

COLLECTIE

Jeroen Fernandez-Alonso
collectiemedewerker,
Amsab-ISG

WEBSITES ALS BRON VOOR DE GESCHIEDENIS VAN MORGEN

Dankzij een projectsubsidie van de Vlaamse overheid konden we ons het afgelopen jaar samen met Liberas verdiepen in de archivering van websites. Het project kreeg de naam *Catching the Digital Heritage*. Twee projectmedewerkers werden aangenomen om zich een jaar in deze materie te verdiepen. Zowel Amsab-ISG als Liberas archiveerden al meerdere jaren websites, maar deze collectie was nog nooit toegankelijk gemaakt voor het publiek. Ook de methode was verouderd.

Met het project wilden we vooral de collectie van reeds gearchiveerde websites toegankelijk maken. Maar hoe begin je daaraan? Voor het behoud en beheer van papieren archief en erfgoedobjecten bestaan heel wat internationale regels en normen, maar dat is niet het geval voor digitale bronnen zoals websites en sociale media. Pas enkele jaren geleden is er voor het eerst een aanzet gegeven tot internationale richtlijnen voor de archivering van websites, maar die richtlijnen zijn nog veel te vaag en eigenlijk meer gericht op bibliotheken. Dit maakte de hoofddoelstelling van het project niet eenvoudig: hoe kunnen we onze registratieachterstand inhalen als er nog geen concrete internationale normen bestaan voor het beschrijven van gearchiveerde websites? We hebben dit zeer pragmatisch

moeten aanpakken en altijd de afweging moeten maken tussen de noden van onderzoekers versus onze mogelijkheden op het vlak van collectiebeheerssoftware. De inventarisatie zelf is deels geautomatiseerd met behulp van de programmeertalen Python en Bash. Het merendeel van de essentiële metadata voor de beschrijving van een gearchiveerde website (URL, datum, softwareversie, extensies ...) konden we namelijk automatisch genereren. Hierdoor verliep de inventarisatie vlot.

Naast de inventarisatie van de reeds bestaande gearchiveerde websites zochten we ook naar een betere werkwijze om in de toekomst websites te archiveren. De methoden die in gebruik waren in Amsab-ISG en Liberas waren namelijk verouderd. De digitale wereld verandert snel en onze archiveringsmethoden moeten dit kunnen bijhouden. We hebben verschillende opensourceprogramma's getest. Uiteindelijk hebben we gekozen voor Heritrix. Het bekende '*internet archive*' maakt ook gebruik van dit programma en het wordt stilaan gezien als de standaard voor website-archivering. Maar de software is niet gemakkelijk. Hij is enkel te gebruiken op Linux-computers en er zijn heel wat aanpassingen nodig in de configuratie om te kunnen garanderen dat de websites volledig

worden behandeld. Websites zijn namelijk op verschillende manieren opgesteld, en niet elke website kan zomaar volledig gearchiveerd worden. Het komt dikwijls voor dat grote delen van een website niet of onvolledig zijn gearchiveerd.

Die kwaliteitsproblemen vormen een grote uitdaging. Met Heritrix is de kwaliteit van onze collectie al beter, maar onze archiefversie van een website zal nooit honderd procent identiek zijn aan de originele website, die destijds online zichtbaar was. Er zullen altijd kleine fouten in zitten. Dit vormt een uitdaging voor onderzoekers die gebruik willen maken van gearchiveerde websites: de authenticiteit kan onvoldoende gegarandeerd worden en de kans bestaat altijd dat de archiefkopie onvolledig is. Daarnaast moeten onderzoekers er rekening mee houden dat een gearchiveerde website slechts een momentopname is in de tijd. Informatie die tijdelijk online stond en alweer verwijderd is, wordt bij de archivering niet opgenomen.

Een ander nadeel voor onderzoekers is dat onze websitecollectie niet online raadpleegbaar mag zijn. Een website bestaat uit duizenden kleine onderdelen (foto's, teksten, video's, software, databanken ...), en elk van deze aspecten kan auteursrechtelijk beschermd zijn. Het is niet omdat een website destijds gratis online raadpleegbaar was, dat wij de archiefkopie ook zomaar online mogen plaatsen. De regels in verband met gearchiveerde websites zijn zeer onduidelijk. In de loop van het project hebben we contact gezocht met verschillende juridische experts, gaande van professoren van de UGent tot de juridische dienst van Sabam, maar niemand kon ons met zekerheid zeggen wat de rechtssituatie is. Hierdoor zijn we gedwongen om onze

collectie enkel intra muros raadpleegbaar te houden. De beschrijvingen en metadata zijn wel online te vinden.

De kennis die is opgedaan tijdens het project hebben we gedeeld via de Cultureel Erfgoed-Standaarden Toolbox en via een workshop. Professor Niels Brügger van de Deense Aarhus Universiteit zorgde op de workshop voor een boeiende keynote over het gebruik van gearchiveerde websites als historische bronnen. De powerpoints van de workshop zijn raadpleegbaar op de Cultureel Erfgoed-Standaarden Toolbox (www.projectcest.be).