

ADOCHS

Controle van data geproduceerd door digitaliseringstechnologieën in de culturele erfgoedsector

DUUR	BUDGET
15/12/2015 – 31/12/2020	596.652 €

PROJECT BESCHRIJVING

In het midden van de Jaren negentig zijn de culturele instellingen het digitale tijdperk binnengetreden. In België keurde de regering in 2004 een eerste digitaliseringsplan goed voor een periode van tien jaar dat resulteerde in de realisatie van negen digitaliseringsprojecten in de federale wetenschappelijke instellingen. Om onvoorziene moeilijkheden te voorkomen vereisten deze projecten belangrijke menselijke en financiële hulpmiddelen. Daarom werd in 2014 een tweede fase gelanceerd die de instellingen toeliet om hun digitaliseringsactiviteiten van de vorige tien jaar verder te zetten. De expertise die werd opgebouwd in de eerste digitaliseringsfase kan ons helpen om toekomstige projecten beter uit te voeren. Het is in deze context dat we dit voorstel lanceren.

Het probleem van kwaliteitscontrole was één van de grootste hinderpalen in de eerste digitaliseringsfase. Het bleek dat veel projecten het belang hiervan, zowel op menselijk als op technisch vlak, hadden onderschat in het algehele digitaliseringsproces. In de meeste gevallen werden de projectmedewerkers geconfronteerd met een gebrek aan methodologische standaardisatie en geautomatiseerde hulpmiddelen. Daarom moest het werk dikwijls manueel uitgevoerd worden, zonder procedurele richtsnoeren aangepast aan de noden van het project. Het is duidelijk dat kwaliteitscontrole een essentieel deel uitmaakt van elke stap van een digitaliseringsproject om de integriteit en consistentie van de geproduceerde bestanden en gegevens te waarborgen en te bewaren voor de lange termijn. Dit is zowel het geval voor externe als voor interne digitaliseringsprojecten. Door werk te maken van kwaliteitscontrole wil het project het gehele digitaliseringsproces bespoedigen en tegelijk de kosten minimaliseren. Het beoogt ook de waarde van de geproduceerde gegevens te verhogen voor op toekomstige projecten.



Hoewel de federale wetenschappelijke instellingen de voornaamste begunstigen zijn beoogt dit onderzoek tegemoet te komen aan jarenlange noden die alle erfgoedinstellingen in België en het buitenland betreffen. De bedoeling is om het probleem in twee stappen aan te pakken. Ten eerste door te aandacht te vestigen op methodologische aspecten en door richtlijnen voor te stellen die kunnen toegepast worden bij de behandeling van erfgoedcollecties. Vervolgens kunnen technische hulpmiddelen ontwikkeld worden die taken met betrekking tot kwaliteitscontrole kunnen automatiseren of de manuele controle vergemakkelijken. Er zal zowel rekening gehouden worden met technische aspecten (beeldresolutie, integriteit van het bestandsformaat) als met de inhoud van de metadata (beschrijvingen van de collecties, overeenstemming met XML schema's, enz.). Naast de internationale wetenschappelijke literatuur kunnen onderzoekers beroep doen op de expertise verworven door de verschillende instellingen tijdens de eerste fase van de digitalisering maar ook op de noden van de nieuwe projecten.

ADOCHS

In deze context werken de onderzoekers op twee soorten collecties: de gedigitaliseerde collecties van de Koninklijke Bibliotheek van België over problemen met tekstdocumenten en de fotocollecties van het CegeSoma die een groep iconografische documenten betreffen. Hoewel het hier over heel verschillende collecties gaat is het scanproces vergelijkbaar zodat we kunnen stellen dat de volgende algemene kwaliteitsproblemen voorkwamen in de eerste fase van de digitaliseringscampagne en opnieuw zullen opduiken in de workflow indien er geen procedure voor kwaliteitscontrole wordt ontwikkeld, en dus zullen ze in de context van dit project aangepakt worden:

- onvolledige scans (ontbrekende pagina's of een ontbrekend deel van een pagina)
- fouten in de ordening van de scans
- kleur i.p.v. grijswaarden en omgekeerd
- veranderingen in de resolutie of het bestandstype
- vormfouten bij de manuele metadatatrascriptie en codering
- gestructureerde metadata kunnen niet geproduceerd worden
- onscherpe beelden (heel het document of een deel er van)
- niet-uniforme kleurweergave
- slechte of ongekropte beelden
- OCR_kwaliteit (fouten bij het converteren van de tekst en detectiezone, bv. titel, uitgavenummer enz.)

Deze fouten kunnen in twee categorieën worden onderverdeeld: 1) menselijke fouten en 2) fouten die voortkomen uit ontoereikende software output. In het laatste geval kan er sprake zijn van een combinatie van a) opnieuw menselijke fouten, b) beperkingen in de gebruikte software of c) materiële degradatie van het voorwerp zelf. Daarom beperken we de vermelde kwaliteitsproblemen naar gelang hun herkomst.

Een VUB PhD student van het Department of Electronics and Informatics (ETRO) is verantwoordelijk voor de kwaliteitscontrole van de beelden (deeltijds aangesteld door de VUB en deeltijds door de KBB). Een ULB PhD student is verantwoordelijk voor de kwaliteitscontrole van de metadata (deeltijds aangesteld door de ULB en deeltijds door het CegeSoma). Beide worden deeltijds aangeworven door de universiteit en deeltijds door de instellingen omdat het wetenschappelijk werk verband houdt met de problemen die in situ voorkomen. De onderzoeksresultaten worden dan ook regelmatig geconfronteerd met case studies.

Twee extra onderzoekers zijn voorzien voor een periode van één jaar elk: één bij het begin van het project en één bij het einde. De eerste onderzoeker zal een inventaris opstellen van goede praktijken in procedures voor kwaliteitscontrole, zowel bij digitale beelden als metadata (zie taakbeschrijving). De tweede zal verantwoordelijk zijn voor de integratie van de onderzoeksresultaten aan het einde van het project zodat een duidelijke procedure voor kwaliteitscontrole kan vastgelegd worden die in de erfgoedinstellingen gevolgd kan worden.



CONTACT INFORMATIE

Coördinator

Florence, Gillet
Belgisch Rijksarchief - CegeSoma
florence.gillet@arch.

Partners

Lemmers Frederic
Koninklijke Bibliotheek van België
frederic.lemmers@kbr.be

Ann Dooms
Vlaamse Universiteit Brussel (VUB)
Department of Electronics and Informatics (ETRO)
Multimedia Forensics Team
ann.dooms@vub.ac.be

Seth Van Hooland
Universit  Libre de Bruxelles (ULB)
Department of Sciences and Technologies of Information
and Communication
svhoolan@ulb.ac.be

LINKS

<http://adochs.be/>

