



de webpagina's

Er zijn verschillende - vaak complementaire - methoden om websites te archiveren:

- archiveren van de broncode. Een eenvoudige methode, maar enkel geschikt voor statische websites.
- archiveren van een snapshot: hiervoor is een *webharvester* of *webcrawler* nodig. Open Source voorbeelden zijn *HTTrack* (<http://www.httrack.com>) en *Heritrix* (<http://crawler.archive.org>). Deze programma's maken een lokale kopie van een website door webpagina's en bijbehorende grafische elementen en documenten naar de harde schijf te kopiëren. Van dynamische websites wordt zo een statische momentopname gemaakt die bewaard kan worden. Deze methode is doorgaans de meest geschikte voor het vastleggen van zowel statische als dynamische websites.
- archiveren van een surfsessie: terwijl een gebruiker een website bezoekt, registreert een screenrecorder alle webpagina's en instructies tijdens het bezoek aan een website. De sessie wordt vervolgens als een videobestand bewaard. Deze methode laat toe dat de originele "look & feel" van moeilijk archiveerbare websites, zoals flash-sites en pagina's geschreven in DHTML, in een open formaat worden vastgelegd.
- archiveren van de unieke responsen, waarbij elke aangevraagde pagina met een wijziging of versieverandering als een nieuwe unieke webpagina wordt beschouwd en vervolgens volledig automatisch gearchiveerd.



de metadata

Voor de op te nemen metadata wordt een uitbreidbaar en flexibel XML schema opgemaakt. De metadata worden waar mogelijk automatisch geëxtraheerd.



deep web archivering

Archiefdocumenten binnen de **deep web** applicaties - CMS, databanken, enz. - kunnen beter rechtstreeks vanuit die applicaties gearchiveerd worden. De bestaande automatische archiveringstools werken immers nog niet perfect op dit vlak. Omwille van de grote verscheidenheid aan applicaties kan er voor het archiveren van het deep web geen eenvormige oplossing worden aangereikt.

Meer informatie in: "Archiveren van websites: een kwestie van waardering en 'capture', Antwerpen."

een portaalsite voor het websitesarchief

Eens de websites gearchiveerd zijn, wil je ze misschien voor gebruikers raadpleegbaar maken. Een eenvoudige manier is het opzetten van een portaalsite die toegang geeft tot de gearchiveerde websites. Door de archiveringsactie is elke band met een server verbroken. Naast een opslagmedium is daarom enkel een browser nodig om de archiefdocumenten te raadplegen.

Nuttig is ook om de bijbehorende metadata via deze portaalsite ter beschikking te stellen. Zo staat de soms onontbeerlijke context informatie meteen ter beschikking van de gebruikers.

juridische vragen

Het toegankelijk maken van een digitaal depot via intra- en/of internet roept een aantal juridische vragen op in verband met het auteursrecht en de privacywetgeving. Een archief te beschikking stellen is geen vanzelfsprekendheid omdat aan de belangen van verschillende betrokkenen wordt ge-raakt.

In het kader van het Digitaal Depot project verschenen een aantal publicaties die antwoord bieden op juridische vragen die gepaard gaan met het archiveren van digitaal cultureel erfgoed. Deze - en vele andere - eDAVID publicaties zijn online terug te vinden:



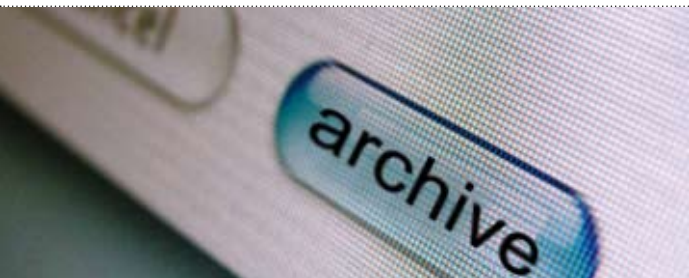
**websites
archiveren
in 4 stappen**



Websites zijn overal en nergens. Ze zijn dynamisch, multimediaal en interactief. Allerhande bedrijven, organisaties en particulieren hebben een eigen stek op het web, ook culturele instellingen en kunstenaars. In het Unesco 'Charter on the Preservation of the Digital Heritage' (2003) worden websites expliciet tot ons digitaal erfgoed gerekend.

Het aantal sites groeit aan een razend tempo maar oude webpagina's verdwijnen net zo snel: ze worden vervangen, geüpdatet of staan plotseling niet langer online.

Juridisch bevindt websitesarchivering zich nog in een schemerzone. Privacy en auteurswetgeving hebben de technologische evolutie niet helemaal bijgebeend. Toch is er geen tijd te verliezen, hoe langer we wachten hoe meer websites verloren gaan. Steeds meer bewaarinstellingen trachten dan ook om uiteenlopende redenen en op verschillende manieren een stukje van het WWW te vereeuwigen.



Efficiënt websites archiveren kan in 4 stappen:

stap 1: wat archiveren: het bepalen van de onderdelen van de website en de bijhorende informatie die gearchiveerd moeten worden.

stap 2: kwaliteitsvereisten: door van bij de ontwikkeling van een website met archivering rekening te houden en een aantal kwaliteitsvereisten in acht te nemen, kunnen de archiveringsacties efficiënter worden uitgevoerd.

stap 3: hoe archiveren: het selecteren van één of meerdere methoden om de website te archiveren.

stap 4: toegankelijk maken: na het archiveren de websites ook voor gebruikers ter beschikking stellen.

Vooraleer we van start gaan is het belangrijk te beslissen waarom, en bijgevolg wat, we willen archiveren. Welke aspecten van de website willen we zeker bewaren? Welke elementen zijn bepalend en maken de website tot wat hij is? Om na te gaan welke elementen essentieel zijn, bekijken we de verschillende onderdelen van de webpagina's, het achterliggende deep-web en de nodige metadata.



de webpagina's

Websites hebben een informatieve waarde. Die informatie is ingebed in **inhoud en structuur** van de afzonderlijke pagina's en de relatie tussen de pagina's onderling.

De pagina's bevatten ook visuele en interactieve elementen. Vaak vormen deze specifieke **'look & feel'** in combinatie met bepaalde **functionaliteiten** zoals hyperlinks en animaties een essentieel onderdeel van de website.



deep web archivering

Veel websites zijn tegenwoordig dynamisch. De webpagina zelf is niet meer dan een interface. De inhoud en lay-out van de pagina's worden *on the fly* opgebouwd. Inhoud, functionaliteit en structuur worden aangeleverd door **deep web** applicaties: Content Management Systemen (CMS), databank management systeem (DMS), documentbeheersystemen en fileservers.

Bij interactieve websites komen ook de **transacties en handelingen** voor archivering in aanmerking. De neerslag hiervan wordt doorgaans in één of meerdere databanken bijgehouden. Het archiveren van deze elementen is dus in de meeste gevallen een kwestie van databankarchivering.



de metadata

Voor een latere archiefgebruiker zal bepaalde **context informatie** onontbeerlijk zijn om de website volledig te kunnen plaatsen en begrijpen. Bijvoorbeeld de link met de actuele gebeurtenissen naar aanleiding waarvan een bepaalde webpagina online werd geplaatst, kan een archiefdocument in een volledig nieuw daglicht plaatsen.

Naast contextgegevens zijn er tal van andere technische en administratieve **metadata** die samen met de webpagina's kunnen worden bewaard: datum waarop de pagina online en offline ging, wie de webdesigners en webmaster waren, informatie over hard- en software van de webserver, enz.

De criteria voor een duurzame en archiveerbare website vallen in grote mate samen met de webtoegankelijkheidsregels. Zo levert het naleven van bepaalde kwaliteitsvereisten niet alleen **tijdwinst** en een archiveerbare website op, maar zorgt het er ook voor dat de website **goed toegankelijk** is.

Enkele eenvoudig toepasbare ontwerpvereisten voor duurzame en archiveerbare **webpagina's** zijn o.a.:

- werk een duidelijke uitbreidbare mappenstructuur voor de website uit
- zorg voor een duidelijke scheiding tussen enerzijds inhoud en structuur (HTML) en anderzijds opmaak (CSS)
- gebruik geen frames in de website
- maak vriendelijke, menselijk begrijpbare URL's
- gebruik absolute pathaanduidingen voor externe links en document-relatieve pathaanduidingen voor interne links

Websites worden steeds vaker aangestuurd vanuit een **CMS**. Ook voor het CMS zijn er richtlijnen. De beschikbare metadata velden (titel/naam, uniek webadres, versienummer, enz.) moeten bijvoorbeeld vrij definieerbaar zijn door de organisatie. Andere voorbeelden van functionele vereisten voor het CMS zijn:

- het is mogelijk de verschillende versies van content-items en hun metadata bij te houden
- het voorziet de mogelijkheid dat gepubliceerde content-items pas na versieverandering worden gewijzigd
- het kan de verschillende versies van online content op statische wijze en als afzonderlijke objecten, en in combinatie met hun metadata, bewaren

Andere vereisten zijn van toepassing op de **databank** gekoppeld aan het CMS, o.a.:

- de gegevens worden beheerd door een open DMS
- de databank heeft een gedocumenteerd, overzichtelijk en uitbreidbaar datamodel



Meer informatie over deze kwaliteitsvereisten vindt u in: "*Digitaal Archiveren: Richtlijn en Advies nr.5: Websitesbeheer en content management voor digitale archivering*"