

SAMENVATTING

Context, doelstellingen en methodologie

Het web is uitgegroeid tot een essentiële omgeving waarin hedendaagse sociale, culturele, politieke en wetenschappelijke ontwikkelingen worden vastgelegd. Websites, online nieuwsplatformen, blogs en sociale media vormen waardevolle bronnen voor het begrijpen van de recente geschiedenis, maar zijn tegelijk bijzonder kwetsbaar en kunnen snel verdwijnen. Hoewel er belangrijke vooruitgang werd geboekt op het vlak van webpreserving, blijven er grote uitdagingen bestaan inzake vindbaarheid, toegang, hergebruik, juridische conformiteit en duurzaamheid op lange termijn.

Het BelgicaWeb-project (2024–2026), gefinancierd door BELSPO, bracht KBR, de Universiteit Gent en de Universiteit Namen samen. Het project bouwde verder op de resultaten van de projecten PROMISE en BESOCIAL en had als doel het Belgische born-digital erfgoed beter vindbaar, toegankelijk, interoperabel en herbruikbaar (FAIR) te maken.

Het project bundelde de expertise van verschillende onderzoekscentra. Aan de Universiteit Gent stond mict in voor gebruikersonderzoek, stakeholderbetrokkenheid en participatief ontwerp. GhentCDH bracht expertise in rond digitale geesteswetenschappen en onderzoekstoepassingen, terwijl IDLab zich richtte op technische ontwikkeling en metadata-enrichment. Aan de Universiteit Namen leverde CRIDS expertise op het vlak van auteursrecht, gegevensbescherming, regelgeving rond artificiële intelligentie en bredere juridische en ethische vraagstukken. KBR coördineerde het project en leidde de activiteiten rond webarchivering, collectieontwikkeling en toegangsvoorzieningen.

BelgicaWeb had zes doelstellingen: het ontwikkelen van duurzame benaderingen voor toegang tot born-digital erfgoed; het uitbreiden van Belgische web- en sociale-mediacollecties; het verbeteren van de vindbaarheid door metadata-enrichment; het ontwerpen van een toegangsplatform en API; het bestuderen van juridische en ethische kaders; en het bevorderen van het bewustzijn en gebruik van born-digital erfgoedcollecties.

Het project hanteerde een interdisciplinaire en gebruikersgerichte aanpak waarin digitale preserving, digitale geesteswetenschappen, informatica, archiefwetenschap en juridisch onderzoek werden gecombineerd. Gebruikersvereisten werden verzameld via enquêtes, workshops en consultaties met onderzoekers, erfgoedprofessionals en webarchiveringsspecialisten. Twee adviesstructuren ondersteunden het project. Het Opvolgingscomité verstreekte periodiek strategisch advies en evalueerde de voortgang van het project, terwijl het Expert Panel een actieve co-creatie rol speelde bij de behoeftenanalyse, collectieontwikkeling, evaluatie van collecties, kwaliteitscontrole en de vormgeving van het toegangsplatform.

Resultaten, belangrijkste conclusies en aanbevelingen

BelgicaWeb toonde aan dat onderzoekers webarchieven steeds vaker beschouwen als onderzoeksinfrastructuren in plaats van louter bewaaromgevingen. Gebruikers benadrukten het belang van transparante collectieontwikkeling, rijke metadata, geavanceerde zoekfunctionaliteiten, downloadbare datasets en toegang via API's. Daarnaast wezen zij op de noodzaak om zowel verkennend als computationeel onderzoek te ondersteunen.

Het project ontwikkelde methodologieën voor het opbouwen van born-digital erfgoedcollecties, waaronder seedlijsten, harvestingworkflows en kwaliteitscontroleprocedures. Het Expert Panel leverde een belangrijke bijdrage aan de diversiteit en relevantie van de collecties. Browsertrix werd geselecteerd als belangrijkste harvestingoplossing vanwege het openbronkarakter en de ondersteuning van moderne webtechnologieën. Tegelijkertijd bevestigde het project de blijvende uitdagingen van sociale-media-archivering, onder meer door technische, juridische en platformgebonden beperkingen.

Een belangrijk resultaat was de ontwikkeling van een prototype van een toegangsplatform dat het ontdekken, doorzoeken en raadplegen van collecties voor verschillende gebruikersgroepen ondersteunt. BelgicaWeb ontwierp en implementeerde bovendien een API die geautomatiseerde toegang tot gearchiveerde inhoud en metadata mogelijk maakt. De technische infrastructuur werd ontwikkeld met openbrontechnologieën zoals Symfony en SolrWayback.

Daarnaast ontwikkelde het project enrichmentworkflows die bijkomende metadata genereren, waaronder kwaliteitsindicatoren, geografische informatie en auteursrechtgerelateerde kenmerken. Ook werden verschillende op artificiële intelligentie gebaseerde technieken onderzocht en geprototypeerd, waaronder taalidentificatie, thematische classificatie, named entity recognition, automatische samenvatting en beeldbeschrijving. Deze technieken kunnen de vindbaarheid en het hergebruik van webarchieven verbeteren wanneer ze transparant worden toegepast en gecombineerd worden met menselijke expertise.

Het juridisch onderzoek benadrukte de noodzaak om openheid en wetenschappelijk hergebruik te verzoenen met de naleving van regelgeving inzake auteursrecht en gegevensbescherming. Daarbij werd bijzondere aandacht besteed aan auteursrecht, wettelijk depot, privacy, text and data mining, licentiemodellen en opkomende regelgeving rond artificiële intelligentie. Een belangrijke conclusie was dat juridische en ethische overwegingen geïntegreerd moeten worden in de volledige archiveringscyclus.

BelgicaWeb toonde aan dat duurzame toegang niet uitsluitend afhankelijk is van technologie, maar ook van governance, institutioneel engagement, interdisciplinaire samenwerking en blijvende betrokkenheid van gebruikersgemeenschappen. Het project liet zien dat webarchieven evolueren naar onderzoeksinfrastructuren die ontdekking, analyse en hergebruik ondersteunen. De resultaten, waaronder gebruikersonderzoek, methodologieën voor collectieontwikkeling, experimentele collecties, een prototype van een toegangsplatform, een API, enrichmentworkflows en juridische aanbevelingen, vormen een stevige basis voor toekomstige toegang tot het Belgische web- en sociale-media-erfgoed.

Belangrijkste aanbevelingen

- Blijf webarchieven verder ontwikkelen als duurzame onderzoeksinfrastructuren.
- Behoud de actieve betrokkenheid van gebruikersgemeenschappen via co-creatie en iteratieve evaluatie.
- Investeer in metadata-enrichment en diensten voor computationele toegang, waaronder API's.
- Integreer juridische en ethische overwegingen in de volledige archiveringscyclus.
- Versterk institutionele samenwerking en governance op lange termijn.
- Blijf verantwoorde toepassingen van artificiële intelligentie voor ontsluiting en hergebruik van archieven onderzoeken.

Trefwoorden: Webarchivering; Born-Digital Erfgoed; Onderzoeksinfrastructuur; FAIR-data; Digitale Preservering