

PROMISE: Het bewaren van veelsoortige online informatie: naar een Belgische strategie

SAMENVATTING

Context

Het internet is geëvolueerd tot een communicatiemiddel dat centraal staat in ons leven van elke dag en ook voor erfgoed is het een bijzonder waardevol instrument geworden. Als collecties van data en documenten vertegenwoordigen websites belangrijk materiaal dat dient te worden gearchiveerd. Net als voor andere publicaties en archieven, waarvan de preservering wordt gewaarborgd door de wetgeving op het wettelijk depot en de archiefwet, moet er voor het Belgische web een beleid van bewaring op lange termijn worden ontwikkeld.

Doelstellingen

Het project PROMISE werd gelanceerd om een antwoord te formuleren op de dringende vraag hoe de preservering van het Belgische web - als een belangrijk onderdeel van de Belgische geschiedenis - voor toekomstige generaties moet worden aangepakt. In het PROMISE project werden vier doelstellingen vooropgesteld:

- Actuele beste praktijken op het vlak van webarchivering ontwikkelen
- Een Belgisch beleid voor webarchivering op federaal niveau uitstippelen
- Webarchivering, toegang tot en gebruik van de pilootversie van het Belgische webarchief voor wetenschappelijk onderzoek sturen
- Aanbevelingen formuleren voor een duurzame Belgischewebarchiveringsstrategie

Methodologie

Op methodologisch vlak begon het project PROMISE data te verzamelen (uit wetteksten, bestaande initiatieven, de rol van instellingen) door middel van een literatuuroverzicht en semigestructureerde interviews met vertegenwoordigers van buitenlandse webarchivering initiatieven. Op basis van de bijeengebrachte informatie werd een realistische strategie en beleid uitgewerkt met het oog op het verzamelen en bewaren van het Belgische web. Dit omvatte een uitgebreide kostenberekening op basis van realistische scenario's voor webarchivering. Tevens werd een enquête gehouden om de vereisten van gebruikers van online archieven te analyseren; met de resultaten daarvan werd ook rekening gehouden bij het opstellen van een strategie.

Vervolgens werd het ontwikkelde model gevalideerd door een piloot-webarchief samen te stellen, bestaande uit de selectie en harvesting van de inhoud en het openstellen van de collecties voor toegang en gebruik. Tot slot formuleerde het project PROMISE een aantal aanbevelingen voor de implementatie van een duurzame dienst voor webarchivering in België, inclusief wettelijke en operationele perspectieven. Op wettelijk vlak werd een grondige studie gevoerd naar de wetgeving voor de

bescherming van persoonsgegevens op Europees (AVG) en Belgisch niveau. Het operationele aspect omvatte de ontwikkeling van een zakenmodel op basis van het Service-Dominant Business Model en het vastleggen van operationele procedures, waarbij rekening werd gehouden met de onderzoeksresultaten die tijdens PROMISE werden verkregen.

Conclusies

Het project PROMISE leverde een gedetailleerd rapport op met een stand van zaken inzake internationale “beste praktijken” voor webarchivering, rekening gehouden met wettelijke, operationele en technische aspecten. De voornaamste conclusies van deze projectfase werden ook gepubliceerd als onderzoeksartikel (Vlassenroot et al., 2019, zie <https://link.springer.com/article/10.1007/s42803-019-00007-7>).

Voor de definitie van een Belgisch, federaal webarchiveringsbeleid werd een rapport geschreven over het rechtskader van Belgische online informatie. Tevens werd een beleidsnota opgesteld voor de raad van bestuur van zowel het Rijksarchief als KBR. Deze nota omvatte de verschillende fasen in het proces van webarchivering (selectie, invoer, incorporatie, conservering, toegang ...) evenals een gedetailleerde kostenanalyse voor een functioneel webarchief op basis van realistische scenario's. Er werd ook een onderzoek gevoerd naar de vereisten van de gebruikers van webarchieven; de resultaten daarvan werden gepubliceerd in de vorm van een interactief dashboard (zie <https://public.tableau.com/profile/eveline.vlassenroot#!/vizhome/PReservingOnlineMultipleInformationtowardsaBelgianstrategy/PReservingOnlineMultipleInformationtowardsaBelgianstrategy>).

De pilootfase van het project behelsde de selectie van te archiveren webinhoud en van de daartoe vereiste instrumenten, de harvesting van die inhoud en het beheer van de toegang tot die collecties. Er werden zogenaamde seed lists (i.e. lijsten met relevante URL's) opgesteld voor het Rijksarchief en KBR. De websites op deze seed lists werden gecapteerd en opgeslagen met de open source software Heritrix, wat resulteerde in een collectie van WARC-bestanden. Het webarchief werd toegankelijk gemaakt met behulp van de open source software PyWB waarvan een instance werd geïnstalleerd op de servers van het Rijksarchief en KBR. De gecapteerde inhoud en de gebruikte instrumenten waren gedurende het project het voorwerp van een iteratieve en informele evaluatie. Dit omvatte een oefening ter evaluatie van de kwaliteit en de volledigheid van het gearchiveerde webmateriaal. In deze context werden antwoorden gezocht op onderzoeksvragen als (i) 'Welk percentage van de Belgische geschiedenis is verloren gegaan als gevolg van het ontbreken van een Belgisch webarchief?', (ii) 'Welke websites bleken bestand tegen de tand des tijds en staan nog steeds online?' en (iii) 'Hoeveel van het oude Belgische web kan worden gereconstrueerd via andere webarchieven of via 'webarcheologie technieken'?'. De resultaten werden (onder andere) gepubliceerd in de conferentie paper 'Unearthing the Belgian web of the 1990's: a digitised reconstruction' die werd voorgesteld ter gelegenheid van de 3de RESAW-conferentie 'The Web That Was: Archives, Traces and Reflections' (Amsterdam, 19-21 juni 2019). Voor andere evaluatieactiviteiten werd gebruik gemaakt van de personae die werden gecreëerd in het

Corpus-project en geschetst in de deliverable ‘Le projet Corpus et ses publics potentiels : Une étude prospective sur les besoins et les attentes des futurs usagers’ (zie <https://hal-bnf.archives-ouvertes.fr/hal-01739730/document>). Deze vijf personae, die aanvankelijk werden beschreven om te helpen met de identificatie van potentiële gebruikers, werden gebruikt ter evaluatie van actuele methoden van toegang tot, en analyse-instrumenten voor, webarchieven.

Op het vlak van aanbevelingen voor duurzame webarchivering werd een aantal acties ondernomen. Om te beginnen werden verschillende aanbevelingen geformuleerd om de Belgische wetgeving inzake het wettelijk depot te herzien en te wijzigen. Vervolgens werd een grondig rapport opgemaakt met juridische beschouwingen betreffende de toegang tot webarchieven. Bovendien werd een lijst met operationele vaak gestelde vragen (Frequently Asked Questions, FAQ) opgesteld die het Rijksarchief en KBR kunnen gebruiken als richtsnoeren voor de bescherming van persoonsgegevens in de context van webarchivering. Ten derde werden “besluitvormingsbomen” gecreëerd voor de evaluatie van copyright voor webarchivering in zowel de selectie- als de toegangsfasen. Ten vierde werd een lijst met FAQ inzake de bescherming van persoonsgegevens opgesteld om de werknemers van KBR en het Rijksarchief te helpen de uitdagingen in verband met webarchivering te begrijpen. Er werd een bedrijfsmodel opgesteld voor KBR en het Rijksarchief en er werden ook operationele procedures geschetst van de volledige werkstroom voor webarchivering.

Trefwoorden

- webarchivering
- digital humanities
- digitale duurzaamheid
- online informatie
- internet