.html> 06-02-2006
- Masanès, Julien "Developing a French web archive." In: *DPC Forum: Web-archiving - managing and archiving online documents and records*. London; 25. März 2002. <http://www.dpconline.org/graphics/events/presentations/pdf/Masanes.pdf> 06-02-2006
- Abiteboul, S.; Preda, M.; Cobéna G. "Computing Web page importance without storing the graph of the Web" (extended abstract). In: *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 25 (1), 27-33 (2002). Via: [http://wotan.liu.edu/docis/dbl/daenbu/2002\\_25\\_1\\_27\\_CWPIWS.html](http://wotan.liu.edu/docis/dbl/daenbu/2002_25_1_27_CWPIWS.html) 06-02-2006
- Abiteboul, S.; Cobena, G; Masanes, J; Sedrati, G.; "A First Experience in Archiving the French Web" in: Maristella Agosti, Costantino Thanos (eds): *Proceedings of the ECDL 2002*. Rome, Italy (September 2002). Springer - Lecture Notes in Computer Science. <ftp://ftp.inria.fr/INRIA/Projects/verso/gemo/GemoReport-229.pdf> 06-02-2006
- Masanès, Julien "Towards Continuous Web Archiving: First Results and an Agenda for the Future." In: *D-Lib Magazine*, v.9, n.12 (December 2002). <http://www.dlib.org/dlib/december02/masanes/12masanes.html> 06-02-2006

### *AOLA: Austrian On-Line Archive*

Nationale Bibliotheek Oostenrijk en de Universiteit in Wenen

Zie: Universiteit: <http://www.ifs.tuwien.ac.at/~aola/> 06-02-2006

- Aschenbrenner, Andreas *Long-Term Preservation of Digital Material - Building an Archive to Preserve Digital Cultural Heritage from the Internet*. Master Thesis, Technical University Vienna (December 2001). <http://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando> 06-02-2006
- Rauber, Andreas; Aschenbrenner, Andreas *Part of Our Culture is born digital – on efforts to preserve it for future generations* via: <http://www.inst.at/trans/10Nr/rauber10.htm> 06-02-2006
- Rauber, Andreas; Aschenbrenner, Andreas "Uncovering information Hidden in Web Archives" in: *D-Lib magazine* (december 2002) via: <http://www.dlib.org/dlib/december02/rauber/12rauber.html> 06-02-2006

- Rauber, Andreas; Aschenbrenner, Andreas *Die Bewahrung unserer Online-Kultur, Vorschläge zu Strategien der Webarchivierung* (2003) via:  
<http://www.onb.ac.at/sichtungen/beitraege/aschenbrenner-a-1a.html> 06-02-2006

### **Kulturarw3 (Sveriges nationalbibliotek)**

Zweden

De Zweedse Nationale bibliotheek archiveert via het Kulturarw3 programma, ook gestart in 1996, de websites van het Zweedse domein van het World Wide Web. Er is sprake van bulkarchivering, inmiddels is twaalf keer een volledige kopie van de websites in het Zweedse domein en van Zweedse websites waarvan de URL eindigt op .org, .com, .net en .nu gemaakt.

Zie: <http://www.kb.se/kw3/ENG> 06-02-2006

- Arvidson, Allan; Persson, Krister, *The Kulturarw3 Project* presentatie tijdens de 66<sup>e</sup> IFLA conferentie in Jeruzalem via: <http://www.ifla.org/IV/ifla66/papers/154-157e.htm> 06-02-2006

### **Nordic Web Archive (NWA)**

#### **NWA Toolset, WERA (Web aRchive Access)**

Scandinavië

De NWA is een forum van de Scandinavische bibliotheken voor coördinatie en informatie-uitwisseling op het gebied van verzamelen en archivering van web documenten. Van november 2000 – juli 2002 heeft het NWA zich bezig gehouden met het ontwikkelen van software voor de toegang tot gearchiveerde web documenten. De NWA toolset bestaat uit een retriever, een exporter en een acces module. Deze NWA toolset is verder ontwikkeld en heeft de basis gevormd voor WERA. WERA is een viewer applicatie waarmee het mogelijk is om in een collectie gearchiveerde websites te zoeken (ook full text).

Zie: <http://nwa.nb.no> 06-02-2006

- *NWA Toolset* Nordic Web Archive (NWA) via: <http://nwatoolset.sourceforge.net> 06-02-2006
- *WERA* via: <http://archive-access.sourceforge.net/projects/wera> 06-02-2006

### **Netarchive.dk**

Koninklijke Bibliotheek Denemarken

Zie voor fase 1: [www.netarchive.dk](http://www.netarchive.dk) voor fase 2: <http://netarchive.dk/fase2index-en.php> 06-02-2006

- Christensen-Dalsgaard, Birte e.a. *Experiences and conclusions from a pilot study: Web archiving of the District and County elections 2001. Final Report for The Pilot Project "netarkivet.dk"* (February 2003).  
<http://netarchive.dk/rap/index-en.php> 06-02-2006
- Christensen-Dalsgaard, Birte "Web Archive Activities in Denmark" in: *RLG DigiNews* (june 15, 2004) [http://www.rlg.org/en/page.php?Page\\_ID=17661#article0](http://www.rlg.org/en/page.php?Page_ID=17661#article0) 06-02-2006
- Christensen, Niels H. *Towards format repositories for web archives* (IWA04, 2004) via:  
<http://netarchive.dk/website/publications/FormatRepositories-2004.pdf> 06-02-2006
- Christensen, Steen S. *Archival Data Format Requirements* (2004) via:  
[http://netarchive.dk/website/publications/Archival\\_format\\_requirements-2004.pdf](http://netarchive.dk/website/publications/Archival_format_requirements-2004.pdf) 06-02-2006
- Clausen, Lars R. *Concerning Etags and Datstamps* (IWA04, 2004) via:  
<http://netarchive.dk/website/publications/Etags-2004.pdf> 06-02-2006
- Clausen, Lars R. *Handling file formats* (2004) via:  
<http://netarchive.dk/website/publications/FileFormats-2004.pdf> 06-02-2006

- Henriksen, Birgit “Danish Legal Deposit on the Internet: Current Solutions and Approaches for the Future. Presentation” In: *European Conference on Digital Libraries (ECDL) 2001: What's next for Digital Deposit Libraries?*. Workshop. Darmstadt, Germany; 8. September 2001.  
<http://netarchive.dk/papers/ECDL2001-bnh-08092001.ppt#310>, 06-02-2006

### **UK Webarchiving Consortium (UKWAC)**

De volgende instellingen vormen samen het UK Webarchive Consortium: The British Library, The National Archives, The Wellcome Library, The National Library of Scotland, The National Library of Wales/ Llyfrgell Genedlaethol Cymru, The Joint Information Systems Committee (JISC).

Zie: <http://info.webarchive.org.uk/index.html> 06-02-2006

- Bailey, Steve and Thomson, Dave *UKWAC: Building the UK's First Public Web Archive* D-Lib Magazine, v.12, n.11 (January, 2006) via:  
<http://www.dlib.org/dlib/january06/thompson/01thompson.html> 06-02-2006
- Charlesworth, Andres *Legal issues relating to the archiving of Internet resources in the UK, EU, US and Australia* (version 1.0, february 2003) via: [http://library.wellcome.ac.uk/doc\\_WTL039231.html](http://library.wellcome.ac.uk/doc_WTL039231.html) 06-02-2005
- Day, Michael *Collecting and preserving the World Wide Web: a feasibility study undertaken for the JISC and Wellcome Trust* (version 1.0, february 2003) via:  
[http://library.wellcome.ac.uk/doc\\_WTL039231.html](http://library.wellcome.ac.uk/doc_WTL039231.html) 06-02-2005

### **Archivserver DEPOSIT.DDB.DE**

Die Deutsche Bibliothek

Zie: <http://deposit.ddb.de> 06-02-2005

- Liegmann, Hans; Berkemeyer, Jörg; *Current activities of Die Deutsche Bibliothek in the field of electronic publications..* <http://www.nla.gov.au/padi/topics/DDB.html> 06-02-2005
- Liegmann, Hans; “Collection of German online resources by Die Deutsche Bibliothek” In: *European Conference on Digital Libraries (ECDL) 2001: What's next for Digital Deposit Libraries?*. Workshop. Darmstadt, Germany; 8. September 2001.  
<http://bibnum.bnf.fr/ecdl/2001/germany/sld001.htm> 06-02-2005

### **DAMP (Digital Archive for Web Publications)**

Project van Universiteit van Zagreb (SRCE) en de National and University Library (NUL) in Zagreb

- Milinovic, Miroslav; Topolscak, Nebojsa “The architecture of DAMP” in: *Wiswisawn* vol3 issue 3 via: [http://widwisawn.cdlib.org/Issues/Vol3/issue3\\_3\\_1.html](http://widwisawn.cdlib.org/Issues/Vol3/issue3_3_1.html) 06-02-2005

### **Politisches Internet Archiv**

Project van 5 Duitse organisaties.

Zie: <http://www.fcs.de/archiv/spiegelung/default.htm> 06-02-2005

- Schmitz, Rudolf *Internet Archivierung* (2004) via: <https://eldorado.uni-dortmund.de/bitstream/2003/21378/1/schmitz.pdf> 06-02-2005
- Schmitz, Rudolf “Aufbau und Struktur eines Internet-Archivs. Die Archivierung von Internet-Auftritten der SPD und ihrer Fraktionen in den Parlamenten” in *Der Archivar* (2004-4) p. 318-320 en via: [http://www.archive.nrw.de/archivar/2004-04/Archivar\\_2004-4.pdf](http://www.archive.nrw.de/archivar/2004-04/Archivar_2004-4.pdf) 06-02-2005

## Nederland

### *Archipol*

Documentatiecentrum Nederlandse Politieke Partijen en Bibliotheek Rijksuniversiteit Groningen

Het archief van websites van politieke partijen, Archipol, is gestart in 2000. Inmiddels is de ontwikkelfase van het project afgerond en bevat het archief snapshots van websites van politieke partijen en hun nevenorganisaties en partijleiders. Het archief is alleen toegankelijk na aanvraag van een wachtwoord. De website bevat een aantal artikelen over de ontwikkeling van Archipol

Zie:

Archipol: <http://www.archipol.nl> 06-02-2005

- Keijzer, A.K.; Hollander, F.J. den; Voerman, G. “Het Archipol-project, het archiveren van websites van Nederlandse politieke partijen” in *Archievenblad* (februari 2002) p. 32 – 33.

### *Koninklijke Bibliotheek*

De KB is in december 2005 gestart met een project dat de mogelijkheden onderzoekt van het duurzaam opslaan en toegankelijk maken van websites via het e-Depot.

Zie: <http://www.kb.nl/hrd/dd/dd.html> en

[http://www.kb.nl/hrd/dd/dd\\_projecten/projecten\\_webarchivering.html](http://www.kb.nl/hrd/dd/dd_projecten/projecten_webarchivering.html) 06-02-2006

- Verhoeven, Hans *Archiving web publications* (IBM/Koninklijke bibliotheek, Den Haag, december 2002) via: [http://www.kb.nl/hrd/dd/dd\\_onderzoek/summary\\_ltpstudy6.html](http://www.kb.nl/hrd/dd/dd_onderzoek/summary_ltpstudy6.html) 06-02-2006
- Hakala, Juha ‘Collecting and Preserving the Web: Developing and Testing the NEDLIB Harvester’ in: *RLG DigiNews* (april 2001) via: <http://www.rlg.org/preserv/diginews/diginews5-2.html> 06-02-2006

## 4.2.3 Archieven

### Australië, Nieuw-Zeeland en Noord-Amerika

#### *National Archives of Australia*

Australië

Het NAA heeft beleid ontwikkeld voor de archivering van websites. Dit beleid is vastgelegd in een Policystatement en in Guidelines. In de Guidelines worden een object georiënteerde en een ‘event’ (interactie) georiënteerde strategie naast elkaar gezet.

Zie:

[http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) 06-02-2006

- NAA *Archiving Web Resources: a policy for Keeping Records of Web-bases Activity in the Commonwealth Government* (januari 2001) via: [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) 06-02-2006
- NAA *Archiving Web Resources: Guidelines for Keeping Records of Web-bases Activity in the Commonwealth Government* (Maart 2001) via: [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html) 06-02-2006

- Philips, Margaret E. Cunningham, Adrian *Keeping online information accessible for E-government and E-democracy* (2005) <http://www.nla.gov.au/nla/staffpaper/2004/phillips1.html> 06-02-2006
- Gibbs, Ross *Archiving Websites: the Australian experience* (CITRA, 2005) via: <http://www.ica.org/biblio/abstract%20Ross%20Gibbs%20ENG.pdf> 06-02-2006

***NARA Web snapshot project***  
***National Archives and Records Administration (NARA)***

Verenigde Staten

Het NARA heeft in januari 2001 snapshots van overheidswebsites verzameld. Aan overheidsorganisaties is gevraagd om een snapshot van de website aan te leveren samen met metadata. Voor de registratie van de metadata heeft het NARA een formulier ontworpen. In 2005 heeft het NARA richtlijnen voor de archivering van websites gepubliceerd.

Zie:

NARA website:

[http://www.archives.gov/records\\_management](http://www.archives.gov/records_management) 06-02-2006

- NARA web snapshot project via: <http://www.archives.gov/records-mgmt/basics/snapshot-public-web-sites.html> 06-02-2006
- NARA *guidance on managing web records* (January 2005) via: [http://www.archives.gov/records\\_management/pdf/managing\\_web\\_records\\_index.pdf](http://www.archives.gov/records_management/pdf/managing_web_records_index.pdf) of via <http://www.archives.gov/records-mgmt/policy/managing-web-records-index.html> 06-02-2006

## **Europa**

***National Archives UK***

Groot-Brittannië

The National Archives UK (voorheen Public Records Office) heeft in 2001 een beleidsdocument gepubliceerd over de archivering van websites. *Managing Web resources. Management of electronic records on websites and intranets: an ERM toolkit* (version 1.0, december 2001)

The National Archives vormen samen met een aantal engelse bibliotheken het UK Webarchiving Consortium. (zie hierboven)

Zie:

National Archives UK: <http://www.nationalarchives.gov.uk/electronicrecords/advice> 06-02-2006

UK Webarchiving Consortium: <http://info.webarchive.org.uk/index.html> 06-02-2006

- *Managing Web resources. Management of electronic records on websites and intranets: an ERM toolkit* (version 1.0, Public Records Office London december 2001) via: of: [http://www.nationalarchives.gov.uk/electronicrecords/advice/pdf/website\\_toolkit.pdf](http://www.nationalarchives.gov.uk/electronicrecords/advice/pdf/website_toolkit.pdf) 06-02-2006

***DAVID***

Digitale archivering in Vlaamse Instellingen en Diensten.

Zie: <http://www.antwerpen.be/david/website/nl/index2.htm> 06-02-2006

- Boudrez, F.; Eynde, S. van den. *DAVID, archivering van websites* (Antwerpen - Leuven, mei 2002) Via <http://www.antwerpen.be/david/website/default.htm> 06-02-2006

- Boudrez, F. *Van back-up tot gearchiveerde website* (Antwerpen, 2002) Via <http://www.antwerpen.be/david/website/default.htm> 06-02-2006
- Boudrez, F. *Archivering van websites: een kwestie van waardering en 'capture'* (Antwerpen, 2005) via <http://www.antwerpen.be/david> 06-02-2006
- Boudrez, F. *Websitesbeheer en contentmanagement voor digitale archivering* versie 2.0 (Antwerpen, 2005) via <http://www.antwerpen.be/david/website/teksten/Richtlijn5.pdf> 06-02-2006
- Boudrez, Filip; Schoups, Inge "Duurzame en archiveerbare websites" in: *Praktijkgids management lokale besturen* (juni 2005) p. 3600/1 t/m 3600/34

## Nederland

### *Occasio*

IISG

In 1995 zijn IISG en Antenna gestart met het archiveren van webdocumenten over de oorlog in Joegoslavië en de nieuwsgroepen van de Association for Progressive Communications (APC). Verzamelen, opslaan ontsluiten en ter beschikking stellen van deze berichten verloopt volledig automatisch.

Zie verder:

<http://www.iisg.nl/occasio> 06-02-2006

- Quast, J. "Het internetarchief van het IISG" in *Archievenblad* (September 2000) p. 16 – 17
- Quast, J. *The Internet Archive of the IISH* (February 2005) via <http://www.iisg.nl/occasio/occasio-uk.pdf> 06-02-2006

### *Gemeentearchief Amsterdam*

Websites zijn er genoeg, maar hoe zit het met het archiveren ervan? De ervaring van het Gemeentearchief Amsterdam met het archiveren van websites, bracht men tot de conclusie dat het noodzakelijk was beleid te formuleren voor het archiveren van websites in Amsterdam. Beleid dat iets zou moeten zeggen over selectie van websites, overbrengingstermijnen, metadata, opslagformaten en hyperlinks.

- Digital Display & gemeentearchief Amsterdam *Het archiveren van websites in de gemeente Amsterdam* (Amsterdam, juni 2003) via: [http://www.ddisplay.nl/docs/2003ddisplay\\_archiveren\\_websites.pdf](http://www.ddisplay.nl/docs/2003ddisplay_archiveren_websites.pdf) 06-02-2006

### *Ministerie van Verkeer en Waterstaat*

Het ministerie van Verkeer en Waterstaat wil haar eigen websites archiveren. Ter voorbereiding hiervan heeft het Ministerie een aantal onderzoeken laten uitvoeren.

- Gun, J.G.E.M. van der *Gevangen van het Net. Onderzoek naar de archivering van intranetsites* (Amsterdam, 2003) zie: <http://www.archiefschool.nl/docs/intranetarchivering.pdf> 06-02-2006
- Hokke, H.A. *Blijvend beschikbaar. Onderzoek naar de Archivering van websites* (Amsterdam, Archiefschool, ministerie van Verkeer en Waterstaat, 2003) zie: [http://www.elo.nl/elo/Images/blijbeschik\\_tcm70-20514.doc](http://www.elo.nl/elo/Images/blijbeschik_tcm70-20514.doc) 06-02-2006
- Hokke, H.A. *Naar archivering van websites. Implementatieadvies* (Amsterdam, Archiefschool, ministerie van Verkeer en Waterstaat, 2003)
- Voorburg, René J.J. *Webontwerp: richtlijnen voor archivering* (Amsterdam, Uselab & ministerie van Verkeer en Waterstaat, 2004) [http://digidiv.amsterdam.nl/handleidingen/digitale\\_duurzaamheid/downloads/webont.pdf](http://digidiv.amsterdam.nl/handleidingen/digitale_duurzaamheid/downloads/webont.pdf) 06-02-2006

- Voorburg, R. J.J.; Goutier, J.L.E *Webarchivering bij het Ministerie van Verkeer en Waterstaat. Verslag van een onderzoek* (Capsis BV, Ministerie van Verkeer en Waterstaat, 2005)  
[http://www.capsis.nl/files/verslag\\_onderzoek\\_webarchivering\\_minvenw\\_capsis\\_publicatie\\_20050822.pdf](http://www.capsis.nl/files/verslag_onderzoek_webarchivering_minvenw_capsis_publicatie_20050822.pdf) 06-02-2006

#### 4.2.4 Overige initiatieven

##### *The Internet Archive*

The Internet Archive verzamelt sinds 1996 websites in het publieke domein van het world wide web. Er is sprake van bulkarchivering, alle beschikbare websites worden periodiek gekopieerd. De verschillende versies van de websites zijn via de Wayback Machine per URL te bezichtigen. Sinds 2005 is het mogelijk zelf een collectie van websites toe te voegen aan The Internet Archive, via Archive IT-org.

Zie:

Internet Archive: [www.archive.org](http://www.archive.org) 06-02-2006

Archive IT-org: <http://www.archive-it.org/index.html> 06-02-2006

- Richard Koman *How the Wayback Machine Works* (1-12-2002) via  
<http://www.oreillynet.com/pub/a/webservices/2002/01/18/brewster.html> 06-02-2006
- “The Internet Archive, an interview with Brewster Kahle” in: *RLG DigiNews* via:  
<http://www.rlg.org/preserv/diginews/diginews6-3.html> 06-02-2006

##### *Europarchive*

Half februari gaat het Europese internetarchief Europarchive.org van start, een initiatief van onder andere Xs4all. De organisatie zal naast het aanleggen van een Europese tegenhanger van de Internet Archive ook gratis collecties "historische" content zoals klassieke muziek ter download aanbieden.

Zie: <http://www.europarchive.org> 06-02-2006

- Neeve, Mick de *Europees Internetarchief deze maand van start* (2006) via:  
<http://www.tweakers.net/nieuws/40997> 06-02-2006
- Artikel “Europa krijgt eigen versie Internet Archive” *Computable* (2006)  
[http://www.computable.nl/nieuws.htm?id=1108097&WT.mc\\_id=nb](http://www.computable.nl/nieuws.htm?id=1108097&WT.mc_id=nb) 06-02-2006

##### *International Internet Preservation Consortium*

In het International Internet Preservation Consortium is opgericht in 2003 door een aantal Nationale Bibliotheken en The Internet Archive. Haar doelstelling is om informatie die gepubliceerd wordt op internet te bewaren en toegankelijk te houden. Het IIPC ontwikkelt o.a. instrumenten en standaards. Via de website zijn een aantal publicaties over webcrawlers en verwijzingen naar verschillende applicaties te vinden.

Zie: Netpreserve: <http://www.netpreserve.org/about/index.php> 06-02-2006

##### *Smithsonian Institution Archives*

Op verzoek van het Smithsonian heeft Charles Dollar een advies uitgebracht over de archivering van de websites van het Smithsonian. Het Smithsonian beheerde 75 websites in 2001. In het advies stelt Dollar dat de websites van het Smithsonian zeker archiefstukken bevatten, omdat ze kunnen dienen als bewijs van het proces van het verspreiden van kennis. (Verspreiden van kennis is de belangrijkste doelstelling van het Smithsonian.) Dollar gaat in het rapport onder andere in op de verschillende webdesign mogelijkheden en de gevolgen daarvan voor archivering.

Zie: [http://siarchives.si.edu/records/electronic\\_records/records\\_erecords\\_publications.html](http://siarchives.si.edu/records/electronic_records/records_erecords_publications.html) 06-02-2006

- Dollar, Ch. *Archival preservation of Smithsonian Web Resources: strategies, principles and best practices* (juli 2001) via: <http://www.si.edu/archives/archives/dollar%20report.html> 06-02-2006
- Dollar, Ch. *Archival Preservation of Web resources, digital quality assurance tools* (October, 2002) <http://www.si.edu/archives/archives/dollarrpt3.pdf> 06-02-2006
- Dollar, Ch. *Archival preservation of Web Resources: HTML to XHTML migration test technical considerations, evaluation, and recommendations* (July, 2002) <http://siarchives.si.edu/pdf/dollarrpt2.pdf> 06-02-2006
- Dollar, Ch. *Archiving Smithsonian Websites: an evaluation and recommendation for a Smithsonian Institution Archives Pilot Project* (May, 2003) <http://siarchives.si.edu/pdf/websitpilot.pdf> 06-02-2006

### **Google Archivering Usenet**

Google beheert sinds februari 2001 de Deja.com's Usenet Discussion Service. In dit archief van usenet-berichten zijn berichten te vinden die terug gaan tot 1981.

Zie

Google groups: [http://www.google.com/googlegroups/archive\\_announce\\_20.html](http://www.google.com/googlegroups/archive_announce_20.html) 06-02-2006

- Sullivan, Danny *Google Acquires Deja Newsgroup Service* (maart 2001) via <http://searchenginewatch.com/sereport/01/03-google.html> 06-02-2006

### **Archiefschool**

De archiefschool doet onderzoek naar het archiveren van websites.

Zie: <http://www.archiefschool.nl/onderzoek/projecten.htm#ar> 06-02-2006

- Goutier, Hans; Hokke, Erika “Blijvend beschikbaar op Internet: Archivering van websites” in: *OD* september 2003
- Gun, J.G.E.M. van der *Gevangen van het Net. Onderzoek naar de archivering van intranetsites* (Amsterdam, 2003) zie: <http://www.archiefschool.nl/docs/intranetarchivering.pdf> 06-02-2006
- Hokke, Erika *Blijvend beschikbaar. Onderzoek naar de Archivering van websites* (Amsterdam, Archiefschool, ministerie van Verkeer en Waterstaat, 2003) zie: [http://www.elo.nl/elo/Images/blijbeschik\\_tcm70-20514.doc](http://www.elo.nl/elo/Images/blijbeschik_tcm70-20514.doc) 06-02-2006
- Hokke, H.A. *Naar archivering van websites. Implementatieadvies* (Amsterdam, Archiefschool, ministerie van Verkeer en Waterstaat, 2003) zie: [http://www.elo.nl/elo/Images/implarchwebs\\_tcm70-20515.doc](http://www.elo.nl/elo/Images/implarchwebs_tcm70-20515.doc) 06-02-2006
- Hokke, H.A. “Archiveren van websites” in: *Archiefbeheer in de praktijk* (Kluwer Alphen a/d Rijn, september 2003, pagina 5780-1 t/m 5780-36)

### **Capsis**

Capsis heeft een applicatie Présurf ontwikkeld voor de archivering van websites.

Zie: [www.capsis.nl](http://www.capsis.nl) 06-02-2006

- *Capsis:Specificaties Présurf v1.0* (2004) [http://www.capsis.nl/files/Specs\\_presurf\\_1.0.pdf](http://www.capsis.nl/files/Specs_presurf_1.0.pdf) 06-02-2006

- Voorburg, R. J.J.; Goutier, J.L.E. *Webarchivering bij het Ministerie van Verkeer en Waterstaat. Verslag van een onderzoek* (Capsis BV, Ministerie van Verkeer en Waterstaat, 2005)  
[http://www.capsis.nl/files/verslag\\_onderzoek\\_webarchivering\\_minvenw\\_capsis\\_publicatie\\_20050822.pdf](http://www.capsis.nl/files/verslag_onderzoek_webarchivering_minvenw_capsis_publicatie_20050822.pdf) 06-02-2006

### 4.3 Congressen

Hieronder een overzicht van links naar verslagen van congressen over archivering van websites. De European Conference on Digital Libraries (ECDL) heeft al meerdere congressen over dit onderwerp georganiseerd. Het overzicht van congressen is chronologisch.

- Preserving the present for the future - Strategies for the Internet, Copenhagen (18.-19. June 2001). <http://www.kb.dk/tilmeld/program.htm> 06-02-2006
- European Conference on Digital Libraries (ECDL) 2001: What's next for Digital Deposit Libraries?. Workshop. Darmstadt, Germany (8.September 2001).  
<http://bibnum.bnf.fr/ecdl/2001/index.html> 06-02-2006
- International Symposium on Web Archiving, NDL - Tokyo (30.January 2002).  
<http://www.ndl.go.jp/en/publication/proceedings/index.html> 06-02-2006  
Archipol – Het web gevangen, symposium over het archiveren van websites, Groningen, 25 april 2002 <http://www.archipol.nl/symposium.html> 06-02-2006
- DPC Forum. Web-archiving: managing and archiving online documents and records. London (March 25, 2002). <http://www.jisc.ac.uk/dner/preservation/webforum.html>
- European Conference on Digital Libraries (ECDL) 2002, Web Archiving Workshop (September 19, 2002). <http://bibnum.bnf.fr/ecdl/2002/index.html> 06-02-2006  
Michael Day, 2nd ECDL Workshop on Web Archiving. Cultivate Interactive n.8 (November 2002). <http://www.cultivate-int.org/issue8/ecdlws2/> 06-02-2006
- ALPSP, 9th Publishing and the Internet Seminar: Archiving - whose problem is it?. The Association of Learned and Professional Society Publishers (ALPSP). London; 29 November 2002. <http://www.alpssp.org/events/previous/default.htm> 06-02-2006
- Erpanet – Preserving the web, Kerkira, 22<sup>nd</sup> - 24<sup>th</sup> May 2003.  
<http://www.erpanet.org/events/2003/kerkira/index.php>
- Archiefschool – Blijvend beschikbaar: symposium over de archivering van websites, 12 juni 2003,  
<http://www.archiefschool.nl/onderzoek/lezingen.htm#sym> 06-02-2006
- 3<sup>rd</sup> ECDL Workshop on Web Archiving, august 21<sup>st</sup>, Trondheim.  
<http://bibnum.bnf.fr/ecdl/2003/index.html> 06-02-2006  
Michael Day: 3rd ECDL Workshop on Web Archiving. *Adiadne Issue 37*, October 2003.  
<http://www.ariadne.ac.uk/issue37/ecdl-web-archiving-rpt/> 06-02-2006
- Archiving Web Resources, Canberra, November 9-11, 2004 <http://www.nla.gov.au/webarchiving/> 06-02-2006
- 4<sup>th</sup> International web Archiving Workshop (IWA04) held in conjunction with the 8<sup>th</sup> European Conference on Research and Advanced technologies for Digital Libraries, September 16, Bath, UK, <http://www.iwaw.net/04/index.html> 06-02-2006  
Julien Msanès & Andreas Rauber - Report on the 4<sup>th</sup> International web Archiving Workshop (*D-lib Magazin*, November 2004) <http://www.dlib.org/dlib/november04/masanes/11masanes.html> 06-02-2006
- National Library of Australia: Archiving Web Resources - Issues for cultural heritage institutions. Canberra, 9-11 November 2004. <http://www.nla.gov.au/webarchiving/> 06-02-2006  
Margaret E. Philips – Archiving Web Resources International Conference: Issues for Cultural Heritage Organisations (*RLG DigiNews* volume 9 number 1, feb 15, 2005)  
[http://www.rlg.org/en/page.php?Page\\_ID=20522#article2](http://www.rlg.org/en/page.php?Page_ID=20522#article2) 06-02-2006

- International Conference on Preservation of Digital Object, September 15-16 September 2005, Göttingen, Germany: <http://rdd.sub.uni-goettingen.de/conferences/ipres/programme> 06-02-2006
- ECDL 2005, September 18-23, Vienna, Austria, <http://www.ecdl2005.org> 06-02-2006
- 5<sup>th</sup> International web Archiving Workshop (IWA05) held in conjunction with the 9<sup>th</sup> European Conference on Research and Advanced technologies for Digital Libraries, September 22 and 23 2005, Vienna, Austria <http://www.iwaw.net/05/> 06-02-2006
- Digital Library Foundation. Fall Forum 2005, November 7-9 2005, Charlottesville, VA, <http://www.diglib.org/forums/fall2005/fall2005abstracts.htm> 06-02-2006
- Digital Memory, Tallinn, 24-25 November 2005, National Library of Estonia <http://www.nlib.ee/31968> 06-02-2006
- Joint Workshop on Future-proofing Institutional Websites, 19–20 January 2006, Wellcome Library, London <http://www.dcc.ac.uk/training/fpw-2006/> 06-02-2006
- Öffentlicher workshop des DFG-Projekts Politisches Internet-Archiv am 7. Februar 2006 [http://www.fes.de/archiv/spiegelung/ver\\_8.htm](http://www.fes.de/archiv/spiegelung/ver_8.htm) 06-02-2006

#### 4.4 Portalsites, bibliografieën, discussielijst.

- Asschenbrenner, Andreas – *Web Archiving Bibliography* (april 2004) <http://www.ifs.tuwien.ac.at/~aola/links/WebArchiving.html> 06-02-2006
- Rick Barry: mybestdocs - *Hot Topics: WWW vs Records*. <http://www.mybestdocs.com> 06-02-2006
- PADI (*Preserving Access to Digital Information, National Library of Australia*): Legal Deposit. <http://www.nla.gov.au/padi/topics/67.html> 06-02-2006  
PADI Topic: Web Archiving <http://www.nla.gov.au/padi/topics/92.html> 06-02-2006
- The December 2004 issue of "*What's new in digital preservation*" has an extensive list of web archiving resources: <http://www.dpconline.org/docs/whatsnew/issue9.html#2.1> 06-02-2006
- Web archive list. This list is focused on on-line content archiving, from the technical, legal and organisational point of view. <http://listes.cru.fr/wvs/info/web-archive> 06-02-2006
- Nestor. Kompetenznetzwerk Langezeitarchivierung 06-02-2006 <http://langzeitarchivierung.de/index.php?newlang=eng> 06-02-2006