

Digitaliseren in de heemkundige kring

We willen onze fotocollectie op het net zetten, we willen onze dure boeken digitaliseren zodat we het origineel niet meer ter beschikking moeten stellen in onze bibliotheek, de kwaliteit van onze videocassettes gaat zienderogen achteruit ... er zijn veel redenen te verzinnen om met het digitaliseren van ons archief te beginnen. Maar hoe beginnen we daar nu aan? En waar moeten we op letten? Welke hard- en software hebben we nodig? Allemaal vragen waarmee we steeds meer geconfronteerd worden in onze heemkundige kring.

Digitaliseren of digitaal archiveren?

Digitaliseren is het omzetten van data van een analoog naar een digitaal medium. Voorbeelden zijn het scannen van oude foto's, het overzetten van muziek van grammofoonplaten naar je pc of video van VHS naar DVD.

Digitaal archiveren is niet exact hetzelfde als digitaliseren. Digitaal archiveren is het bewaren van digitale bestanden voor een lange termijn. Het gaat hier zowel om gedigitaliseerde bestanden als om digital born bestanden. Digital born bestanden zijn bestanden die van bij hun 'geboorte' digitaal zijn, bv. een tekst die je in word hebt geschreven of een foto van je digitale camera.

Wanneer we gaan digitaliseren volgen we best een vastgelegd stappenplan. Eerst en vooral gaan we een goede planning opmaken. Daarnaast zorgen we voor een goed uitgewerkte mappenstructuur waar we onze digitale bestanden gaan onderbrengen. Deze twee stappen voeren we best uit vóór we effectief gaan digitaliseren. Tijdens en na het digitaliseren zelf voegen we metadata toe aan onze digitale bestanden. Tot slot zorgen we er ook voor dat onze digitale bestanden op een correcte manier bewaard worden.

Planning

Eén van de belangrijkste stappen in het digitaliseringsproces is het opmaken van een goede plan-

ning. Nochtans wordt deze stap vaak overgeslagen, en begint men op goed geluk met het digitaliseren van de eerste stukken. Uiteindelijk eindigt men met een bont allegaartje digitale bestanden.

Om dit te voorkomen is het belangrijk om vóór de start van het proces een aantal vragen te stellen:

Wat wil ik digitaliseren?

Breng in kaart welk soort documenten je precies wilt digitaliseren. Wil je enkel de muziek van je 78-toerenplaten overzetten, of wil je de hoes ook mee digitaliseren? Bestaat je videocollectie enkel uit VHS-cassettes, of zitten er ook andere formaten (bv. Betamax) bij? Allemaal vragen die best gesteld worden voor aanvang van het digitaliseringsproces. Zo kan je meteen nagaan of je alle expertise en middelen in huis hebt om deze documenten te digitaliseren.

Hoe wil ik het digitaliseren?

Doe ik het zelf of laat ik het doen? Als ik het zelf doe moet ik nagaan of ik alle know-how en materiaal in huis heb? Is dat niet het geval moet ik nagaan of ik die alsnog kan binnenhalen?

Ook wie zijn materialen laat digitaliseren door een firma of instelling, moet een goede planning maken. Door het maken van een goed lastenboek zorgt hij er voor dat de externe partner weet wat hij moet aanleveren. Ben je als klant niet tevreden over het eindproduct, kan je op je lastenboek terugvallen bij een dispuut over de kwaliteit van de geleverde bestanden.

Waarom wil ik het digitaliseren?

Misschien wel de belangrijkste vraag die men zich moet stellen. Een foto scannen voor langetermijnbewaring vergt een heel andere aanpak dan een foto scannen om een affiche te maken voor je jaarlijkse tentoonstelling.

Ga je voor langetermijnbewaring, dan moet je werken volgens de geldende open standaarden, een de-

gelijke mappenstructuur uitwerken (zeker voor een grote groep documenten), de digitale bestanden op een correcte manier beschrijven, en tenslotte alles op een zo duurzaam mogelijke manier bewaren. Alles bij elkaar is dit de meest tijdrovende manier van digitaliseren, maar ook de meest duurzame, en een ideaal vertrekpunt voor andere doelstellingen. Zo kan je van een bestand dat aangemaakt is voor langetermijnbewaring gerust een kopie maken om te gebruiken voor reproductie in een boek of om het online te delen.

Is het niet noodzakelijk om de documenten voor een lange termijn te bewaren, dan kun je een aantal stappen overslaan. Wil je bv. een reproductie van een foto maken voor een boek, dan is het voldoende om de foto op een redelijke resolutie (bv. 300 ppi) in te scannen. Werken volgens open standaarden is dan niet noodzakelijk. Je moet er gewoon voor zorgen dat de drukker je bestanden kan openen. Verder hoeft je geen aandacht te schenken aan een mappenstructuur of uitgebreide beschrijving. Wanneer je dus werkt volgens het procédé van langetermijnbewaring kan je je bestanden bijna voor alle andere doeleinden gebruiken, behalve wanneer je een reproductie op groot formaat wil maken. De aanbeveling voor langetermijnbewaring is om foto's met een resolutie van 300 ppi (pixels per inch) in te scannen. Wil je echter een affiche of foto op groot formaat maken, dan zal je de bestanden op 600 ppi en meer moeten inscannen. Waarom dan niet meteen alles inscannen op 600 ppi of meer? Omdat je bestanden meteen een heel stuk zwaarder zijn en je dus veel meer opslagruimte nodig hebt. Daarom hebben archivariissen bepaald dat 300 ppi de minimale norm is.

Zijn er mogelijke partners?

Digitaliseren is een dure zaak. Het vraagt eerst en vooral veel tijd. Maar ook middelen. Denk maar aan een degelijke scanner, pc, software, ... Voor je begint met een grootschalig digitaal archiveringsproces is het dan ook aan te raden om informatie in te winnen bij mogelijke partners.

Eerst en vooral moet je nagaan of de documenten nog niet gedigitaliseerd zijn. In dat geval is het misschien mogelijk om de gedigitaliseerde bestanden gratis of tegen een redelijke prijs te verkrijgen.

Is dat niet het geval kan samenwerken toch nog voordelen opleveren. Zijn er partners waar je hard-

ware (bv. een boekscanner) of software kan lenen? Of om deze dure aankopen gezamenlijk aan te kopen? Misschien kom je door samenwerking wél in aanmerking voor subsidies waar je anders niet voor in aanmerking komt? Allemaal redenen om op zoek te gaan naar één of meerdere goede partners.

Documenteren

Tot slot is het ook belangrijk dat men elke stap in het proces documenteert. Eerst en vooral voor de medewerker(s) aan het digitaliseringsproces zelf. Door een goede documentatie voorkomt men na verloop van tijd vergeten te zijn waarom men bepaalde stappen heeft gezet, of hoe ze tot een bepaalde werkwijze zijn gekomen. Maar ook voor latere gebruikers is het belangrijk te weten welke stappen je hebt ondernomen in je digitaliseringsproces. Heb je foto's opgesmukt (bv. krassen verwijderd), stukken video van slechte kwaliteit weggeknipt, ... allemaal informatie die voor een toekomstige gebruiker van belang kan zijn. Wanneer men om puur archivistische redenen gaat archiveren is het trouwens aan te raden om zo dicht mogelijk bij het origineel te blijven.

De documentatie bewaren we ook volgens de geldende open standaarden. We bewaren deze informatie dus niet in word, maar in xml. De documentatie schrijven we ook best telkens mee naar het opslagmedium waar we onze gedigitaliseerde bestanden bewaren.

Mappenstructuur opbouwen

Wanneer we een groot aantal documenten willen digitaliseren om ze op lange termijn te bewaren, dan is het aan te raden om eerst een goede mappenstructuur uit te werken. Een goede mappenstructuur is logisch opgebouwd. Bestanden die samen horen, worden bewaard binnen dezelfde map. Voor meer informatie over het opzetten verwijs ik graag naar de website van eDavid (<http://www.edavid.be/david-project/teksten/Richtlijn3.pdf>).

Digitaliseren

Hardware

Aangeven welke toestellen het best geschikt zijn om te gaan digitaliseren, is afhankelijk van veel factoren: je budget, je doelstellingen, je kennis, ... Het is onmogelijk om in het bestek van dit artikel alle soorten toestellen te overlopen. Toch wil ik een aantal aspecten aangeven waar je best op let.

Voor het scannen van foto's, dia's of fotofilms is een flatbedscanner een betaalbaar en geschikt toestel. Voor dia's en films bestaan er ook aparte toestellen, maar die zijn duurder in aankoop. Waar moet je op letten bij de aankoop van een scanner: hij moet minstens een resolutie van 300 ppi aankunnen (voor dia's zelfs 1200 ppi). Wil je kleuren scannen, gebruik je best een scanner die een kleurdiepte van minimaal 24bit/pixel aankan.

Allroundtoestellen waarmee je ook kan printen, faxen, ... bieden over het algemeen minder kwalitatieve scans. De digitale revolutie gaat echter zo snel dat ook dat misschien wel verandert in de toekomst.

Wil je dure boeken gaan scannen, dan is het gebruik van een flatbedscanner sterk af te raden. De lichtintensiteit die vrijkomt tijdens het scanproces is namelijk nefast voor je papier. Dure boeken worden het best ingescand met een 'dure' boekscanner. Heb je geen mogelijkheid om die te gebruiken, dan kan je al veel bereiken met een semi-professioneel fototoestel en een lichtsterke lens. Wanneer je gaat digitaliseren met een fototoestel, let er dan op dat de hoek tussen het object en de lens 90° graden is, en de externe lichtinval egaal verdeeld is. Gebruik in geen geval een flits, want die zorgt opnieuw voor een heel schadelijke lichtstraal (om die reden mogen we geen foto's nemen in musea). Overigens is een digitale camera enorm geschikt om grote formaten (bv. charters, kaarten, vlaggen ...) te gaan digitaliseren.

Voor het digitaliseren van audio bestaan er tegenwoordig platen- of cassettespelers met een usb-uitgang. Indien je een geluidskaart in je pc hebt kan je die dan op je pc aansluiten. Via Audacity kan je dan de audio op je pc opnemen.

Voor het digitaliseren van video zijn er twee manieren. Ofwel sluiten we onze videorecorder aan op onze dvd-recorder en nemen we op via de tv. Tweede mogelijkheid is om onze videorecorder aan te sluiten op onze grafische kaart van onze pc. De grafische kaart moet dan wel een analoge video-ingang hebben. Heeft de grafische kaart dit niet, dan moeten we een converter gebruiken om het analoge signaal in een digitaal signaal om te zetten. Daarna nemen we de videobestanden op met behulp van een videobewerkingsprogramma. Voor

digitaal archiveren is de tweede manier de beste, omdat je meer controle hebt over je proces. Je kan meer bestandsformaten kiezen, en metadata toevoegen.

Software

Om te digitaliseren is heel wat gratis software beschikbaar. In Gimp kan je foto's bewerken. XnView en IrfanView kunnen gebruikt worden om je scanner mee aan te sturen. Voor audiobestanden kan men Audacity gebruiken.

Gratis software bevat meestal een groot deel van de functionaliteiten van betalende software. Betalende software daarentegen is vaak gebruiksvriendelijker. Voorbeelden van betalende software die men tijdens het digitaliseringsproces kan gebruiken zijn Adobe Photoshop (fotobewerkingsprogramma), Adobe Lightroom (fotobeheerprogramma), Apple Aperture (fotobeheerprogramma), Adobe Premiere (video), Nero (video) ... Betalende software hebben wel een bijkomend nadeel ten opzichte van open sourceprogramma's: ze werken niet altijd volgens de open standaarden. Wie foto's beschrijft in Adobeprogramma's, zal een deel van zijn metadata verliezen wanneer hij de foto in een niet-Adobeprogramma opent. Je bent dus afhankelijk van één producent en hebt meer kans dat je informatie op lange termijn niet (helemaal) leesbaar is.

Proces

Wanneer we gedigitaliseerde bestanden opslaan, doen we dat best in open en gestandaardiseerde bestandsformaten. Open bestandsformaten zijn voor iedereen beschikbaar en niet afhankelijk van één enkele leverancier. Gestandaardiseerde bestandsformaten zijn bestandsformaten die wereldwijd bekend zijn, en door veel mensen worden gebruikt. Tot slot is het voor langetermijnbewaring ook aan te raden bestandsformaten te gebruiken die niet vertrekken van een compressie. Om die reden raden archivarissen het gebruik van jpeg en mp3 af.

Welke bestandsformaten zijn dan allemaal geschikt? Voor tekstbestanden gebruiken we best het xml- of pdf-bestandsformaat. Foto's worden best in tiff of jpeg2000 opgeslagen. Voor audio gebruiken we best het wav-formaat. Voor video gebruiken we MXF, AAF, AVI of MOV voor de wrapper en MJPEG2000 of MPEG-2 voor de videocodecs .

Beschrijven

Wanneer we een groot aantal documenten gaan digitaliseren, is het belangrijk dat we die later makkelijk kunnen terugvinden. Dat kan enerzijds door een goede mappenstructuur (zie hoger), en anderzijds door het toevoegen van metadata aan onze bestanden. Metadata zeggen iets meer over je bestanden. Metadata zijn eigenlijk data over data. Sommige metadata kunnen we automatisch laten genereren (bv. de datum en het uur waarop we foto's nemen met onze digitale camera). Andere metadata moeten we zelf toevoegen. Voorbeelden van metadata die we zelf moeten invoeren zijn trefwoorden. In de meeste programma's kun je metadata toevoegen (bv. onder bestandsinformatie). Helaas werken heel wat programma's met een eigen metadataschema. Wanneer die bestanden in een ander programma geopend worden, zijn niet alle metadata leesbaar. Microsoft- of Adobeprogramma's werken vaak met een eigen metadataschema. Voor het digitaal archiveren is het echter noodzakelijk om metadata te gaan gebruiken die wel door alle andere programma's kunnen worden gelezen. Men test aan het begin van een beschrijvingsproces

du best even welke metadata door verschillende programma's kunnen worden gelezen.

Bewaren

Al tijdens ons digitaliseringsproces is het noodzakelijk om de documenten op een goede manier te bewaren. We willen namelijk niet dat al ons werk teniet wordt gedaan wanneer onze pc crasht of een andere calamiteit zich voordoet.

Daarom is het interessant om geregeld een backup te maken van onze gedigitaliseerde bestanden. Tijdens het proces gebruiken we hiervoor best een externe harde schijf. We gebruiken geen backup-software. Die maakt namelijk een compressie van onze bestanden. Bij decompressie is er telkens een klein beetje dataverlies, waardoor onze documenten sneller onleesbaar worden. We maken dus een exacte kopie van onze bestanden op deze externe harde schijf. Deze externe harde schijf bewaren we best op een andere plaats dan waar onze pc staat. Anders ben je bij brand of diefstal beide opslagmedia kwijt.



Wanneer we klaar zijn met ons digitaliseringsproces, kunnen we de bestanden ook gaan overschrijven op CD-R, DVD-R of zelfs Blu-Ray. We moeten hier enkele vuistregels in acht nemen:

- We branden de schijfjes zo traag mogelijk. Hoe trager het brandingsproces, hoe beter.
- We gebruiken het best schijfjes met een gouden laag (de zogenaamde 'golden' CD-R of DVD-R). We gebruiken zeker geen CD-RW of DVD-RW. RW staat voor rewritable of herbeschrijfbaar.
- We schrijven niks op het schijfje zelf, zelfs niet met daarvoor ontwikkelde stiften. Wanneer we iets schrijven op het schijfje maken we namelijk kleine krasjes in de bovenlaag. De informatie bevindt zich slechts enkele nanometer onder die bovenlaag. De krasjes zorgen dus ook voor data-verlies.
- We kleven ook niks op het schijfje. Wanneer de klever niet volledig perfect hangt (en dat is bij 99,99% van de gevallen zo), zal het schijfje niet perfect draaien in de lezer. Dat zal op termijn voor schade aan het schijfje zorgen.
- We bewaren de doosjes rechttop in een kast. We zetten niks op de doosjes.

Online bewaren is ook een mogelijkheid, al zou ik dit afraden als eerste of zelfs tweede opslagmedium. Wanneer we online bewaren, geven we de controle over onze bestanden namelijk uit handen. Gaat de provider morgen failliet, dan kunnen we misschien niet meer aan onze bestanden.

Nog minder te gebruiken als bewaarplaats voor onze digitale bestanden zijn web 2.0-toepassingen

Interessante websites

Algemene informatie

<http://www.projectcest.be>

<http://www.edavid.be>

<http://www.den.nl/debasis>

<http://www.jiscdigitalmedia.ac.uk>

Gratis programma's

GIMP: <http://www.gimp.org>

XnView: <http://www.xnview.com/en/features.html>

Irfanview: <http://www.irfanview.com>

Audacity: <http://audacity.sourceforge.net/?lang=nl>

Md5Summer: <http://www.md5summer.org>

als Flickr (foto + film), Picassa (foto), Youtube (film), google docs (documenten), ... Wanneer je documenten naar de toepassingen oplaadt, gaan deze ingrijpen op de documenten. Meestal zullen de bestanden worden gecomprimeerd. Daarnaast zullen ook andere zaken aangepast worden, zoals de naam van het document of de extensie (Flickr bv. zet tiff automatisch om in jpeg). Deze toepassingen kunnen wel interessant zijn om onze documenten aan de wereld te tonen.

Digitaal archiveren is een relatief jonge discipline. Daardoor is er nog veel onduidelijkheid over de levensduur van opslagmedia. In de handel zijn CD-R's verkrijgbaar die garanderen dat je informatie 300 jaar leesbaar zal zijn. Over 290 jaar zullen we kunnen zeggen of dat effectief zo is. Ook harde schijven verliezen na verloop van tijd hun data. Het is dus aan te raden om kritisch om te springen met je digitale opslagmedia, en geregeld een controle uit te voeren van je bestanden. Dat kan met behulp van een programma om checksums te berekenen, zoals md5summer. Dat programmaatje berekent de exacte opeenvolging van ééntjes en nulletjes in een bestand. Wanneer het bestand onleesbaar begint te worden (bv. door degradatie van de drager), dan zal deze checksum veranderen. Wanneer dat voorkomt, kan je ingrijpen door een nieuwe kopie te maken.

Dit artikel beschrijft in een notendop hoe je van start gaat met het digitaliseren van je heemkundige collectie. Digitaliseren is echter een complexe materie. Wil je dus zelf aan de slag, dan is het noodzakelijk zich nog verder te documenteren. Je vindt hiervoor een aantal interessante websites bij dit artikel.

Francis Vlieghe

Hoofdmedewerker Bibliotheek Hogeschool Gent