

Hoofdstuk 4 · Het wetenschappelijk gebruik van digitale archieven

YANA-FRAUKE VANDENDRIESSCHE, LIESBETH VAN MELLE, INGE VAN NIEUWERBURGH
(UNIVERSITEITSBIBLIOTHEEK GENT)

Een wetenschapper die archieven consulteert, moet soms creatief zijn om het waarheidsgehalte van een bron te onderzoeken. Hij heeft daarvoor een aantal werkmiddelen voorhanden, maar het is niet steeds evident. Veel hangt af van wat bewaard is gebleven, de context van archiefstukken en het beleid van het archief. Bij een digitaal archief is de situatie vaak nog wat complexer. Daarbij is de gebruiker ongetwijfeld een belangrijke schakel in de hele opzet ervan. Zonder gebruikers, nu of in de toekomst, lijkt een langetermijnarchief immers volstrekt zinloos. Bovendien is het noodzakelijk een continue wisselwerking tussen instelling, archief en gebruiker te onderhouden aangezien gebruikers en hun noden, naargelang de context, de periode en de evoluerende (technologische) mogelijkheden, voortdurend veranderen.

Een multimediaal langetermijnarchief zal ook een wetenschappelijk interessant archief zijn aangezien het kan fungeren als een rijke en waardevolle bron voor wetenschappelijk onderzoek. Inzicht in de wetenschappelijke gebruikersnoden helpt dan ook de kwaliteit van het archief te verhogen en de toekomstige interpretatie ervan te vergemakkelijken. Aan welke eisen moet een digitaal archief aldus voldoen om die noden van 'wetenschappelijkheid' en 'betrouwbaarheid' in te willigen? Volgens Dame Lynne Brindley, CEO van The British Library, zijn de drie belangrijkste eisen in de context van digitale langetermijnbewaring: *Trusted*, *Safe* en *Easy Accessible*. Bovendien moet, bij het deponeren van data in een archief, het doelpubliek, de zogenaamde *Designated Community*, steevast goed voor ogen gehouden worden. In de lijn van Brindleys uitspraak, formuleren we drie criteria waaraan het werk en zijn auteur, het digitaal archief of de bewaarplaats en de wetenschappelijk gebruiker moeten beantwoorden om als *wetenschappelijk* getypeerd te worden: kwaliteit (*safe and trusted*), toegankelijkheid (*easy accessible*) en een welomschreven doelpubliek (*designated community*).

Er wordt in deze bijdrage uitgegaan van een meer genuanceerde en ruime definitie van *wetenschappelijkheid*. *Wetenschappelijk* zijn die data, of is de methode van een

auteur of gebruiker, indien er in de eerste plaats een aanwijsbare garantie voor *kwaliteit* is. *Wetenschappelijk* is in deze context dus niet gelijk aan *academisch*².

Dit hoofdstuk behandelt het *wetenschappelijk gebruik* van een digitaal archief en de criteria waaraan zo een archief moet voldoen om wetenschappelijk gebruik toe te laten. Allerlei soorten archieven kunnen aan de criteria voldoen, dus niet enkel het wetenschappelijk *repository*, waarin alleen data zijn opgeslagen die uit academisch onderzoek voortvloeien.

1. *Kwalitatieve content*

De wetenschappelijke integriteit van digitale data is het resultaat van rigoureuze kwaliteitscontrole op drie niveaus: bij de aanlevering door de dataproducent of auteur, tijdens het beheer en de opslag in het archief en bij de omgang met en het (her)gebruik van de data door de gebruiker. Deze drie initiële concepten of niveaus zijn ontleend aan het OAIS-model (Open Archival Information System)³, een ISO-standaard die een uitgelezen conceptueel raamwerk biedt om de kwaliteit van een digitale bewaarplaats grondig aan te toetsen. Vandaag vormt het OAIS een referentiemodel voor heel wat digitale archieven.

DE AUTEUR EN DE INHOUDELIJKE KWALITEIT VAN DE BRON

De kwaliteit van de instroom bepaalt in hoge mate de kwaliteit van het archief en het gebruik. Onderscheid tussen data die in een wetenschappelijke context zijn ontstaan en data die met een ander doel gecreëerd zijn, dient dus best bewaard te blijven. Terwijl kwaliteit bij de eerste categorie zowel inhoudelijk als vormelijk en technisch aan een aantal bijzondere eisen moet voldoen, kan men de kwaliteit van de tweede categorie eigenlijk enkel op het formele en technische vlak meten. Dit komt later aan bod in de context van technische kwaliteit. Hier komt vooral de wetenschappelijke bron aan bod.

Er bestaan enkele parameters waaraan een wetenschappelijk werk inhoudelijk moet voldoen om geloofwaardig en kwalitatief te zijn. Het oordeel van experts en collega-onderzoekers binnen een vakgebied is alvast een bepalende factor. De meest courante methode is *peer review*, de beoordeling door 'gelijken', wetenschappers die werkzaam zijn in dezelfde discipline. Deze methode is dus kenmerkend

voor de evaluatie van wetenschappelijk tekstmateriaal en lijkt misschien in de context van een digitaal archief minder van toepassing. *Peer review* of daarop geïnspireerde principes zullen echter steeds aanzienlijker worden in digitale omgevingen, ook met betrekking tot audiovisueel materiaal. Het is immers niet ondenkbaar dat wetenschappers niet enkel elkaars publicaties (in welke vorm dan ook) beoordelen maar ook complete datasets. Belangrijke kwaliteitscriteria zullen dan zijn: wie is de auteur? Bij welke instellingen is de auteur betrokken? Met welk onderzoek, welke publicaties, welke data kan men hem in verband brengen? In welk vakgebied liggen zijn onderzoeksinteresses? Wat kan men afleiden uit de persoonlijke bibliografie van deze auteur, zijn referenties naar andere auteurs (met hun respectieve wetenschappelijke uitstraling, aangetoonde expertise in het vakgebied)? De wetenschappelijke gebruiker kan dan zelf oordelen of deze informatie voldoet aan de wetenschappelijke eisen van zijn vakgebied.

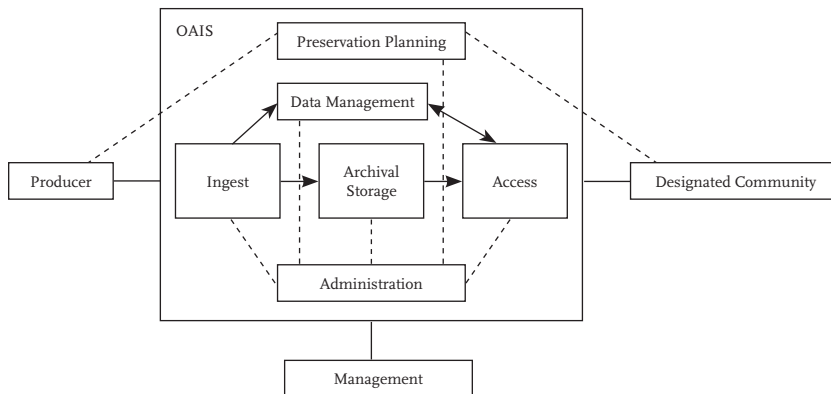
Citatieanalyse gaat dan weer in op de vraag wat de impact van een onderzoek is. Hoe vaak wordt naar dit onderzoek verwezen? Door wie wordt deze auteur geciteerd? Wat is de wetenschappelijke uitstraling van de bron? Er wordt met andere woorden 'gemeten' hoe belangrijk het onderzoek is voor het vakgebied. Deze methode is echter binnen de sociale en humane wetenschappen moeilijker toe te passen omdat men er andere publicatiemethodes toepast dan bij STM (Science, Technology en Medicine). Niettemin kunnen experts eventueel zelf een lijst tijdschriften of bronnen opstellen die zij voor een bepaald vakgebied als relevante maatstaf beschouwen voor hun oordeel over de wetenschappelijkheid van een document. In het digitale en online tijdperk is ook een digitale pendant van citatieanalyse ontstaan. Zo geldt het aantal downloads van of hyperlinks naar een elektronische bron als een parameter voor de zichtbaarheid, de relevantie en aldus de kwaliteit van het onderzoek. Maar het is duidelijk dat ook hier het aantonen van objectiviteit problematisch is geworden en met de nodige omzichtigheid benaderd moet worden. Een combinatie van de verschillende methodes is dan ook aan te raden.

Behalve de wetenschappelijke achtergrond van de auteur is het ook van belang of de gevolgde (wetenschappelijke) methode in het werk duidelijk verantwoord is en of de relevantie van het onderzoek binnen het vakgebied kan worden aangetoond. Ook moet men beoordelen of de specifieke richtlijnen of standaarden, met het oog op bijvoorbeeld digitalisering, dataverzameling, presentatie van onderzoeksgegevens, die in het betreffende vakgebied gelden, goed zijn nagevolgd. Met deze laatst genoemde criteria wordt al vooruitgewezen naar de vormelijke kwaliteitsvoorwaarden van een onderzoek. Zo zal de noodzaak aan voldoende metadata een belangrijke formele en technische voorwaarde zijn voor de waarborg van wetenschappelijkheid op lange termijn.

HET ARCHIEF

Het archief zal instaan voor het behoud van kwaliteit van de data op lange termijn. Noodzakelijkerwijs moet het digitaal archief zelf zijn bestaan op lange termijn kunnen garanderen. Het moet bijgevolg de geschikte financiële, organisatorische en juridische omkadering hebben om het beheer, onderhoud en de verbetering van het systeem en de data op lange termijn te blijven garanderen. Continuïteit is dus een basisvereiste. Een welomschreven beleidsplan is daarbij onontbeerlijk. De nodige maatregelen voor duurzame archivering moeten nauwkeurig onderzocht en uitgevoerd worden en er zijn criteria en voorschriften nodig om de aanlevering van digitale data correct en uniform te laten verlopen. In een dergelijk beleidsplan rond digitale bewaring moet rekening gehouden worden met de voortdurende evoluties en (internationale) onderzoeken rond digitale langetermijnbewaring. Enkele aandachtspunten zijn ongetwijfeld de aansluiting bij en keuze voor standaarden, gestandaardiseerde en overwogen ontsluitingswijzen, de blijvende beschikbaarheid van elektronische bronnen, onderbouwde selectiemechanismen, regelmatige gebruikersonderzoeken, het onderhoud van de technische inrichtingen en aandacht voor de kwetsbaarheid van dragers en formaten⁴, rechtenbeheer, en zo meer. De beleidskeuzes uit het beleidsplan moeten weerspiegeld zijn in de metadata: welke (sectorspecifieke) standaarden worden aanbevolen? Hoe zit het met rechten? Welke technische metadata zijn noodzakelijk? Ook met het oog op zoekmodaliteiten en een goed begrip van de data zijn metadata onontbeerlijk. Bovendien is het een essentiële taak van het digitaal archief om aan de gebruikers de optie aan te bieden zelf metadata toe te voegen, maar dit uiteraard volgens bijzondere richtlijnen.

Al deze overwegingen hangen nauw samen met de kwaliteit van de technische infrastructuur van het archief. Voor DANS, het Nederlandse instituut voor 'Data Archiving and Networked Services'⁵, is een 'databewaarplaats' kwalitatief en betrouwbaar (zie de oorspronkelijke term *Trusted Digital Repository*) indien ze expliciet de gedefinieerde taken en functies van het OAIS-referentiemodel kan vervullen⁶. Hier staat de kwaliteit van het 'management' of beheer van het archief centraal.



FIGUUR 1: OAIS model

Het OAIS vigeert als een functioneel model (zie figuur 1) waarvan de kern de 'archival storage' is. OAIS geeft namelijk aan hoe de informatie, de data en hun metadata, als informatiepakketten opgeslagen moeten worden en welke gegevens noodzakelijk zijn om de bewaring ervan te garanderen. Een onderdeel van de informatie is beschrijvende informatie. Daarin zijn onder meer gegevens te vinden met betrekking tot de 'provenance', de historiek van het dataobject. Het gaat dan over de registratie van de laatste wijzigingen of van welke software en hardware met respectieve instellingen voor de creatie van de data gebruik werd gemaakt, unieke identificatiecodes zoals ISBN's, DOI's of PURL* maar ook de bibliografische referenties zoals titel en auteur.

TECHNISCHE KWALITEIT

Behalve de garantie op een kwalitatieve inhoud, doet ook de kwaliteit van de vorm waarin de data geleverd en opgeslagen worden, er toe. Met het oog op preservering is de technische kwaliteit cruciaal wegens het gevaar op veroudering van digitale dragers. De inhoud mag dan nog zo kwalitatief bevonden zijn, als de data binnen enkele decennia niet meer bekeken of beluisterd kunnen worden, is de langetermijnbewaring ervan mislukt.

Aandacht voor geschikte opslagformaten is zowel een taak van de auteur als van het archief. Door de keuze voor de juiste opslagformaten staat het archief ervoor

* Respectievelijk 'International Standard Book Number', 'Digital Object Identifier' (uniek id voor een digitaal document) en 'Persistent Uniform Resource Locator'.

in dat de formaten leesbaar blijven en dat zo de technische infrastructuur goed onderhouden en beheerd wordt. Ten slotte moet de toekomstige eindgebruiker weten over welke applicaties hij dient te beschikken om de gewenste data te kunnen consulteren en te gebruiken. Het is de taak van het archief om aan de auteur of producent van data opslagformaten aan te bevelen. In die aanbevelingen wordt rekening gehouden met de specifieke kenmerken van de formaten, hun kans op veroudering en oplossingen om de leesbaarheid te blijven garanderen. Idealiter wordt geopteerd voor producentonafhankelijke bestandsformaten. Documentatie over de gebruikte formaten, software- en hardware-instellingen moet de eindgebruiker informeren over het ontstaansproces van de digitale data. Het is daarom van groot belang voldoende *technische metadata* met de data mee te leveren met het oog op blijvende leesbaarheid en bruikbaarheid.

KWALITEIT VAN DE METADATA

Uit de vorige paragrafen is al gebleken wat de meerwaarde is van documentatie van digitale data. In de context van wetenschappelijke integriteit kan men twee soorten inhoudelijke metadata onderscheiden. Enerzijds is er sprake van ‘professionele metadata’ en anderzijds duikt het fenomeen van *social tagging* in huidige digitale omgevingen meer en meer op.

Professionele metadata zijn metadata die toegekend zijn door experts, gebruikmakend van standaarden en eventueel ook van aanbevolen thesauri (gecontroleerde woordenlijsten). Afhankelijk van de sector zijn verschillende metadatastandaarden en thesauri courant. Professionele metadata betreffen informatie waarvan de herkomst duidelijk identificeerbaar en betrouwbaar is en waar experts of auteurs zelf verantwoordelijk voor zijn geweest⁷.

Het is de taak van het archief om aan de auteur of degene die de data aanlevert richtlijnen mee te geven over de minimaal vereiste metadata en dit met het oog op een geslaagde bewaring van de data. Het OAIS-model schrijft conceptueel voor welke metadata noodzakelijk zijn (cf. supra): technische, administratieve (bv. rechtenbeheer), structurele, contextuele (bv. links tussen verschillende datasets) en provenance (ontstaansgeschiedenis van het digitale object) gegevens.

In het licht van web 2.0-technologie en de steeds actievere participatie van de gebruiker is de gebruiker geëvolueerd van een lezer en toeschouwer naar een *co-creator*. Steeds meer online bronsystemen en databanken laten de (anonieme) gebruiker toe om een eigen bijdrage, commentaar of tag toe te voegen. Maar in welke mate komt de garantie van kwaliteit en dus ‘wetenschappelijkheid’ hierdoor in het

gedrang? En op welke manier kan men oplossingen bieden om de authenticiteit, de controle en betrouwbaarheid van een digitaal document te blijven verzekeren?

De meerwaarde van *social tagging* is groot en de gebruikersannotaties kunnen onmiskenbaar een verrijking betekenen op het gebied van de ontsluiting, visibiliteit en doorzoekbaarheid van de data. *Social tagging* biedt ongetwijfeld een aanvulling op de metadata die in een meer professionele of wetenschappelijke context gegenereerd zijn. En niet enkel de gewone gebruiker zal die meerwaarde ondervinden, ook de wetenschappelijke onderzoeker of expert zal door de *user generated metadata* de digitale data vanuit nieuwe invalshoeken kunnen benaderen, interpreteren en (her)gebruiken. Er bestaan reeds veel online initiatieven die *social tagging* in hun opzet integreren. Extra waakzaamheid bij de interpretatie van deze metadata is echter aangeraden.

Ondanks de bestaande twijfels vanwege instellingen ten opzichte van *social tagging* is in meerdere projecten de waarde ervan aangetoond. Een mooi voorbeeld is *Steve: The Museum Social Tagging Project*. In dit project werd een van de meest recente grootschalige onderzoeken naar de invloed van *social tagging* in de museumsector gevoerd. De kunstcollectie staat online en de gebruikers worden uitgenodigd om de kunstwerken te *taggen*. Op die manier krijgt men inzicht in de visie van de gewone man of vrouw en de wijze waarop die verschilt met die van de 'specialist' of 'academicus'.

Het fenomeen van *social tagging* kan nog verder worden doorgedreven dan enkel het vrij toevoegen van metadata. Een gebruiker kan namelijk ook op basis van de geconsulteerde data in het digitaal archief komen tot eigen onderzoeksresultaten die hij vervolgens aan hetzelfde archief wil aanleveren en erin bewaren. De gebruiker bouwt met andere woorden mee aan het archief door data te hergebruiken en zo nieuwe data toe te voegen. De eisen voor een duidelijke kwaliteitsgarantie kunnen zich dus tegelijkertijd op de drie niveaus afspelen. De gebruiker aan de outputzijde van het archief moet zich kunnen vergewissen van de kwaliteit van de (meta)data in het archief en hiermee rekening houden wanneer hij op basis hiervan auteur van nieuwe data wordt aan de inputzijde van het archief⁸.

VERSIEBEHEER

Het beheer van de historiek of 'provenance' van een digitaal object biedt zonder twijfel zeer verrijkende informatie over een digitale bron. Meer nog dan een analogoog object kan een digitaal object oneindig bewerkt, aangepast en aangevuld

worden. Het kan in een andere context voorkomen, verwerkt worden in multimediale collages en voortdurend hergebruikt worden. Hoewel de flexibiliteit van digitaal materiaal doorgaans als een voordeel ten opzichte van fysieke bronnen wordt beschouwd, is digitaal materiaal eigenlijk veel kwetsbaarder. Vaak vallen de aanpassingen of wijzigingen achteraf niet meer waar te nemen of te corrigeren. De betrouwbaarheid van digitale documenten kan dan ook gauw in vraag worden gesteld⁹.

Een belangrijke taak van het archief bestaat er bijgevolg in de authenticiteit van de digitale objecten en hun metadata te waarborgen en eventuele aanpassingen te documenteren. Inhoud, vorm, structuur, digitale integriteit en volledigheid van het origineel mogen niet wijzigen en ze moeten blijven beantwoorden aan het uitzicht van het digitale object op het moment van zijn ontstaan of aanlevering aan het archief. Het is een taak van het digitaal archief ervoor te zorgen dat de ‘geschiedenis’ van het object, zoals informatie over de originaliteit, herkomst en eventuele versies voldoende gedocumenteerd en bijgehouden wordt. Op die manier kan de gebruiker zich optimaal van de geloofwaardigheid en betrouwbaarheid van het object vergewissen en ook binnen vijftig of honderd jaar nog de precieze ‘digitale genese’ van de data reconstrueren en de mate van authenticiteit kennen. De technische infrastructuur van het archief moet er tevens op gericht zijn onrechtmatige wijzigingen en eventuele manipulaties tijdig te detecteren en ongedaan te maken. Zo is het bijvoorbeeld van belang dat van ieder document een originele versie of master bewaard wordt die ‘onaantastbaar’ is en dat daarnaast versies voorhanden zijn die men eventueel kan wijzigen, waarover dan weer de nodige informatie over wordt bijgehouden. Bovendien is het bij versiebeheer noodzakelijk om stevast de relaties tussen de data(sets) en hun versie(s) aan te geven (onder meer aan de hand van relationele en structurele metadata).

2. Toegang

Een ontoegankelijk archief is geen bruikbaar archief. De opzet van een langetermijnarchief bestaat er in om in eerste instantie goed gedocumenteerde data voor de toekomst te bewaren zonder dat men a priori met mogelijke ontsluitingsmodellen of zoekinterfaces voor verschillende gebruikersgroepen zal rekening houden. Er moet met andere woorden een duidelijk onderscheid in acht genomen worden tussen het toegankelijk *houden* op lange termijn en het toegankelijk *zijn*, nu en in

de toekomst. Daartoe zal voldoende documentatie mee gearchiveerd moeten worden om die toegankelijkheid en ontsluiting mogelijk te maken.

TOEGANKELIJKHEID

Vanzelfsprekend moet eerst aan een aantal technische vereisten voldaan zijn om de data veilig en betrouwbaar beschikbaar te maken. De besproken voorwaarden voor een betrouwbaar archief (zie boven) zijn daarbij cruciaal. Maar techniek alleen regelt de toegankelijkheid niet. Rechtenbeheer is een niet te onderschatten informatiepakket bij geleverde data. Mag iedereen de data zien, of enkel een beperkte groep? De auteur bepaalt wat het archief met de data mag aanvangen. Rechtenbeheer is in het archief dan ook een volwaardig, onmisbaar deel van de metadata. Een mooi voorbeeld om een bepaald gebruik van data te regelen, zijn de *creative commons* licenties¹⁰, waarbij de gebruiker op voorhand van de auteur toestemming krijgt tot bepaalde soorten hergebruik. Het archief dient dit soort informatie aan de hand van metadata en/of beveiligingsmechanismen aan de gebruiker duidelijk te maken.

Een bijkomende voorwaarde voor toegankelijkheid is dat de data begrijpelijk zijn. Maar voor wie? Het volstaat dat de *designated community* de data begrijpt en kan interpreteren. Het is immers niet haalbaar om ieder object inhoudelijk voor het onbepaalbare (toekomstige) grote publiek begrijpelijk of toegankelijk te maken. Wel is het mogelijk en noodzakelijk om de data steeds voldoende te documenteren en in te passen in een bepaald discours of kennisdomein van een vooraf gedefinieerd doelpubliek, bijvoorbeeld in een wetenschappelijke discipline, vakgebied of instelling.

ONTSLUITING

Ontsluiting heeft betrekking op de wijze waarop de data met hun metadata in het digitaal archief georganiseerd en geïndexeerd zijn en opengesteld worden aan het publiek. De ontsluitingswijze bepaalt namelijk in sterke mate de doorzoekbaarheid van de data. Het gebruik van gestandaardiseerde of gecontroleerde woordenlijsten en thesauri bevordert ongetwijfeld de zoekmogelijkheden van data, maar dit is een arbeidsintensief werkproces. Ook het gebruik van metadatastandaarden (en mappings tussen standaarden) en voldoende relationele metadata met het oog op contextualisering van de data binnen het archief, verhogen de kans op een optimaal zoekresultaat.

Idealiter wordt daarom een evenwicht gevonden in een aanbod van zowel eenvoudige als meer geavanceerde zoekopties. Niets belet het archief echter om verschillende zoekinterfaces, bovenop de algemene, voor specifieke doelgroepen op te zetten. De onderliggende metadata en onderlinge mappings zijn hierbij van groot belang.

OPEN ACCESS

In deze context is het onmogelijk over toegankelijkheid binnen een wetenschappelijke context te spreken zonder het Open Access-principe te vermelden. Volgens de verklaring van het *Budapest Open Access Initiative* (BOAI)¹¹ is de kern van Open Access een wereldwijde, elektronische verspreiding van *peer-reviewed*, wetenschappelijke tijdschriftartikelen zonder enige beperking (dus zonder prijs- en copyright-barrière). Deze basisdefinitie werd later bekrachtigd en uitgebreid waarbij vooral *The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*¹² een brede ondersteuning geniet. Het doel is 'to promote the Internet as a functional instrument for a global scientific knowledge base and human reflection and to specify measures which research policy makers, research institutions, funding agencies, libraries, archives and museums need to consider'. Het internet wordt dus gepropageerd als het internationale forum voor de verspreiding van wetenschappelijke kennis. Hierbij gaat het ondertussen niet enkel meer om traditionele wetenschappelijke artikels maar ook om ruwe data, onderzoeksgegevens en audiovisueel materiaal waar zowel academische onderzoekers als culturele erfgoedinstellingen zoals musea en archieven gebruik van kunnen maken. Vrije toegang tot onderzoeksgegevens, zij het van academische of van artistieke, audiovisuele oorsprong, zet immers verder aan tot creativiteit en de ontwikkeling van nieuwe kennis. De open en vrije consultatie van onderzoeksgegevens mag vanzelfsprekend enkel gebeuren met respect voor de juridische beperkingen inzake copyright en privacy die door nationale overheden bepaald zijn. Van informatie die in *Open Access* gepubliceerd is, zijn bepaalde auteursrechtelijke beperkingen opgeheven aangezien de auteur op voorhand zijn fiat geeft voor toegang tot de data. Ook hier wordt regelmatig teruggegrepen naar een *creative commons* licentie om het gebruik te regelen.

Aanvullend op de Open Access-beweging, die vooral de wetenschappelijke data aanbelangt, biedt *Open Content* audiovisueel materiaal van allerlei oorsprong digitaal aan met de mogelijkheid data vrij te kopiëren, te bewerken of te hergebruiken in eigen documenten. Behalve raadplegen en lezen kan de gebruiker de data doorgaans ook opslaan en printen. *Open Content*-systemen bieden bovendien

vaak tools aan om de deelname en bijdrage van gebruikers te optimaliseren³. Zo kunnen gebruikers bijvoorbeeld hun oordeel over de content opgeven door het toekennen van een aantal sterren of cijfers, waardoor vanzelf een algemene ranking of volgorde ontstaat. Ook bestaan systemen waarmee auteurs toch nog enkele copyright-beperkingen kunnen opleggen, eventueel naar bepaalde gebruikersgroepen of gebruikswijzen toe. Het bekendste *Open Content*-voorbeeld is ongetwijfeld Wikipedia, waarbij men vrij de inhoud van artikelen kan wijzigen, aanvullen of hergebruiken⁴.

Op basis van het *Open Content*-principe van hergebruik, kan men bovendien nieuw onderzoek genereren en zo een grotere dynamiek binnen het onderzoek laten ontstaan. De wereldwijde uitwisseling van expertise, standaarden en 'best practices' wordt gestimuleerd en richtlijnen voor de behandeling en het gebruik van data worden beschikbaar gemaakt. *Open Access* versnelt onderzoek, verrijkt onderwijs, bevordert de return van investeringen in onderzoek en geeft gelijke kansen aan arme en rijke landen⁵. Een uitspraak van Christopher Surridge is hier toepasselijk: 'Science happens not just because of people doing experiments but because they're discussing those experiments.'⁶ Auteurs kunnen daarenboven rekenen op een uitgebreider lezerspubliek en een verhoging van zichtbaarheid en impact van hun werk. Volgens Alma Swan is de belangrijkste motivatie van een auteur om tot publiceren over te gaan immers nog steeds de drang om hun vakgebied te verrijken: 'they publish to have an impact on their field.'⁷ Door de publicatie in *Open Access* worden de waarde en kwaliteit van data niet enkel beoordeeld door een select en invloedrijk groepje van 'peers', maar in principe worden OA-gegevens onderworpen aan de kritische evaluatie van de wereldwijde onderzoeksgemeenschap. Hierdoor wordt hun visibiliteit ook op langere termijn gegarandeerd*.

3. De gebruiker

Uiteraard mag de gebruiker in de hele opzet van een digitaal archief niet vergeten worden. Een wetenschappelijk gebruiker, en zoals verantwoord gaat het hier niet enkel over de academische gebruiker, moet zich kunnen verzekeren van de kwaliteit van data en van het archief, en er tevens op een wetenschappelijke manier mee kunnen omgaan. Als de data, de auteur en het archief aan de genoemde criteria

* Brindley gaat zelfs zover door te stellen dat de optimalisering van vrije toegang of open access in feite belangrijker is dan digitale langetermijnbewaring. Cf. noot 1

voldoen en de gebruiker de mogelijkheid heeft de nodige informatie over de data eenvoudig te raadplegen, dan is beantwoord aan de vereiste kwaliteit van de data voor wetenschappelijk gebruik. Uiteindelijk staat het de gebruiker wel vrij om de bronnen al dan niet wetenschappelijk te gebruiken.

In het OAIS-model is voor het specifieke doelpubliek de term *designated community* gereserveerd. Al in de inleidende paragraaf van de *CCSDS Recommendation for an OAIS Reference Model*, waarin de definitie van OAIS wordt beschreven, wordt het belang van en de focus op de *designated community* benadrukt: 'An OAIS is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community.'¹⁸ En even verder volgt de definitie van *designated community* als '[a]n identified Group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities.'¹⁹

De *designated community* hoeft dus in geen geval het grote, onbepaalde en ondefinieerbare publiek te zijn dat de gearchiveerde digitale data nu of binnen enkele decennia zal willen raadplegen. Terwijl dit anonieme publiek op voorhand moeilijk in te schatten is, moet het wel mogelijk zijn om de aard van de *designated community* te kennen, voor wie bepaalde informatie of digitale data met hun metadata bedoeld zijn. De *designated community* is met andere woorden een specifiek type *consumer*. Ze bezit een bepaalde basiskennis, een discours of kennisdomein, waardoor de interpretatie van voor hen bestemde digitale data op lange termijn gegarandeerd blijft.

In ieder geval is het belangrijk dat het archief op voorhand duidelijk de doelstelling en het doelpubliek in acht neemt, zodat de *designated community* zonder hulp van de auteurs die de data hebben aangeleverd, de data kunnen begrijpen. De definitie van het geïntendeerde publiek wordt idealiter zichtbaar gemaakt voor de gebruiker. De dataproductent wordt bijvoorbeeld verplicht om bij de aanlevering van zijn data aan te geven voor welk doelpubliek de betreffende data en metadata best begrijpelijk zijn. Afhankelijk van de achtergrond van de gebruiker en van de sector zullen immers andere (vakspecifieke) criteria van kracht zijn en andere verwachtingen scheppen bij wetenschappers en onderzoekers met betrekking tot het archief. Zo wil een kunsthistoricus zeker zijn van de authenticiteit van het beeld, de kleuren of de kunstenaar. Juristen en rechtbanken die gebruikmaken van digitaal materiaal willen zich er dan weer van vergewissen dat het gebruikte materiaal de nodige wettelijke bewijskrachtige vereisten bevat²⁰. Bovendien is het dus noodzakelijk dat de *designated community* inzicht krijgt in de toegangsrechten en licentievoorwaarden voor het raadplegen en/of gebruik van een bepaalde collectie.

Aangezien het archief uiteenlopend materiaal kan bevatten, dat bedoeld is voor verschillende gebruikersgroepen en hun gepaard gaande noden, is het mogelijk dat 'different policies might be needed for different communities as well as for different collection types.'^{*} In verband met toegangsrechten moet een gebruiker eventueel kunnen aantonen dat hij deel uitmaakt van een bepaalde *designated community* om een digitaal object te kunnen raadplegen.

4. Conclusie

Alvorens een wetenschapper gebruik zal maken van een archief, zal hij eerst de kwaliteit van het archief en de daarin bewaarde data nagaan. Een goede wetenschapper neemt immers niet alles wat een archief aanbiedt klakkeloos aan. Er is een aantal criteria opgesteld die de kwaliteit, de veiligheid en de toegankelijkheid van een archief moeten garanderen. Het Open Archival Information System vormt daarbij de leidraad.

Sluit dit uit dat ongestructureerde, ongenueanceerde of toevallige data worden opgenomen in het archief? Of dat de wetenschapper nog 'ontdekkingen' kan doen in een digitaal archief? Het wordt minder evident, maar als we op zolder een oude floppy vinden, zullen we wellicht ook nog pogingen ondernemen om de mysterieuze inhoud te achterhalen.

* 'An Audit Checklist for the Certification of Trusted Digital Repositories. Draft for Public Comment', RLG/NARA, 2005, p. 36. Ook op het vlak van toegang en ontsluiting kunnen afhankelijk van de designated community andere zoekmodaliteiten aangeboden worden. Zo wijst Annemieke de Jong in *Informatieprofessional*, pp. 25-26 op het feit 'dat niet alle gebruikers en raadplegers van audiovisueel materiaal op dezelfde manier kijken. Velen zijn alleen op zoek naar specifieke onderdelen van een programma, zoals een fragment of een citaat, en hebben geen boodschap aan informatie over het hele programma. Anderen zoeken 'abstracter', bijvoorbeeld op gegevens die een serie, genre of reeks programma's betreffen. Deze klanten willen niet lastig gevallen worden met informatie over de afzonderlijke afleveringen, laat staan over de individuele shots en items. Data over fysieke en administratieve kenmerken en bewerkingen van materialen moet [sic] ook weer apart kunnen worden opgevraagd en losstaan van elke beschrijving van de inhoud. En aan de invoerkant, de positie van de documentalisten, moeten deze verschillende facetten van een audiovisuele productie juist als een geïntegreerd, samenhangend geheel kunnen worden benaderd.'

EINDNOTEN

1. Dame Lynne Brindley - CEO, The British Library, keynote IPRES 2008.
2. Zie het rapport BOM-vl WP1 Gebruikersnoden – Taak 3 Wetenschappelijk archief waarop dit artikel is gebaseerd, <http://hdl.handle.net/1854/LU-764058>
3. Reference Model for an Open Archival Information System (OAIS), blue book CCSDS 650.0-B-1, <http://public.ccsds.org/publications/archive/650xob1.pdf>
4. Zie onder andere het Planets project <http://www.planets-project.eu/> (29/01/2010) en hoofdstukken 4 en 5 van Weenink, K., Waaijers, L., van Godtsenhoven, K. (2007). *A Driver's Guide to European Repositories : Five Studies of Important Digital Repository Related Issues and Good Practices*. Amsterdam: Amsterdam University Press. <http://dare.uva.nl/aup/nl/record/260224> .
5. Data Archiving and Networked Services : <http://www.dans.knaw.nl>
6. Laurents, S., Van Horik, R., Harmsen, H. (2008). *Datakeurmerk.nl kwaliteitsrichtlijnen voor digitale onderzoeksdata in Nederland*. Den Haag: DANS.
7. 'Want zonder metadata en contextgegevens kan je de onderzoekers net zo goed de woestijn insturen', volgens Willem Vanneste, Vanneste, W. (2000). *Selectie En Bewaring Van Digitale Archieven*. <http://ls.kuleuven.be/cgi-bin/wa?A2=indoo05&L=elardo&P=723> .
8. Niettemin dient de impact van UCC of social tagging ook gerelativeerd te worden. Zo wees een studie van Universal (Universal McCann (2008). *Power to the People. Social Media Tracker Wave 3*). uit dat ondanks de enorme toename van UCC-diensten, het toch hoofdzakelijk om 'toeschouwers' gaat en in veel mindere mate om 'spelers'. Zie ook Witteman, R. (2008). Hoe gaat User Created Content in Europa overleven? *Frank Watching*, 17.11.2008.
9. zie De Jong, A. (2002). *Preservation of the Web. Issues for Audiovisual Archives*. FIAT/IFTA en Boudrez, F. (2005). *Digitale handtekeningen en archiefdocumenten*. Antwerpen: David. p1: 'the "respect des fonds", the principle of the original order, that implies that we are dealing with fixed entities, obviously cannot be maintained'. <http://www.creativecommons.be/>
10. Cf. Budapest Open Access Initiative (2002): Dit is een verklaring opgesteld op 14 februari 2002 als uitkomst van een congres in Boedapest in december 2001 en is inmiddels ondertekend door 4911 individuen en 462 organisaties waarvan 5 Belgische organisaties en 211 individuen, Cf. <http://www.soros.org/openaccess/search.cfm?q=belgium>
12. Lees de volledige verklaring: <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>
13. Men kan hier refereren aan de gamesindustrie, aangezien hier nogal wat voorbeelden te noemen zijn waarbij spelers de zelf gecreëerde content beschikbaar stellen en met elkaar uitwisselen. In sommige online (multiplayer) games kan men zelfs die content verder vormgeven. Het spel The Sims (Cf. <http://thesims2.ea.com/exchange/>) bijvoorbeeld, aangeboden door spelmaker Electronic Arts, laat spelers toe hun huizen, interieur, kostuums, enzovoort zelf te ontwerpen en uit te wisselen. Zie ook De Waal, M. (2006). Open Content. *De Nieuwe Reporter*.
14. De omvang van Wikipedia is enorm. Alleen al de Nederlandstalige versie van de internetencyclopedie telt sinds 30.11.2008 500.000 artikelen. Cf. Al half miljoen artikelen op Nederlandstalige Wikipedia. *De Standaard*, 30.11.2008, http://www.standaard.be/Artikel/Detail.aspx?artikelId=DMF30112008_040&ref=nieuwsoverzicht
15. zie Dekeyser, R. (2005). *Alternatieve vormen van publiceren. (R)evolutie in de wetenschappelijke communicatie*. K.U.Leuven, p. 26. In de OECD Principles and Guidelines for Access to Research Data from Public Funding wordt het OA-principe nog onderschreven als: 'The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research', OECD (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. p. 11.
16. Waldrop, M. (2008). Science 2.0 - Is Open Access Science the Future? Is Posting Raw Results Online, for All to See, a Great Tool or a Great Risk? *Scientific American*.
17. Swan, A., Sheridan B. (2005). *Open Access Self-Archiving: An Author Study*. ex. sum.
18. CCSDS (2002). *Reference Model for an Open Archival Information System (OAIS)*. p. 10.
19. CCSDS (2002). *Reference Model for an Open Archival Information System (OAIS)*. p. 19.
20. Jones, M., Beagrie, N. (2008). *Preservation Management of Digital Materials: The Handbook*. Londen: British Library Publishing. p. 37.