

Onderzoeksproject 'Van Horen Zeggen fase III'

Haalbaarheidsstudie naar een innovatieve applicatie
voor de ontsluiting van mondelinge bronnen

Erik Mannens, UGent-MMLab
Tom Paridaens, UGent-MMLab
Laurence Hauttekeete, UGent-MICT
Tom Evens, UGent-MICT
Jan Gysels, UGent-MICT

Oktober 2007

E. Mannens, T. Paridaens, L. Hauttekeete, T. Evens, J. Gysels, *Haalbaarheidsstudie naar een innovatieve applicatie voor de ontsluiting van mondelinge bronnen*.
Gent, Universiteit Gent - IBBT, 2007. 171 p.

Van Horen Zeggen fase III is een erfgoedproject in coördinatie van de Universiteit Gent, vakgroep Nieuwste Geschiedenis, i.s.m. en met de steun van de Erfgoedcellen van Antwerpen, Brugge, Brussel, Gent, Hasselt, Ieper, Leuven, Mechelen, Tongeren en het Waasland en het Vlaams Centrum voor Volkscultuur, i.s.m. AMSAB-ISG, ADVN, KADOC, Stadsarchief Antwerpen, VRT Geluidsarchieven radio. Met de steun van de Vlaams minister van Cultuur, Jeugd, Sport en Brussel.

Inhoud

1	Management samenvatting	5
2	Context, historiek en motivatie	13
2.1	<i>Onderzoeksopdracht</i>	14
2.2	<i>Onderzoeksopzet en -activiteiten</i>	16
3	Mondelinge historische bronnen: een begripsbepaling	17
4	Een kritische analyse van de huidige situatie op het gebied van opname, bewaring en ontsluiting van mondelinge historische bronnen in Vlaanderen	20
4.1	<i>Aard en omvang van de collectie</i>	20
4.2	<i>Randdocumentatie</i>	21
4.2.1	Transcripties	22
4.2.2	Rechtenbeheer	23
4.3	<i>Digitalisering</i>	24
4.4	<i>Ontsluiting</i>	26
4.4.1	Index	27
4.4.2	Metadata	28
4.4.3	Beschikbaarheid via internet	29
4.5	<i>Gezamenlijke ontsluiting</i>	30
4.6	<i>Gezamenlijke opslag</i>	32
4.7	<i>Krachtlijnen en beschouwingen</i>	33
5	Internationale cases	38
5.1	<i>Instituut voor Beeld en Geluid in Nederland</i>	38
5.1.1	Historiek	38
5.1.2	Infrastructuur	38
5.1.3	Gebruikers	39
5.1.4	Huidige werking	39
5.1.5	Digitaliseringsplannen	42
5.2	<i>CATCH projecten (Continuous Access To Cultural Heritage)</i>	44
5.2.1	Onderzoeksopzet	44
5.2.2	Financiering	45
5.2.3	Realisaties	45
5.2.4	CHOICE (uitgebreid)	47
5.2.5	CHORAL (uitgebreid)	49
5.2.6	Bemerkingen	49
5.3	<i>British Library (Sound Archive) in Verenigd Koninkrijk</i>	50
5.3.1	Historiek	50
5.3.2	Infrastructuur	50
5.3.3	Gebruikers	51

5.3.4	Huidige werking-----	51
5.3.5	Digitaliseringsplannen-----	52
5.3.6	Digital Object Management Programme-----	52
5.3.7	Archival Sound Recordings 1 & 2-----	53
5.4	<i>National Library in Australië</i> -----	57
5.4.1	Historiek-----	57
5.4.2	Infrastructuur-----	57
5.4.3	Gebruikers-----	57
5.4.4	Huidige werking-----	58
5.4.5	Digitaliseringsplannen-----	58
5.4.6	Digital Services Architecture-----	60
6	Algemene vereistenanalyse -----	64
6.1	<i>Functionele vereisten</i> -----	64
6.2	<i>Niet-functionele vereisten</i> -----	69
6.3	<i>Algemeen plan van aanpak</i> -----	71
7	Haalbaarheidsstudie metadata -----	72
7.1	<i>Containerformaten</i> -----	73
7.1.1	Audiocontainers-----	73
7.1.2	Multimediacontainers-----	74
7.2	<i>Metadastandaarden</i> -----	78
7.2.1	MARC/MARC21-----	79
7.2.2	Dublin Core-----	81
7.2.3	EAD-----	83
7.2.4	METS-----	86
7.2.5	LOM-----	89
7.2.6	BibTeX-----	91
7.2.7	MPEG-7-----	91
7.2.8	Ondersteunende standaarden-----	93
7.2.9	MPEG 21-----	95
7.2.10	Overzicht-----	97
8	Haalbaarheidsstudie software systemen -----	100
8.1	<i>Open Source Media Asset Management raamwerken</i> -----	100
8.1.1	Fedora-----	100
8.1.2	DSpace-----	108
8.1.3	JeromeDL-----	115
8.1.4	Bricks-----	120
8.1.5	Samenvatting-----	124
8.2	<i>Commerciële Media Asset Management Systemen</i> -----	127
8.2.1	Pictura-----	127
8.2.2	Adlib-----	131
9	Streaming technologieën -----	137
9.1	<i>Streaming Servers</i> -----	137

9.1.1	Apple Darwin Streaming Server -----	137
9.1.2	Adobe Macromedia Flash Media Server 2-----	137
9.1.3	Windows Media Server 9 -----	137
9.1.4	RealNetworks Helix Server/Helix DNA Server-----	137
9.1.5	Apache-----	138
9.1.6	Orb -----	138
9.1.7	VLC -----	138
9.1.8	Samenvatting-----	139
9.2	<i>Audio searching</i> -----	140
9.2.1	Zoeken op metadata-----	140
9.2.2	Zoeken via Speech-to-Text -----	141
9.2.3	Doorzoeken door karakteristieken (pauzes,...)-----	142
10	Beheermodellen -----	143
11	Literatuurlijst -----	145
12	Bijlagen -----	150
12.1	<i>Bijlage 1: schriftelijke vragenlijst</i> -----	150
12.2	<i>Bijlage 2: Respons schriftelijke vragenlijsten</i> -----	171
12.3	<i>Bijlage 3: Rondetafelgesprekken</i> -----	172

1 Management samenvatting

Europa, en inzonderheid Vlaanderen, heeft een zeer rijk cultureel verleden. Musea, archieven, bibliotheken en lokale erfgoedcellen beheren aldus heel waardevolle en diverse verzamelingen erfgoedstukken, die ons toelaten om het verleden te begrijpen en het heden te duiden. Tot voor kort waren deze culturele schatten echter min of meer onzichtbaar. Het was heel moeilijk om uit te zoeken wat er zich bevond in welke verspreide collecties, behalve dan bij de grote instellingen. Meer nog, een persoonlijk bezoek was nodig als iemand een bepaald stuk wou zien of raadplegen.

Nieuwe digitale technologische mogelijkheden stellen ons nu echter in staat om deze culturele bronnen, inzonderheid mondelinge historische bronnen, te ontsluiten voor een breder publiek. Dit ontsluiten van culturele bronnen heeft veel voordelen. Locale en internationale onderzoekers zullen nieuwe vormen krijgen om bepaalde theses te staven. De culturele instellingen zelf zullen ineens nieuwe markten en een groter publiek kunnen aanspreken. Ook uitgevers, conservators, intendanten en exposanten zullen nieuwe opportuniteiten ontdekken. Er zijn echter ook nog een aantal obstakels. Er bestaat tot op heden geen gemakkelijke manier om het bestaan en locatie van erfgoedmateriaal te identificeren. Ook de complexe materie van het beheer van intellectuele eigendomsrechten maken het moeilijk om erfgoedmateriaal in een digitale vorm aan te bieden en te exploiteren. Verder is er ook nog een scala aan technische en andere standaarden dat moet doorgrond worden door de instellingen die erfgoedmateriaal in een digitale vorm wensen aan te bieden. Dit zijn slechts enkele van de belangrijkste problemen die moeten overwonnen worden om het volledige ontsluitingspotentieel van bijvoorbeeld mondelinge historische bronnen ten volle te kunnen benutten.

Daarom sloegen een aantal partners de handen in elkaar om een haalbaarheidsstudie uit te voeren. Het onderzoeksproject 'Van Horen Zeggen III' (vervolg van de projecten Van Horen Zeggen I en II) werd opgezet onder de coördinatie van prof. Dr. Bruno De Wever (Vakgroep Nieuwste Geschiedenis, Universiteit Gent). De nodige onderzoeksexpertise werd aangeleverd door MMLab (technische componenten) en MICT (behoefteonderzoek), beiden behorend tot Universiteit Gent en eveneens het Interdisciplinair Instituut voor BreedBand Technologie (IBBT). Het onderzoek verliep in samenwerking met het Vlaamse Centrum voor Volkscultuur vzw (VCV) en werd ondersteund door de minister van Cultuur, Jeugd, sport en Brussel.

Het startpunt van dit vervolproject is dan ook een klare kijk op de vereisten van de culturele sector in zake archivering en ontsluiting van mondelinge historische bronnen. Veel organisaties die deze culturele bronnen beheren zijn klein en hebben slechts een beperkt budget ter beschikking. Meestal hebben ze ook niet de nodige experts in huis als het gaat over het digitaliseren of het beheren van intellectuele eigendomsrechten. De meeste culturele huizen hebben nood aan een duidelijk advies in verband met intellectuele eigendomsrechten van de bronnen die zij beheren. Verder is er ook nood aan een raamwerk die het transparant beheren van licenties mogelijk maakt. Veel organisaties vragen ook duidelijke informatie over hoe ze best technisch te werk gaan tijdens het digitaliseringsproces en welke standaarden ze daarbij het best gebruiken.

De grote vraag naar en de sterke interesse in een gemeenschappelijk systeem voor de archivering en ontsluiting van mondelinge historische bronnen schraagt de relevantie van een haalbaarheidsstudie voor de ontwikkeling van een innovatieve applicatie voor de ontsluiting van mondelinge historische bronnen in Vlaanderen.

Het feit dat de bevroagde actoren een systematisch en gestructureerd archiefbeleid voeren met de bedoeling ontsluiting van historische materialen te maximaliseren kan ongetwijfeld als een van de grote **sterktes** worden ervaren. De grote mate van ontsluiting leidt er toe dat mondeling historische bronnen snel en efficiënt teruggevonden kunnen worden door interne medewerkers of extern geïnteresseerden.

Binnen de archief- en erfgoedsector bestaat een sterke behoefte aan een gestandaardiseerd model voor het annoteren van historisch materiaal, zowel voor het toekennen van formele als inhoudelijke beschrijvingen. Geen enkele bestaande standaard voldoet immers (volledig) aan de vereisten voor een 'algemeen Vlaams beschrijvingsmodel'. De huidige tendens om collecties summier en minder diepgaand te beschrijven, dient weerspiegeld te worden in een complementaire en modulaire metadatastandaard.

Samen met de grote vraag naar gestandaardiseerde metadata vertoont de archief- en erfgoedsector een sterke interesse voor een gemeenschappelijk ontsluitingssysteem voor mondelinge historische bronnen. Het voorstel om de verschillende historische collecties gezamenlijk beschikbaar stellen, diverse databanken te koppelen en deze via één zoekscherm te ontsluiten, wordt met veel enthousiasme onthaald door de bevroagde actoren.

De al bestaande (voorzichtige) samenwerking tussen de diverse archief- en erfgoedinstellingen zoals in het kader van specifieke databanken (ODIS, Archiefbank Vlaanderen...) of projecten (bv. Expo '58), vormen alvast een vruchtbare bodem voor de toepassing van een uniform metadatamodel en de ontwikkeling van een gemeenschappelijk ontsluitingssysteem.

Bij **zwaktes** detecteren we de gebrekkige situatie op het gebied van digitalisering van mondelinge historische bronnen. Het omzetten van historisch gegroeide collecties analoge materialen naar digitale vorm wordt wegens de specifieke inspanning (tijd, personeel, budget...) zeker niet bij de grote archiefinstellingen als prioritair aangegeven. Het niet digitaal beschikbaar zijn van een groot aandeel van mondeling historische bronnen in Vlaanderen is echter geen optimaal startpunt voor de ontwikkeling van een gemeenschappelijk digitaal ontsluitingssysteem.

Het gebrek aan een uniform metadatasysteem vormt een tweede zwakte in het huidige beleid op het gebied van bewaring en ontsluiting van mondeling historische bronnen. Specifieke, projectmatige, inconsistente en/of afwezige metadata zorgen voor gebrekkige beschrijvingen, suboptimale ontsluiting en moeilijke samenwerking met andere instellingen en complexe integratie van beheerssystemen.

Een substantiële zwakte in het huidige archiveringsbeleid van mondeling historische bronnen betreft het gebrekkige beheer van de intellectuele en auteursrechtelijke bepalingen die erop gelden. De juridische situatie van de bronnen is veelal niet opgenomen in de metadata en vaak is er ook geen koppeling tussen contract en historische bron. Tot slot bevatten de overeenkomsten al te vaak geen regelingen betreffende verdere verspreiding of ontsluiting van de bronnen zodat de valorisatie en ontsluiting via het internet niet vanzelfsprekend wordt.

Tot slot wordt de Vlaamse archief- en erfgoedsector door een sterke fragmentatie en versnippering gekenmerkt. De talrijke en diverse instellingen hebben over de jaren heen eigen beheer- en ontsluitingssystemen opgezet en eigen praktijken op het gebied van beschrijving opgebouwd zodat een gemeenschappelijk ontsluitingsbeleid niet vanzelfsprekend is. Bovendien leidt dergelijke decentralisatie tot het inefficiënt aanwenden van middelen.

De huidige situatie biedt aldus op bepaalde vlakken een ideaal startpunt voor de ontwikkeling van een gemeenschappelijke ontsluiting van mondelinge historische bronnen in Vlaanderen maar brengt eveneens enkele fundamentele problemen naar boven die de verdere toekomst van dit project kunnen bemoeilijken of zelfs tegenhouden. Het lijkt ons evident dat aan deze zwaktes een meer dan speciale aandacht wordt besteed.

Een enorme **kans** is weggelegd voor maximale beschikbaarheid, consulteerbaarheid, toegang en hergebruik van mondelinge bronnen. Gemeenschappelijke ontsluiting via het internet maakt dit unieke erfgoed beter zichtbaar voor diverse doelgroepen en doeleinden: wetenschappelijk, educatief, cultureel, media... Een vlot toegankelijk, gebruiksvriendelijk, performant, gestructureerd en geactualiseerd ontsluitingssysteem kan in belangrijke mate bijdragen tot de verdere valorisatie en democratisering van het rijke orale verleden.

Een dergelijk systeem kan eveneens een oplossing aanreiken voor de huidige versnippering op het gebied van het gehanteerd metadatamodel en kan mogelijks leiden tot een gedeeld model dat aan eenieders wensen en noden beantwoordt, eventueel geïnspireerd door internationale praktijken. De huidige internationale archiefstandaard beantwoordt immers niet ten volle aan de vereisten op het gebied van inhoudelijke en formele beschrijvingen. Standaarden dienen zich eveneens aan te passen aan de huidige trend om historische materialen summier en economisch te beschrijven.

Een dergelijke gemeenschappelijke minimale dataset moeten eveneens kansen bieden voor een betere samenwerking tussen de diverse archief- en erfgoedactoren, geconcretiseerd in het eenvoudig uitwisselen van bronnen. Krachtenbundeling kan opnieuw leiden tot een betere zichtbaarheid van de waardevolle collectie en de representatie van de sector in het educatieve, sociaal-culturele en politieke veld versterken.

De ontwikkeling van een gemeenschappelijk ontsluitingssysteem biedt een vruchtbare bodem voor het ontwikkelen van nieuwe methodes voor (vol)automatische toekenning van metadata, een probleem waar men binnen de eigen systemen mee geconfronteerd wordt. Door eliminatie van het tijdrovend omzetten van bestandsformaten en dragers kan eveneens de workflow binnen de bevroegde organisaties geoptimaliseerd worden.

Bij de mogelijke **bedreigingen** identificeren we de mogelijke incompatibiliteit van de huidige systemen. Het hanteren van een gemeenschappelijk ontsluitingssysteem impliceert de integratie van de huidige systemen in één globaal model. De afwezigheid van een uniform

metadatamodel en de diversiteit van gebruikte bestandsformaten verhoogt het risico op een complex technisch integratieproces.

Ernstige aandacht dienen we ook te besteden aan het behoud van de samenhang van de verschillende historische bronnen en de context waarin ze opgenomen, bewaard en ontsloten zijn. Het reduceren van historische bronnen tot consulteerbare bestanden doet afbreuk aan de historische waarde ervan, die mede beïnvloed wordt door de context en samenhang met andere collectiestukken.

Bovendien dient de exacte rol van de betrokken actoren nauwkeurig vastgelegd te worden en moet de onduidelijkheid over de kostenstructuur van het digitaal archiveren en ontsluiten uitgeklaard worden. Er zijn met andere woorden financiële afspraken vereist met respect voor de relatieve positie van de diverse instellingen in het erfgoedveld.

Een andere techn(olog)ische uitdaging betreft de continue ondersteuning en aanpassing van het gemeenschappelijke systeem. Systemen dienen zich aan te passen aan de veranderende omstandigheden en daardoor bestaat het gevaar op verstarring van het systeem. Het te ontwikkelen systeem moet toekomstgericht zijn en bovenal werkbaar blijven. In het ideale geval kan een zelfstandig onderzoekscentrum of een onderzoekscel behorend tot een bestaande structuur de continue ondersteuning en actualisering van ontsluitingssystemen begeleiden.

Aldus biedt het project een aantal kansen die een mogelijke oplossing bieden voor de huidige problemen op het gebied van opname, bewaring en ontsluiting van mondelinge historische bronnen in Vlaanderen. Tenslotte dienen we bij het overwegen en realiseren van de ontwikkeling van dergelijk beheerssysteem eveneens oog te hebben voor enkele potentiële valkuilen die de ambitieuze toekomstplannen kunnen doorkruisen.

Ten eerste kan men zich bij dit alles de vraag stellen of het niet onwezenlijk is een applicatie te ontwerpen voor de exclusieve ontsluiting van mondelinge historische bronnen? De bevroegde instellingen hebben immers zelf te kennen gegeven dat mondeling historische bronnen slechts een relatief marginaal aandeel van de totale collectie uitmaken en geen prioriteit vormen binnen de reguliere werking. Kan de invalshoek van dit ontwikkelingsproces niet worden opengetrokken naar het ruime spectrum van archiefmaterialen waarbij niet louter de mondeling historische bronnen van de betreffende instellingen maar meteen de volledige

archieffcollecties aan elkaar gekoppeld worden en in één zoekscherm zoek- en opvraagbaar worden gemaakt?

Het moet immers de ultieme doelstelling zijn de rijke diversiteit van het Vlaamse archief- en erfgoedpatrimonium in één zoekscherm te ontsluiten; de huidige trend om verschillende databases betekenisvol aan elkaar te linken, mag daarom niet genegeerd worden. Geïnteresseerde gebruikers kunnen dan ‘op zijn Googles’ door middel van identieke argumenten tekst- en mediafragmenten doorzoeken. Het koppelen van het Vlaamse iconografische, audiovisuele en tekstuele geheugen zou een enorme kans bieden om het huidige versnipperde erfgoedveld in één muisklik samen te brengen.

Een grote uitdaging daarbij betreft ongetwijfeld het ontwikkelen van een gemeenschappelijk metadatasysteem dat toepasbaar is op de diverse types historische bronnen. Het beschrijven van bewegende beelden, beeldbanken, tekst... aan de hand van een minimale (inhoudelijke en formele) dataset lijkt inderdaad een onmogelijke en tegelijk massieve opdracht. Niettemin zijn wij ervan overtuigd dat één globaal ontsluitingssysteem een belangrijk model voor de toekomst uitmaakt.

Een tweede opmerking betreft de kwestie in welke mate een nieuwe applicatie voldoende meerwaarde schept tegenover de huidige systemen. Een aantal archiefinstellingen is nu al verenigd door middel van bepaalde databanken en ontsluitingssystemen, zodat in plaats van een nieuwe applicatie, ook de mogelijkheid voor een extensie op bestaande systemen minstens moet gewogen en overwogen worden.

Een zelfde opmerking kan gemaakt worden voor de betreffende modellen op het gebied van metadata. De huidige ISAD(G)-standaard voldoet niet aan eenieders wensen en noden maar vormt niettemin een ideaal startpunt voor een vraaggedreven en efficiënt model. Daarom pleiten we voor een extensie op de bestaande modellen zodat compatibiliteit gegarandeerd is. Verder moet de mogelijkheid onderzocht worden om technische metadata automatisch te generen, eventueel zelfs uit spraak zoals bij het Nederlandse onderzoekstraject CHoral.

Wegens de tijdrovende aard van catalogeren en beschrijven, is al langer de evolutie merkbaar waarbij buitenstaanders (vrijwilligers, senioren...) worden ingeschakeld bij het beschrijven van erfgoed. Met de opkomst van Web 2.0 wordt deze trend echter geradicaliseerd en zien we een verschuiving van passieve gebruiker naar interactieve deelnemer. De gebruiker kan zelf inhoud aanleveren, heeft zeggenschap over de vormgeving en bepaalt mee hoe iets beschreven wordt (*social tagging*). Folksonomies

duiden een beschrijvingsproces aan waarbij het publiek vrij gekozen trefwoorden toekent aan bepaalde inhoud; tags vormen immers samen geen ontologie. Dit in tegenstelling tot taxonomieën waarbij gebruik gemaakt wordt van een beperkte terminologie, op voorhand vastgelegd door een groep experts. Ondanks het succes van folksonomies vrezen archivarissen echter dat een ongecontroleerd vocabularium tot inconsistentie en onvolledigheid leidt. Welke invloed heeft deze paradigmawissel op de autoriteit van een erfgoedinstelling? En in welke mate draagt de interactieve deelnemer wezenlijk bij tot maximale beschikbaarheid en vooral terugvindbaarheid [1]? De grote populariteit van YouTube of Flickr is uiteraard niet representatief voor de Vlaamse cultuurprojecten. Toch dienen folksonomies taxonomieën niet uit te sluiten. Niets belet immers om een hybride vorm van interactie toe te passen. Bovendien maakt het semantisch web (Web 3.0) langzamerhand opgang. In Web 3.0 worden verschillende terminologieën aan elkaar gekoppeld zodat betekenisvermeerdering optreedt. (Van zodra term A gerelateerd is aan term B is het mogelijk om zoekresultaten naar boven te halen die uitsluitend gekoppeld zijn aan term B, ook al heeft de gebruiker uitsluitend term A als zoekargument gebruikt.) In Web 3.0 is er niet alleen meer structuur, door het leggen van betekenisvolle relaties kan de gebruiker gemakkelijker en efficiënter in databases zoeken [2].

Uit recent onderzoek binnen Vlaanderen blijkt bovendien dat de virtuele cultuurdeelnemer wel degelijk bestaat [3] [4]. Met virtuele cultuurdeelnemer bedoelen we de deelnemer die via internet en nieuwe media kennismaakt met en zowel actief als passief deelneemt aan kunstuitingen (zoals in film, theater, muziek, beeldende kunsten) en cultureel erfgoed. Bovenstaande studies tonen aan dat de virtuele cultuurdeelnemer niet alleen op zoek is naar culturele informatie maar ook graag via het web van kunst geniet. Cultuurparticipatie wordt nog té vaak bekeken vanuit de fysieke participatie waarbij virtuele kunst- en culturomgevingen niet zozeer als complementair maar eerder als concurrentieel wordt beschouwd. De digitale ontsluiting van mondelinge bronnen biedt echter een unieke kans om de interesse in en het gebruik van lokaal en oraal erfgoed te stimuleren. Een ruimere bekendheid brengt de rijkdom van het beheerde Vlaamse erfgoed aan het daglicht en leidt tot betere conservatie, intensiever gebruik en verdere wetenschappelijke valorisatie, mogelijkheden die voorheen buiten bereik lagen. Het komt de publiekswerking van de archief- en erfgoedinstellingen ten goede: nieuwe vormen van beschikbaarheid en een groter aanbod gaan gepaard met het bereiken van een groter en nieuw publiek. Niet alleen leidt virtuele kennismaking tot een sterkere interesse, de verbeterde toegankelijkheid laat ook participatie toe voor personen die fysiek niet (langer) in staat zijn om de diverse kunst- en erfgoeduitingen fysiek te consulteren.

2 Context, historiek en motivatie

Er bestaat momenteel in Vlaanderen een achterstand op het vlak van de ontsluiting van mondelinge bronnen. Met mondelinge bronnen bedoelen we o.a. (historische) opnames van interviews en getuigenissen (mondelinge historische bronnen), verhalen of liederen, dialecten of andere vormen van ‘historisch klankarchief’. Een mondelinge, historische bron is een belangrijk eindproduct van mondelinge geschiedenisprojecten en is een cruciaal onderdeel van mondelinge geschiedenis als praktijk. Interviews zonder specifieke historische vraagstelling, opgenomen toespraken, geregistreerde (volks)verhalen, dialectvarianten en straatliedjes, ... zijn vormen van oraal erfgoed, die historisch relevant zijn. Mondelinge bronnen worden analoog of digitaal geregistreerd op geluidsdragers (op band, cassette, plaat, cd, flash card, minidisc, DVD,...). Er is een breed spectrum van ‘producenten’ van mondelinge bronnen, en vermoed wordt dat slechts een beperkt deel in bewaring gegeven wordt bij diverse archiefinstellingen.

Er is in Vlaanderen geen instelling of organisatie die expliciet de opdracht heeft de ‘problematiek’ van de ontsluiting, opslag en bewaring van mondelinge bronnen te ondersteunen. De situatie is dan ook weinig rooskleurig. Over de stand van zaken verscheen een rapport van Roel Vande Winkel, *Mondelinge historische bronnen in Vlaanderen*, December 2004 (Onderzoek in het kader van het project *Van Horen Zeggen I*, Universiteit Gent, vakgroep Nieuwste Geschiedenis, i.s.m. het Vlaams Centrum voor Volkscultuur vzw - VCV). In dit onderzoeksproject werd een kritische bevraging gedaan naar de status inzake het bewaren en ontsluiten van mondelinge, historische bronnen in Vlaanderen. Daartoe werden diverse archiefinstellingen bezocht en experts geïnterviewd. Het rapport kan gedownload worden op: <http://www.mondelingegeschiedenis.be> (onderdeel: ‘bibliografie’).

De voorbije jaren zijn mondelinge bronnen meer onder de aandacht gekomen via diverse initiatieven, waarbij de ontwikkeling van verhalenbanken in het erfgoedveld het meest in het oog springt. (Bv. de Brugse verhalenbank – Erfgoedcel Brugge, de Volksverhalenbank – Seminarie voor Volkskunde van de KUL, de verhalenbank van ‘Brussel Behoort ons toe’, enz.). Diverse erfgoedorganisaties ontwikkelen projecten mondelinge geschiedenis die resulteren in mondelinge bronnen. Het steunpunt VCV ontwikkelde een werking rond mondelinge geschiedenis en gaf een publicatie uit: *Gestemd verleden: mondelinge geschiedenis als praktijk*, De Wever, B., & François, P., Brussel: VCV, 2003. Het VCV zette ook de website <http://www.mondelingegeschiedenis.be> op en gaf in 2005 in het kader van

het project Van Horen Zeggen II i.s.m. de UGent een DVD met praktische handleiding uit over mondelinge geschiedenis.

Maar er moet verder gedacht (en gehandeld) worden om de (kwetsbare) mondelinge bronnen te inventariseren en beter te ontsluiten, te bewaren en te valoriseren. Het is daarbij belangrijk te vermijden dat er op verschillende niveaus aan dure, parallelle ontwikkeling wordt gedaan van systemen voor de ontsluiting van mondelinge bronnen. Er is in dit stadium misschien een kans om sector- en instellingsoverschrijdend en schaalvergroterend te werken aan een gezamenlijk traject dat ook meer kansen en mogelijkheden biedt voor een kwalitatieve ontsluiting op langere termijn. Op een overleg dat tijdens het project *Van Horen Zeggen II* werd georganiseerd op 23 september 2005, werd door de aanwezige instellingen positief gereageerd om een onderzoekstraject op te starten om een aantal aspecten rond meer bepaald de ontsluiting van mondelinge bronnen verder te onderzoeken en om na te gaan wat de kansen en de voorwaarden zijn voor een nieuw te ontwikkelen applicatie.

Ondertussen zijn er internationaal diverse ontwikkelingen gaande waarbij geavanceerde applicaties worden ontwikkeld voor de ontsluiting en valorisatie van 'geluidscollecties' via het internet en die op professionele wijze dit type bronnen inventariseren. Deze *best practices* worden in dit onderzoek meegenomen.

2.1 Onderzoeksopdracht

Het onderzoeksproject 'Van Horen Zeggen III' werd uitgevoerd onder de coördinatie van prof. Dr. Bruno De Wever (Vakgroep Nieuwste Geschiedenis, Universiteit Gent). De nodige onderzoeksexpertise werd aangeleverd door MMLab (technische componenten) en MICT (behoefteonderzoek), beiden behorend tot Universiteit Gent en eveneens het Interdisciplinair Instituut voor BreedBand Technologie (IBBT). Het project werd opgezet i.s.m. en met de steun van de Erfgoedcellen van Antwerpen, Brugge, Brussel, Gent, Hasselt, Leper, Leuven, Mechelen, Tongeren en het Waasland en het Vlaams Centrum voor Volkscultuur, i.s.m. AMSAB-ISG, ADVN, KADOC, Stadsarchief Antwerpen, VRT Geluidsarchieven radio en met de steun van de Vlaams minister van Cultuur, Jeugd, Sport en Brussel (projectsubsidie Erfgoeddecreet).

Het opzet is om een (haalbaarheids)studie uit te voeren i.s.m. de participerende partners naar de mogelijkheid om in Vlaanderen een applicatie te ontwikkelen dat op flexibele wijze tegemoet komt aan de invoer- en ontsluitingspraktijken van mondelinge bronnen, en

hieromtrent een (technologisch en methodologisch) innovatief traject te ontwikkelen. Daarbij gaat ook aandacht naar de digitale lange termijnbewaring van gedigitaliseerde (of born digital) opnames en naar de toegankelijkheid en de valorisatie, bv. door consultatie via internet (website). Het is de bedoeling van het onderzoekstraject om (de verschillende componenten van) het systeem te beschrijven en de haalbaarheid ervan in te schatten.

De focus ligt hierbij zowel op de korte termijn als op de lange termijn:

- *op korte termijn* moet ingespeeld worden op de directe vraag van de partners, en meer bepaald van de betrokken erfgoedcellen, om een gezamenlijk systeem op te zetten dat kan ingezet worden om de mondelinge bronnen die worden geproduceerd in de context van allerlei projecten of in verhalenbanken, te ontsluiten. Men wenst hierbij te vermijden dat er parallel allerlei databanken worden ontwikkeld. Hierbij zal dus vertrokken moeten worden vanuit een operationele context, waarbij relatief snel een concreet resultaat moet worden geboekt: een concreet plan voor een 'basisapplicatie', die echter ook voor andere actoren dan erfgoedcellen een bruikbaar werkinstrument is of kan worden.
- *op lange termijn* kunnen strategieën worden geëxploreerd naar de innovatieve (of 'experimentele') componenten van een dergelijk systeem, waarbij in samenspraak met andere partners (ook in internationale context) onderzoekstrajecten kunnen worden opgezet of geïnitieerd die niet op korte termijn in concrete toepassingen moeten worden vertaald, maar die later ingeschoven kunnen worden in een gefaseerde verdere ontwikkeling van de 'basisapplicatie'.

De opdracht omvat concreet o.a. de volgende aspecten (desk research en prospectie):

- evaluatie van de actuele situatie in Vlaanderen m.b.t. de ontsluiting van mondelinge bronnen bij archiefinstellingen, de VRT en andere culturele en wetenschappelijke instellingen, erfgoedorganisaties en -verenigingen (o.a. de erfgoedcellen). Hiervoor zal verder gewerkt worden op basis van het rapport: R. van de Winkel, *Mondelinge, historische bronnen in Vlaanderen*, December 2004 (resultaat van het project fase 1 van 'Van Horen Zeggen')
- nagaan wat de (potentiële) gebruikersbehoeften en -toepassingen zijn (van onderzoekers, erfgoedprofessionals, onderwijs en andere doelgroepen). Hierbij zal

- vooral een beroep gedaan moeten worden op bestaande (gebruikers)onderzoeken en ervaringen in het veld
- studie van de (internationale) praktijken en methoden m.b.t. ontsluiting van mondelinge bronnen, meer bepaald:
 - modellen en standaarden van beschrijving van geluidsbronnen
 - problematiek van metadata
 - prospectie software oplossingen (bestaande applicaties)
 - prospectie hardware systemen (bestaande applicaties)
 - zoeksystemen op basis van spraakherkenningstechnieken en innovatieve ontsluitingspraktijken
 - digitalisering, digitale opslag en lange termijnbewaring van geluidsbronnen
 - prospectie van bedrijven, instellingen en onderzoeksinstanties, zoals meer bepaald het IBBT, VRT (partner in het project) en diverse (universitaire) onderzoeksinstituten en vakgroepen, die expertise (kunnen) ontwikkelen op dit terrein
 - contacten met organisaties in het buitenland die ervaring hebben met het beheer en ontsluiting van mondelinge bronnen.

2.2 Onderzoeksopzet en -activiteiten

Voor een evaluatie van de huidige situatie op het gebied van ontsluiting van mondelinge bronnen en het in kaart brengen van de functionele gebruikersvereisten werd een beroep gedaan op onderzoeksgroepen Media & ICT (MICT-IBBT, UGent) en Multimedia Lab (MMLab-IBBT, UGent). Er werd geopteerd voor een duale onderzoeksmethode met enerzijds een schriftelijke consultatie van de betrokken erfgoed- en archiefinstellingen en documentatiecentra, anderzijds werd tijdens een aantal expertinterviews gepeild naar de huidige en toekomstige visie op het gebied van het ontsluiten van mondelinge bronnen.

Teneinde de huidige situatie op het gebied van bewaren en ontsluiten van mondelinge bronnen in kaart te brengen, werd in een eerste fase aan alle partners een vragenlijst opgestuurd. Deze vragenlijst omvatte in totaal 57 vragen en behandelde aard en omvang van de collectie, digitalisering, (gezamenlijke) ontsluiting, technische infrastructuur, rechtenbeheer en behoeften van de gebruiker (zie [bijlage 1](#)). In totaal werden 21 vragenlijsten verzonden waarvan 19 (ingevuld) werden teruggestuurd (zie [bijlage 2](#)). Deze vragenlijst diende als eerste indicatie voor de huidige situatie op het gebied van bewaring en ontsluiting en gold eveneens als vruchtbare bodem voor de latere onderzoeksfases.

Aan de hand van de resultaten van deze bevraging werden een aantal contactmomenten met de betrokken instellingen voorzien. Op 14 mei werd een rondetafel met zowel Erfgoedsector als Archiefinstellingen belegd, op 16 mei werd een gesprek aangegaan met het VRT Geluidsarchief en SOMA (zie [bijlage 3](#)).

Om de resultaten van dit onderzoek in een ruimere context te plaatsen, werd tot slot ook over de taal- en landsgrens gekeken en werden werkbezoeken gebracht aan relevante buitenlandse onderzoekscentra. Zo was op 23 mei bij Beeld en Geluid (Hilversum, Nederland) en op 21 augustus bij British Library Sound Archive (Londen, Engeland) een delegatie van MICT en MMLab te gast om wederzijdse ervaringen uit te wisselen.

Verder werden ook de niet technische vereisten door Multimedia Lab (MMLab-IBBT, UGent) in kaart gebracht door technische diepte-interviews van enkele hierboven vermelde partners, door het technisch screenen van hun gebruikte Software en door uitgebreide deskresearch van “state of the art” compressieformaten, containerformaten, metadatamodellen en “open source” media asset management systemen.

3 Mondelinge historische bronnen: een begripsbepaling

De termen mondelinge geschiedenis (*oral history*) en mondelinge (historische) bron zijn onderling verwant en worden vaak maar niettemin onterecht als synoniemen gehanteerd. Academics hanteren ‘**mondelinge geschiedenis**’ als gemeenschappelijke term voor een aantal samenhangende onderzoeksactiviteiten die erop gericht zijn het verleden te reconstrueren door het interviewen van getuigen. Mondelinge geschiedenis is als dusdanig een geïntegreerde onderzoekspraktijk waarin een aantal stappen in min of meer onderstaande volgorde worden genomen [5] [6] :

- het formuleren van historische vraagstelling(en) waarop men een antwoord wil vinden door getuigen te interviewen
- het zich documenteren over de materie (periode, concrete gebeurtenis) waarnaar men onderzoek verricht door andere (bijvoorbeeld geschreven) bronnen te raadplegen
- het contacteren van potentiële getuige(n)
- het interviewen van deze respondent(en), gebruik makend van vooraf opgestelde vragen- en themalijsten

- het documenteren van deze getuigenissen door (voor of na het interview) identificatiefiches van de respondent(en) op te stellen
- het ontsluiten van de interviews, enerzijds door het afsluiten van een overeenkomst (met de respondent) over de toegankelijkheid van het interview, anderzijds door het opstellen van een volledige of verkorte transcriptie
- het aanwenden van de verworven informatie in een historisch betoog, dat diverse vormen kan aannemen (geschreven publicatie, tentoonstelling, audiovisueel productie, website enz.)

Daarenboven streeft de onderzoeker er doorgaans naar het interview op een audiovisuele informatiedrager (beeld- of geluidsbron) te registreren. Het geheel van documenten dat uit deze onderzoekspraktijk ontstaat, omschrijven we als een **mondellinge (historische) bron**.

Een mondellinge, historische bron is dus het eindproduct van een onderzoeksmethode, i.e. mondellinge geschiedenis [6] en bevat in ideale omstandigheden volgende randdocumentatie:

- een themalijst
- een vragenlijst
- het verzamelde materiaal uit andere historische bronnen en literatuur
- een identificatiefiche van de getuige
- het interview op band of een andere beeld- of geluidsdrager
- een interviewfiche
- een transcriptiefiche
- een bandinhoudsfiche
- een contract tussen interviewer en getuige waarin wederzijdse rechten en verplichtingen worden vastgelegd

Vandaag de dag groeit de belangstelling om het recente verleden te documenteren aan de hand van mondellinge getuigenissen. Deze interesse blijft niet alleen beperkt tot universiteiten en academische instellingen, ook bij het brede erfgoedveld (musea, archieven, erfgoedcellen, heemkundige kringen...) is het besef voor historische registraties exponentieel gegroeid. Tegelijk groeit ook internationaal het besef dat optimale bewaring en ontsluiting van mondellinge historische bronnen een cruciaal onderdeel is bij mondellinge geschiedenis. Niet alleen is het essentieel dat het kostbare orale erfgoed voor latere generaties verzekerd wordt, ook een snelle toegankelijkheid van dergelijke bronnen biedt kansen voor een ruimere verspreiding van het huidige erfgoed en verleden.

Tot slot wensen we nog te vermelden dat in klank- en beeldarchieven tal van vormen van oraal erfgoed aanwezig zijn die strikt gezien niet aan de definitie van mondelinge historische bron voldoen maar niettemin historisch relevant zijn. Dat geldt voor de door radio- of televisieomroepen opgenomen materialen maar ook eveneens voor audiovisuele sporen uit het verleden zoals opnamen van muziekconcerten, fictiefilms... Mondelinge historische bronnen vormen dus slechts een onderdeel van ruimere begrippen als oraal erfgoed of audiovisueel archivalia. Het lijkt ons evident dat de technische problematiek eigen aan het bewaren en ontsluiten van mondeling historische bronnen eveneens van toepassing is op andere audiovisuele dragers en archivalia [6].

4 Een kritische analyse van de huidige situatie op het gebied van opname, bewaring en ontsluiting van mondelinge historische bronnen in Vlaanderen

4.1 Aard en omvang van de collectie

Het aantal gestockeerde en ontsloten interviews loopt zeer sterk uiteen naargelang de instelling die bevestigd werd: de erfgoedcellen hebben minder bronnen dan de archiefinstellingen en beschikken in slechts één geval over meer dan 100 bronnen, bij de archiefinstellingen is er dan weer slechts één die over minder dan 100 bronnen beschikt. Twee erfgoedcellen hebben helemaal geen materiaal. De discrepantie tussen de bevestigde instellingen blijkt verder duidelijk uit de hoeveelheid historische materialen: sommige erfgoedinstellingen beschikken slechts over een handvol mondelinge bronnen, het rijkelijk gestoffeerde archief huisvest dan weer enkele tienduizenden interviews. Zeven van de 14 instellingen met mondelinge historische bronnen hebben enkel audio-opnames en geen videomateriaal.

We zien een wezenlijk verschil wat betreft de aangroei van de collectie en het verwerven van nieuwe mondeling historische bronnen. Of er nog nieuw materiaal aan de collectie wordt toegevoegd, is meestal afhankelijk van het lopen van nieuwe projecten (en bijhorende subsidies). Zo zijn een aantal erfgoedcellen vandaag betrokken bij het project Expo '58 naar aanleiding van een grote tentoonstelling voor zijn vijftigste verjaardag. De collectie van de archiefinstellingen is hoofdzakelijk historisch gegroeid maar wordt niet meer geactualiseerd. Enkel KADOC realiseert nog nieuwe interviews via lopende projecten. Bij andere archiefinstellingen vormt (het uitbreiden van) de collectie mondeling historische bronnen geen prioriteit.

In het algemeen kunnen we stellen dat de staat waarin het materiaal zich bevindt goed tot zeer goed is. Slechts één instelling gaf aan dat het bewaarde videomateriaal zich in slechte staat bevindt. Analoog materiaal wordt zonder uitzondering in niet-geacclimatiseerde ruimtes bewaard.

De tevredenheid over de manier van opslag daarentegen is bij de erfgoedcellen minder dan de archiefsector. De erfgoedcellen hebben weinig bronnen voorhanden en hebben ook geen uitgebreid en uitgebouwd systeem voor bewaring en beheer van de opnames. De erfgoedcellen zien de opslag van interviews ook niet als hun kerntaak en geven hun

materiaal bij voorkeur door aan een bewaarinstelling. Dit aspect draagt misschien bij tot de mindere waardering van de collectie.

De originele mondelinge audiobronnen worden op een diversiteit van dragers opgenomen en bewaard. Ook binnen één archief- of erfgoedinstelling zijn verschillende dragers terug te vinden. Zo is het materiaal opgeslagen op onder andere fonogram, magnetische band, magnetische draad, compact cassette, 78-toeren vinyl, DCC (digital compact cassette), DAT (digital audio tape). Een paar erfgoedcellen beschikken enkel over digitaal materiaal. Zeven instellingen hebben daarnaast ook videofragmenten in bezit. Deze zijn opgeslagen op pellicule, videoband, betacam, 1", DV, MiniDV en Digital 8. Ook op DVD wordt er materiaal bewaard, maar dit is dan digitaal materiaal.

Twee derde van de instellingen houdt een kopie bij als back-up van de originele mondelinge historische bron. Bij verschillende van deze instellingen is het echter zo dat men niet voor de volledige collectie een kopie bezit. Er is dus nog heel wat materiaal dat op slechts 1 exemplaar aanwezig is. Een groot aantal instellingen plant binnenkort wel enkele wijzigingen door te voeren betreffende het digitale beheer van de collectie.

4.2 Randdocumentatie

Het belang van contextdocumenten in het archiefwezen is vrij groot. De koppeling tussen identificatiefiche, bandinhoudsfiche, transcriptie en contract vormt een wezenlijk aspect in het situeren, identificeren en evalueren van de historische bron. Algemeen gezien probeert de bevraagde sector per interview een dossier op te stellen waarin alle contextinformatie is vervat en alle randdocumenten aan elkaar gekoppeld worden. Op vlak van algemeen beleid betreffende randdocumentatie scoren de bevraagde archief- en erfgoedactoren behoorlijk goed, ook al is er een duidelijk onderscheid te merken, afhankelijk van de oorsprong van de opname (in- of extern verworven).

De aanwezigheid van randdocumentatie is zoals gezegd sterk afhankelijk van de oorsprong van de bron. Interviews ontstaan binnen projecten en daardoor ook door eigen medewerkers gerealiseerd, zijn meestal voorzien van de nodige randdocumentatie. Bij extern verworven interviews (in bewaargeving) ontbreekt het echter geregeld aan contextinformatie zoals juridische afspraken en transcripties. Vaak probeert men de ontbrekende randdocumentatie te vervolledigen maar dit lukt niet steeds.

In het algemeen is de tevredenheid over het gevoerde beleid op het gebied van randdocumentatie behoorlijk en eigenlijk analoog aan de tevredenheid over de manier van bewaring: ook hier scoren de archiefinstellingen naar eigen zeggen beter dan de erfgoedcellen. Gevraagd naar toekomstige wijzigingen willen sommige actoren in de toekomst annotaties bij de opnames voorzien, juridische regelingen treffen en de randdocumenten met elkaar koppelen.

Het bijhouden van bandinhoudsfiches verloopt vrij gevarieerd binnen de sector. Iets meer dan de helft van de instellingen bezit voor driekwart van de materialen een bandinhoudsfiche, de resterende instellingen maken zelden dergelijke fiche op. Gemiddeld gezien is ongeveer de helft van alle mondeling historische bronnen gekoppeld aan een bandinhoudsfiche. De identificatiefiche is vaker aanwezig dan de bandinhoudsfiche. Ongeveer 70% van alle bronnen wordt gekoppeld aan een dergelijke fiche.

4.2.1 Transcripties

De aanwezigheid van transcripties verschilt sterk over de bevroegde instellingen. Bij vijf van de 14 instellingen beschikt men over geen enkele volledige transcriptie. Bij slechts 4 van de 14 instellingen is meer dan de helft van het materiaal volledig getranscribeerd. Deze transcripties worden in de meeste gevallen zowel op papier als digitaal bewaard, hetzij ingescand hetzij als een origineel computerbestand. Slechts een kwart van de instellingen beschikt niet over digitaal beschikbare transcripties. Een beperkt aantal actoren bezit transcripties waarvan de opnames verloren zijn gegaan. Doorgaans gaat het echter om oud en beperkt materiaal.

Een aantal instellingen geeft aan te (willen) stoppen met het volledig transcriberen van interviews. Het volledig uitschrijven of -tikken is een omvangrijke karwei – ‘begijnenwerk’ volgens één van onze respondenten – en blijkt té arbeidsintensief. Vaak worden enkel de relevante of historische waardevolle episodes uitgeschreven.

Het klopt inderdaad dat een dergelijke taak niet meer van deze tijd is. Binnen enkele jaren moeten video/audio mining of spraakherkenningstechnologie een wezenlijk alternatief vormen voor dit tijdrovende transcriptieproces. In Nederland werd dergelijke technologie in het kader van het CHoral-onderzoekstraject al – met wisselend succes – toegepast.

4.2.2 Rechtenbeheer

Het is een vaststelling dat juridische afspraken tussen interviewer en geïnterviewde niet steeds aanwezig zijn. De erfgoedcellen, recenter begonnen met het werven van mondeling historische bronnen, scoren hierin beter dan de archiefinstellingen, die vaak over een berg (ouder) historisch materiaal beschikken waarbij geen juridische afspraken zijn opgenomen. Algemeen gezien is slechts bij 54% van alle materialen een contract voorzien. De gebrekkige koppeling tussen contract en interview is echter meer problematisch.

De tevredenheid over de huidige situatie op het gebied van rechtenbeheer is niettemin behoorlijk. Instellingen geven vooral aan in toekomstige contracten meer specifieke consultatievoorwaarden (bijvoorbeeld voor ontsluiting via het internet) te voorzien. Actoren die nu zonder contracten werken (bijvoorbeeld via mondelinge afspraken) wensen in de toekomst deze afspraken toch vooral schriftelijk vast te leggen.

De kennis van de bestaande rechten die gelden op de audiovisuele collectie zijn behoorlijk goed gekend. Het valt niettemin op dat de instellingen geen overdreven risico's willen nemen en het belang van rechten groter vinden dan de publieke consulteerbaarheid ervan. Deze restrictieve voorzorgsmaatregelen hebben er mede toe geleid dat nog geen enkele instelling aansprakelijk is gesteld voor het onrechtmatig (her)gebruik van verworven materialen. Hoofdzakelijk in de archiefsector waakt een juridisch medewerker over de correcte toepassing van de wettelijke bepalingen.

Belangrijker is echter de relatieve complexiteit van het rechtenbeheer betreffende mondeling historische bronnen. Naast eigendomsrechten moeten de instellingen ook rekening houden met de privacy van de geïnterviewde, auteurs- en uitzendrechten... Doordat mondeling historische bronnen niet vrijelijk verdeeld worden (consultatie gebeurt meestal ter plaatse en gecontroleerd), zijn maatregelen tegen ongeoorloofde verspreiding voorlopig niet van toepassing.

Tot slot is het cruciaal te vermelden dat het gebrek aan contracten en vooral de gebrekkige voorwaarden voor consultatie en verspreiding in deze contracten een uitdaging vormen voor gemeenschappelijke ontsluiting via het internet. Een groot aantal contracten zijn opgesteld in een periode waarin van het internet nog geen sprake was en zodoende geen opening laten voor nieuwe vormen van (her)gebruik, consultatie en verspreiding. Ondanks de enorme waarde van het bewaarde historische materiaal, zorgt de feitelijke toestand voor beperkte

mogelijkheden betreffende gebruik en hergebruik van deze bronnen en wordt de culturele, educatieve en wetenschappelijke valorisatie enigszins belemmerd.

4.3 Digitalisering

Hoewel het een belangrijke voorwaarde is tot het opzetten van een innovatief (de)centraal ontsluitingssysteem, blijkt digitalisering duidelijk nog niet sterk doorgedrongen in de bevroagde erfgoed- en archiefinstellingen. Hoewel registraties momenteel veelal meteen digitaal gebeuren (Minidisk), blijken de meeste instellingen over een groot aantal analoge mondelinge historische bronnen te beschikken. Het valt meteen op dat de archiefinstellingen, die echter over een omvangrijkere collectie dan de erfgoedsector beschikken maar geen actief registratiebeleid voeren, nog niet zo ver als de erfgoedsector staan op het gebied van digitaal registreren en/of converteren van mondelinge historische bronnen.

Omdat het omzetten van analoge naar digitale bestanden nog te beperkt voorvalt, bepaalt het bezitten van born digital materiaal in sterke mate de huidige situatie op het gebied van digitalisering. Archiefinstellingen bezitten vaak een grote collectie van analoge bronnen (beschikbaar op tapes) die in vroegere jaren door schenkingen of eigen inspanning werd opgebouwd. Het digitaliseren van dergelijke hoeveelheid materialen vergt uiteraard een specifieke inspanning (tijd, personeel, budget...) van de betreffende instellingen maar wordt momenteel zelden als prioriteit binnen de werking aangegeven. De analoge tapes worden bewaard in niet-geacclimatiseerde ruimtes maar betekenen wél een potentieel gevaar voor ontbinding en vernietiging doordat deze niet digitaal gedupliceerd (bijvoorbeeld op server) zijn.

Een ander verhaal is te noteren bij de erfgoedsector die pas recent met het registreren van mondelinge historische bronnen gestart is en in vele gevallen over born digital materiaal beschikt. De beperktere omvang van de (analoge) collectie laat ook toe deze in snellere mate te digitaliseren aangezien de inspanningen daartoe veel beperkter zijn in vergelijking met de archiefinstellingen.

Deze tendens laat zich echter niet onder alle archiefinstellingen veralgemenen omdat enkele instellingen over een 'speciale missie' blijken te beschikken. Zo streeft VRT in het kader van het DIVA-project (Digitaal Vlaams/VRT Archief) naar de realisatie van een Digitale Mediafabriek die een deel van het archief van de openbare omroep naar het grote publiek toe moet ontsluiten. Terwijl opnames voor radio uitsluitend digitaal verlopen, worden televisieprogramma's nog vaak analoog aangeleverd. Hoewel de openbare omroep

momenteel deze uitzendingen naar digitale vorm omzet, zorgt dit voor een grijze fase waarbij zowel analoog als digitaal materiaal parallel het archief zal binnenstromen. Pas als alle programmaveveranciers ook digitaal zullen produceren, kan de VRT de stap naar volledige digitalisering wagen. Een ander opmerkelijk initiatief betreft eveneens het Digitaal Depot van KADOC die van het digitaliseren (en ontsluiten) van (mondeling) historische bronnen een prioriteit wil maken.

In grote lijnen kunnen we echter stellen dat digitalisering geen grote prioriteit blijkt onder de deelnemende erfgoed- en archiefinstellingen (met uitzondering van KADOC en VRT). Of mondeling historische bronnen al dan niet digitaal beschikbaar zijn, is vaak afhankelijk van de wijze waarop de registratie verlopen is en veelal een historisch gegroeide situatie. Sommige instellingen beschikken over een groot aantal analoge mondelinge historische bronnen maar slagen er niet in deze naar digitale vorm om te zetten. Organisaties als AMSAB geven aan digitalisering louter te overwegen als daar de nodige additionele middelen worden voor gegenereerd. Wegens de reguliere werking van de archieven, slagen deze instellingen er zelden in personeel en middelen vrij te maken voor het digitaal beschikbaar maken van deze (schat aan) informatie. Zo is SOMA pas recent begonnen met het digitaliseren van een beperkt aantal bronnen in het kader van een project. Het lijkt er sterk op dat in een eerste fase de nood aan middelen en kennis om de historisch verworven analoge collectie te digitaliseren sterker is dan de behoefte aan het opstarten van nieuwe projecten rond mondelinge geschiedenis.

De redenen om deze bronnen te digitaliseren lopen echter niet ver uiteen. Digitale bronnen blijken immers noodzakelijk wil men deze via het internet beschikbaar stellen, bovendien zijn digitale bestanden eenvoudiger in beheer en vormt het digitaliseren een relatief eenvoudig proces. De keuze van het materiaal dat het eerst in aanmerking komt voor digitalisering is afhankelijk van enkele criteria: materiaal in slechtste staat en het meest geconsulteerde materiaal zijn de meest gangbare, soms wordt ook prioriteit gegeven aan materiaal in het kader van projecten, evenementen en tentoonstellingen (thematisch criterium).

Organisaties die de stap zetten naar het digitaliseren van het archiefmateriaal, verkiezen verliesloze formaten zoals AIFF, WAV en FLAC. Het is opmerkelijk dat voor de opslag van digitale bestanden verliesloze standaarden worden gebruikt om optimale geluids- en beeldkwaliteit te garanderen. De keuze voor dergelijke standaarden is vaak gebaseerd op buitenlandse gevallen of op de kennis van een (technisch) medewerker die zich een tijdje heeft verdiept in de materie (bijvoorbeeld via internetfora). Wanneer dergelijke bronnen door

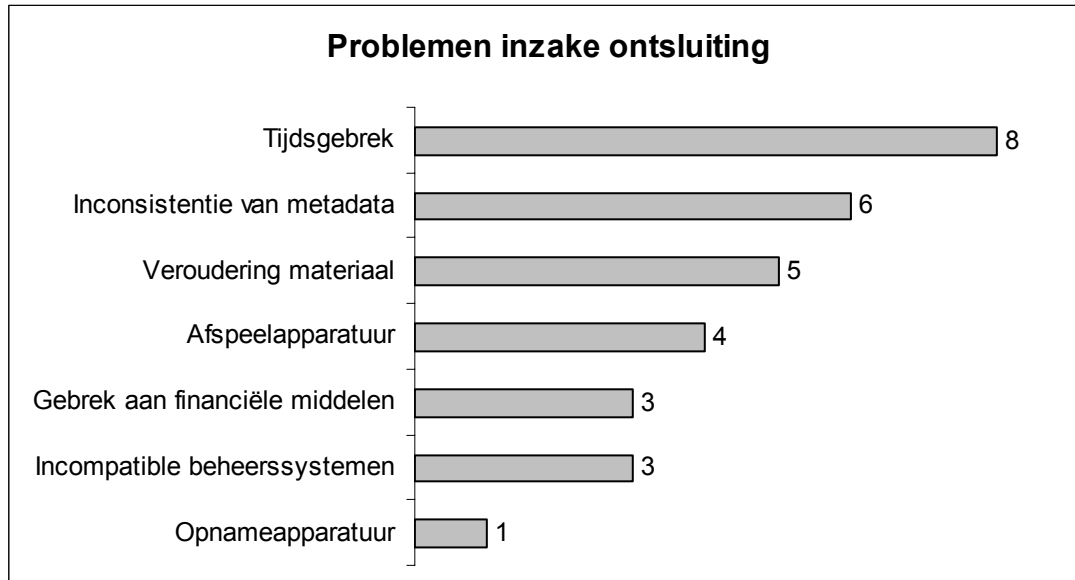
derden opgevraagd of geconsulteerd worden, worden deze omgezet naar verlieslatende standaarden zoals MP3 die gebruik maken van compressietechnieken. De digitale bestanden (born digital of gedigitaliseerd) worden gewoonlijk opgeslagen en bewaard op interne dataserver of externe (hard)disks (sterk variërend in opslag grootte) waarvan eveneens een back-up bestand voorhanden is.

4.4 Ontsluiting

De grote variatie in belang dat aan het digitaliseren van de collectie mondelinge historische bronnen wordt gehecht, wordt eveneens weerspiegeld in de mate dat deze worden ontsloten en publiek beschikbaar worden gesteld. Met ontsluiting wordt hier verwezen naar het door middel van een classificatiesysteem gestructureerd beschrijven zodat bronnen snel en efficiënt teruggevonden kunnen worden door medewerkers en/of externen. Evenals digitalisering is een afdoende en gestandaardiseerde ontsluiting vereist voor het beschikbaar stellen van de mondelinge historische bronnen in een internetomgeving.

In tegenstelling tot de digitalisering waar de erfgoedsector flexibeler bleek dan de archiefinstellingen, staan deze laatste veel verder op het gebied van de ontsluiting en beschrijving van de bewaarde materialen. Algemeen kan echter worden gesteld dat de bevraagde actoren een systematisch en gestructureerd archiefbeleid voeren met de bedoeling ontsluiting van historische materialen te maximaliseren.

Niettemin de grote mate van ontsloten materialen ondervinden heel wat instellingen problemen met deze praktijken. Onderstaande tabel geeft duidelijk weer dat het gebrek aan tijd (en dus onrechtstreeks prioriteit) en het niet-consequent hanteren van (inconsistente) metadata de grootste drempels zijn op het gebied van het ontsluiten van mondelinge historische bronnen. We komen daar bij de bespreking van de metadata nog op terug.



Ontsluiting vereist uiteraard een overzicht van alle materialen uit de collectie, die aan de hand van een consistent metadatasysteem beschreven zijn en aldus gemakkelijk terugvindbaar en/of consulteerbaar zijn. Om maximale beschikbaarheid na te streven kunnen archief- en erfgoedinstellingen hun collectie ook aan derden aanbieden, het internet biedt daarvoor onder andere een laagdrempelig potentieel.

4.4.1 Index

Een volledig overzicht van alle mondelinge historische bronnen in de collectie is bij vrijwel elke instelling voorhanden. Er is met name een 'index' beschikbaar waardoor alle bronnen opgelijst en aldus terugvindbaar zijn. Dergelijke databases zijn vaak niet meer dan Microsoft Office-documenten (Excel, Access, Word...) die een overzicht geven van alle geregistreerde interviews. Deze index is echter niet rechtstreeks gekoppeld aan de geluidsbronnen zelf. In een beperkt aantal gevallen wordt dergelijke index ook op het internet aangeboden zodat ook de occasionele gebruiker doorheen het archief kan bladeren. Een aantal archiefinstellingen worden eveneens in ODIS, LIBIS-Net en/of AdLib opgenomen.

Ook hier verdienen een aantal beheers- en ontsluitingssystemen extra aandacht. Zo biedt SOMA zijn catalogus 'Pallas' eveneens online aan. Pallas, ontwikkeld binnen SOMA, laat toe archiefbronnen, bibliotheekmaterialen én fototheek met één muisklik te doorzoeken. Dergelijke koppeling van diverse bronnen via één zoekscherm blijkt uniek binnen de archiefsector. Doordat de diverse bronnen worden beschreven aan de hand van trefwoorden, biedt een thesaurus hulp bij het consulteren van Pallas.

De collectie van KADOC is ontsloten via diverse geautomatiseerde databanken. In een samenwerkingsverband met de Katholieke Universiteit Leuven wordt een digitaal depot met archiefbeschrijvingssysteem opgezet onder de naam LIAS (Leuvens Integraal Archiveringssysteem). Voor de realisatie van dit project werd geopteerd voor DigiTool en de ontwikkeling van een EAD-module (in samenwerking met LIBIS). Dit systeem biedt de oplossing voor het beheer van digitale bestanden en laat toe digitale collecties te creëren, beheren, conserveren en ter beschikking te stellen.

Het ter beschikking stellen van de collectie gebeurt behoudens interne medewerkers ook voor externen, meestal in de hoedanigheid van onderzoekers, studenten of andere archiefwerkers. In sommige gevallen is eerst toelating vereist van de geïnterviewde personen vooraleer de aanvrager de betreffende materialen kan consulteren.

4.4.2 Metadata

Een adequaat ontsluitingssysteem vereist een gestandaardiseerd metadatamodel of tenminste het consistent toepassen van trefwoorden en labels. Het beschrijven van de archiefmaterialen (zowel inhoudelijk als vormelijk) moet tot een optimale doorzoekbaarheid van de collectie leiden en historische bronnen gemakkelijker uitwisselbaar maken. Binnen de bevroegde archief- en erfgoedinstellingen worden historische bronnen doorgaans beschreven aan de hand van gestandaardiseerde metadata. Een aantal organisaties grijpt terug naar bestaande modellen en standaarden, andere instellingen kiezen ervoor een eigen systeem in het leven te roepen.

In de archiefsector worden de internationale ISAD(G) en ISAAR-standaarden vrij frequent toegepast (bij KADOC en ADVN), AMSAB daarentegen heeft zelf een standaard uitgewerkt, gebaseerd op het systeem van de British Library Sound Archive. De erfgoedsector, die nog niet zo ver geëvolueerd is als de archiefinstellingen wat ontsluiting betreft, hanteert maar met mondjesmaat metadatasystemen. Erfgoedcel Meetjesland hanteert Dublin Core terwijl Erfgoedcel Brugge een eigen model heeft ontworpen dat evenwel kan geïntegreerd worden met Dublin Core. Andere erfgoedinstellingen hebben geen concrete plannen voor ontsluiting en/of wachten nog op een hanteerbare standaard.

Metadata biedt de mogelijkheid collecties op een gestructureerde wijze te beschrijven aan de hand van vormelijke en inhoudelijke kenmerken. Technische metadata worden momenteel veelal niet in de beschrijving opgenomen terwijl de huidige inhoudelijke standaard niet altijd aan de huidige noden voldoet. Daarenboven stelt zich een probleem bij extern aangeleverde

mondelinge historische bronnen (in bewaargeving) waarbij de controle op de toegevoegde metadata – net zoals bij de randdocumentatie – quasi nihil is. Eigen bronnen worden vaak wel beschreven, opnames door externen bieden echter minder garantie op een kwaliteitsvolle beschrijving.

Het mag duidelijk zijn dat er binnen de archief- en erfgoedsector een sterke behoefte bestaat aan een gestandaardiseerd systeem voor de inhoudelijke én technische beschrijving van historische materialen, dit om de werking van de nationale archieven en erfgoedcellen op elkaar af te stemmen. De huidige ISAD(G)-standaard is goed werkbaar omdat het naast een beperkt aantal verplichte identificatiegegevens ook in een aantal facultatieve uitgebreide velden voorziet, hoewel een inhoudelijke aanpassing sterk gewenst is. Het leveren van technische metadata moet geïntegreerd kunnen worden binnen een ontsluitingssysteem die deze automatisch aan het historisch materiaal toekent.

Waar ISAD(G) uitgebreid en diepgaande beschrijvingen toelaat, is de huidige tendens om historische bronnen eerder oppervlakkig en goedkoop te beschrijven. Uit efficiëntieoogpunt wordt ook bij de bevraagde actoren steeds vaker voor summiere en algemene metadata gekozen. Extensieve beschrijving van (mondeling) historische bronnen blijkt met andere woorden geen prioriteit wegens de tijrovende en arbeidsintensieve aard van het beschrijven.

4.4.3 Beschikbaarheid via internet

Teneinde maximale beschikbaarheid en consultatie na te streven, vormt het aanbieden van online consulteerbare mondelinge historische bronnen een enorme opportuniteit. Dit vereist uiteraard het digitaal aanleveren van de historische bronnen en een consistent beschrijvingssysteem teneinde optimale zoekmogelijkheid en terugvindbaarheid te garanderen.

Hoewel voor enkele organisaties een ideaal einddoel, benut enkel Erfgoedcel Brugge deze mogelijkheid door het online consulteerbaar maken van mondelinge historische bronnen. De Verhalenbank is een elektronische gegevensbank die digitale geluidsopnames met Bruggelingen bevat. Bedoeling is uiteraard het creëren van een digitaal klankarchief van de Brugse herinnering waarin zowel historici, cultuurmakers als het brede publiek naar specifieke verhalen kunnen zoeken. De Verhalenbank wordt mede ondersteund door partners als Stadsarchief Brugge en de Brugse Radio VBRO die eveneens opnames aanleveren.

Gelijkaardige projecten zijn ook te vinden bij Erfgoedcel Hoogstraten, 'Bruxelles nous appartient – Brussel behoort ons' (BNA-BBOT) en 'het HUIS van ALIJN' die eveneens verhalen via streaming audio aan de bezoeker aanbieden. Het Koninklijk Museum voor Midden-Afrika ontsloot onlangs 2550 uren muziek en vormt daardoor het belangrijkste digitale klankarchief voor Centraal-Afrika ter wereld. Deze projecten behoorden echter niet tot de scope van dit onderzoekstraject.

De mogelijkheid om ook andere documentatie digitaal beschikbaar te stellen, moet eveneens overwogen worden. Nú al bieden tal van instellingen (een deel van) hun collectie afbeeldingen voor de webbezoeker aan. Eventueel kan ook bekeken worden in welke mate de transcripties online kunnen gesteld worden. Niet zozeer technische maar eerder auteursrechtelijke aspecten kunnen hier een drempel creëren. Zo zien we nu al dat extern aangeleverde opnames – wegens gebrek aan contractuele afspraak – vaak onvoldoende ontsloten zijn. Ook bij zelfgemaakte opnames ontbreekt echter vaak een juridische afspraak die verdere valorisatie van mondeling historische bronnen in de weg staat.

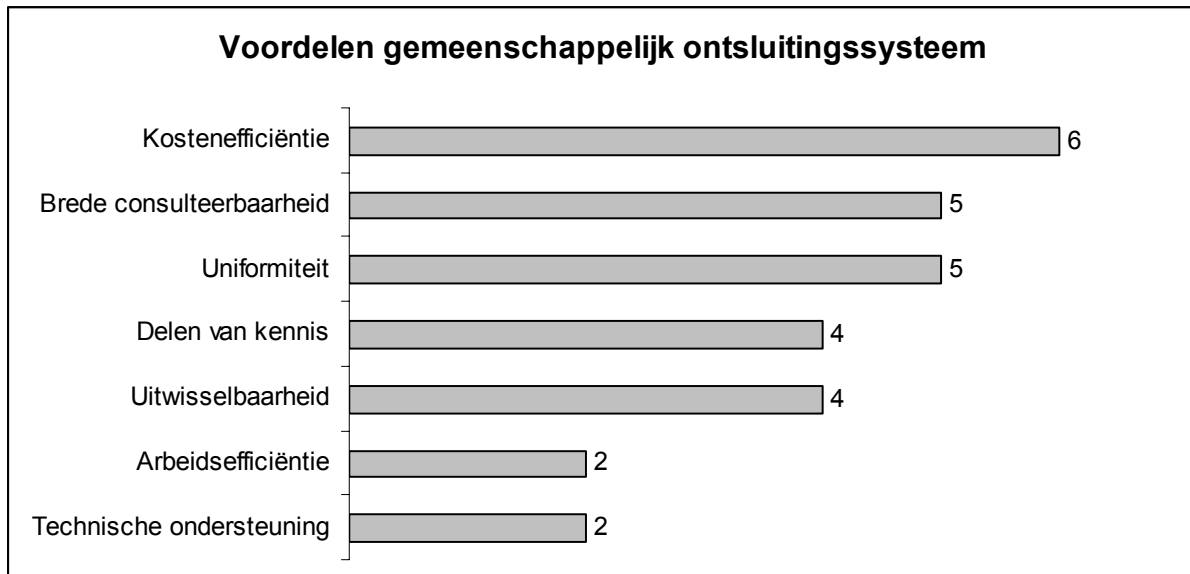
Momenteel worden mondelinge historische bronnen digitaal consulteerbaar gesteld door middel van een (beperkte) streaming. Het lijkt geen goed idee deze opnames ook ter download te stellen (o.a. omwille van intellectuele en eigendomsrechten), een opmerking die ook op andere bronnentypes toepasbaar is. Wie een kopie op een drager (DVD, CD-ROM...) wenst aan te vragen, kan zich dan eventueel wenden tot de archiefinstelling of erfgoedcel in kwestie. Het begrip 'institutionele controle' dient hier dan ook om de nodige controle over het historisch materiaal bij de betreffende instellingen te houden.

4.5 Gezamenlijke ontsluiting

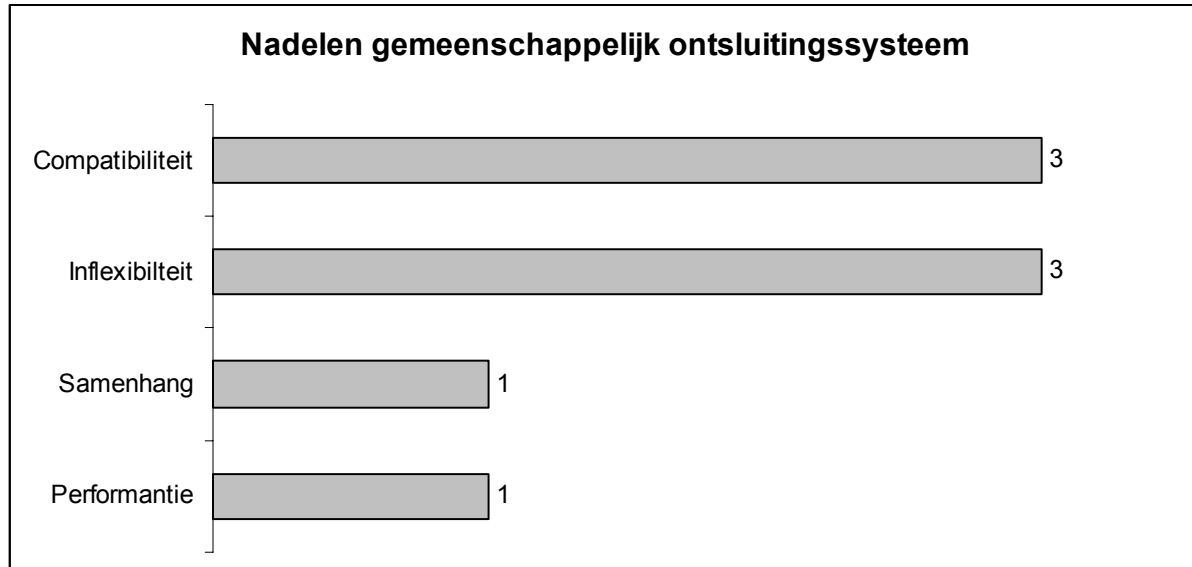
De doelstelling van VHZ III behelst de technische, organisatorisch en strategische haalbaarheid van een innovatieve, (de)centrale applicatie voor de ontsluiting van mondelinge bronnen en geluidsarchieven in kaart te brengen. Eén van de mogelijke gevolgen die dergelijke applicatie met zich meebrengt, is het opzetten van een gezamenlijk systeem waarbij mondeling historische bronnen over alle archief- en erfgoedinstellingen heen in één zoekscherm worden gekoppeld. Eventueel kan ook de gezamenlijke opslag van en controle over de materialen overwogen worden.

Zowel binnen de archief- als erfgoedsector bestaat grote eensgezindheid over een maximale beschikbaarheid en consulteerbaarheid van mondeling historische bronnen. De interesse voor een applicatie die de collecties van voornoemde instellingen samen doorzoekbaar

maakt, is dan ook behoorlijk groot bij de bevroagde instellingen. Enkel VRT en Stadsarchief Antwerpen gaven aan geen interesse te hebben voor dergelijk gemeenschappelijk ontsluitingssysteem. Het spreekt voor zich dat een gezamenlijke ontsluiting van mondeling historische bronnen in een gelijkaardig toepassingsysteem van metadata dient te voorzien. Het afstemmen van de onderlinge systemen of het hanteren van een (inter)nationale standaard is dan ook een absolute voorwaarde voor een gezamenlijk ontsluitingssysteem.



Dergelijk centraal ontsluitingssysteem is volgens de bevroagde instellingen kostenefficiënt: zowel ontwikkelings- als operationele kosten kunnen door meerdere organisaties gedragen worden. Ook de uniformiteit (via standaardisatie metadata), uitwisselbaarheid van bronnen en het delen van kennis op het gebied van bewaring en ontsluiting worden als mogelijke voordelen aangegeven. Voor de institutionele en eindgebruikers betekent een bredere toegankelijkheid eveneens een belangrijk aspect in de ontsluiting van historisch materiaal.



Wegens de hoge interesse voor een gemeenschappelijk ontsluitingssysteem zijn er nauwelijks gepercipieerde nadelen aan de applicatie verbonden. Enkele instellingen stellen zich niettemin vragen over de vermeende comptabiliteit met het bestaande beheersysteem: een eventuele overschakeling naar een nieuw systeem zal uiteraard flexibel en zonder al te grote complicaties dienen te verlopen. Bovendien moet een dergelijk grootschalig systeem inflexibiliteit en stugheid vermijden en dient het te worden ondersteund door een performante server. Tot slot kan een dergelijk systeem een gevaar betekenen voor de samenhang van de archiefcollectie.

4.6 Gezamenlijke opslag

Uit voorgaande is duidelijk gebleken dat de interesse voor een gemeenschappelijk ontsluitingssysteem groot is bij de bevroagde organisaties. Voor een model voor gemeenschappelijke opslag lopen de archief- en erfgoedinstellingen echter minder warm. Motieven daarvoor slaan terug op enkele algemene principes van de archivaliek.

Het herkomstbeginsel stelt dat elk archief als één geheel moet gezien worden en moet bewaard worden op naam van de archiefvormer. Archieven mogen niet met andere archieven worden vermengd omdat dit leidt tot het verlies van een pak historische informatie. Elk materiaal ontleent immers zijn belang en betekenis aan zijn band met de archiefvormer. Het structuurbeginsel bepaalt dan weer dat de organisch en historisch gegroeide structuur en ordening van een archief gerespecteerd dient te worden. Dit impliceert onder andere dat documenten niet uit hun context mogen worden weggenomen, archieven niet ontmanteld mogen worden om op thematische wijze te ordenen, documenten niet bij een archief mogen

gevoegd worden waarin ze van nature niet thuishoren en de structuur van het archief en de samenhang tussen de verschillende bronnen bewaard dient te blijven [7].

Een groot aantal instellingen wenst geen gemeenschappelijk opslagsysteem omdat in dat geval de context waarin de bron ontstaan en gearchiveerd is, dreigt te verdwijnen. Hoewel relatief marginaal in de volledige collectie kunnen – net zoals andere types van materialen – ook mondeling historische bronnen niet geïsoleerd worden van hun context want het is net deze context die de opname, bewaring én ontsluiting op bepaalde wijze mee vorm geeft.

Een tweede – meer impliciet aangegeven – reden die achter deze terughoudendheid moet gezocht worden, betreft ongetwijfeld de notie ‘institutionele controle’. Het blijken vooral de grotere instellingen te zijn die zich het meest categoriek afzetten van een centraal opslagmodel. Individuele en decentrale opslag laat toe volledige controle te behouden over de eigen collectie. Dergelijke controle wordt bij een aantal instellingen belangrijk geacht teneinde bevoegdheid en beslissingsmacht over het aanwenden van de collectie op institutioneel niveau te behouden.

4.7 Krachtlijnen en beschouwingen

Bij wijze van besluit bespreken we de huidige situatie inzake bewaring en ontsluiting van mondelinge historische bronnen in Vlaanderen via het gebruik van een SWOT-analyse. Dit instrument brengt de sterktes (strengths), zwaktes (weaknesses), kansen (opportunities) en bedreigingen (threats) in kaart. Deze vier kwadranten worden ingevuld aan de hand van een zogenaamde environmental scan (omgevingsanalyse). De omgevingsanalyse in het kader van dit rapport bestond uit de schriftelijke vragenlijst, gesprekken met diverse archief- en erfgoedinstellingen en kennisname van bestaande literatuur en projecten. In een SWOT-analyse is het gebruikelijk om de interne factoren te rangschikken onder de noemers ‘sterktes’ en ‘zwaktes’ terwijl externe factoren beschreven worden onder de noemers ‘kansen’ en ‘bedreigingen’. Per kwadrant werden slechts een beperkt aantal elementen opgenomen louter om de overzichtelijkheid van de factoren te bewaren. De begeleidende tekst gaat evenwel dieper in op een aantal bijkomende aspecten.

Het feit dat de bevraagde actoren een systematisch en gestructureerd archiefbeleid voeren met de bedoeling ontsluiting van historische materialen te maximaliseren kan ongetwijfeld als een van de grote **sterktes** worden ervaren. De grote mate van ontsluiting leidt er toe dat mondeling historische bronnen snel en efficiënt kunnen teruggevonden worden door interne medewerkers of extern geïnteresseerden.

Binnen de archief- en erfgoedsector bestaat een sterke behoefte aan een gestandaardiseerd model voor het annoteren van historisch materiaal, zowel voor het toekennen van formele als inhoudelijke beschrijvingen. De vigerende ISAD(G)-standaard voldoet immers niet aan de vereisten voor een algemeen Vlaams beschrijvingsmodel. De huidige tendens om collecties summier en minder diepgaand te beschrijven, dient weerspiegeld te worden in een complementaire en modulaire metadatastandaard.

Samen met de grote vraag naar gestandaardiseerde metadata vertoont de archief- en erfgoedsector een sterke interesse voor een gemeenschappelijk ontsluitingssysteem voor mondelinge historische bronnen. Het voorstel om de verschillende historische collecties gezamenlijk beschikbaar stellen, diverse databanken te koppelen en deze via één zoekscherm te ontsluiten, wordt met veel enthousiasme onthaald door de bevroegde actoren.

De reeds bestaande (voorzichtige) samenwerking tussen de diverse archief- en erfgoedinstellingen zoals in het kader van specifieke projecten of catalogi (ODIS, Archiefbank Vlaanderen...) vormen alvast een vruchtbare bodem voor de toepassing van een uniform metadata-model en de ontwikkeling van een gemeenschappelijk ontsluitingssysteem.

Bij **zwaktes** detecteren we de gebrekkige situatie inzake digitalisering van mondelinge historische bronnen. Het omzetten van historisch gegroeide collecties analoge materialen naar digitale vorm wordt wegens de specifieke inspanning (tijd, personeel, budget...) zeker niet bij de grote archiefinstellingen als prioritair aangegeven. Het niet digitaal beschikbaar zijn van een groot aandeel van mondelinge historische bronnen in Vlaanderen is evenwel geen optimaal startpunt voor de ontwikkeling van een gemeenschappelijk digitaal ontsluitingssysteem.

Het gebrek aan een uniform metadata-systeem vormt een tweede zwakte in het huidige beleid inzake bewaring en ontsluiting van mondelinge historische bronnen. Specifieke, projectmatige, inconsistente en/of afwezige metadata zorgen voor gebrekkige beschrijvingen, suboptimale ontsluiting en moeilijke samenwerking met andere instellingen en complexe integratie van beheerssystemen.

Een substantiële zwakte in het huidige archiveringsbeleid van mondeling historische bronnen betreft het gebrekkige beheer van de intellectuele en auteursrechtelijke bepalingen die erop gelden. De juridische situatie van de bronnen is veelal niet opgenomen in de metadata en vaak is er ook geen koppeling tussen contract en historische bron. Tot slot bevatten de overeenkomsten al te vaak geen regelingen omtrent verdere verspreiding of ontsluiting van de bronnen zodat de valorisatie en ontsluiting via het internet niet vanzelfsprekend wordt.

Tot slot wordt de Vlaamse archief- en erfgoedsector door een sterke fragmentatie en versnippering gekenmerkt. De talrijke en diverse instellingen hebben over de jaren heen eigen beheer- en ontsluitingssystemen opgezet en eigen praktijken inzake beschrijving opgebouwd zodat een gemeenschappelijk ontsluitingsbeleid niet vanzelfsprekend is. Bovendien leidt dergelijke decentralisatie tot het inefficiënt aanwenden van middelen.

Een enorme **kans** is weggelegd voor maximale beschikbaarheid, consulteerbaarheid, toegang en hergebruik. Gemeenschappelijke ontsluiting via het internet maakt Vlaanderens unieke erfgoed beter zichtbaar voor diverse doelgroepen en doeleinden: wetenschappelijk, educatief, cultureel, media... Een vlot toegankelijk, gebruiksvriendelijk, performant, gestructureerd en geactualiseerd ontsluitingssysteem kan in belangrijke mate bijdragen tot de verdere valorisatie en democratisering van het rijke orale verleden.

Een dergelijk systeem kan eveneens een oplossing aanreiken voor de huidige versnippering inzake het gehanteerd metadatamodel en kan mogelijks leiden tot een gedeeld model dat aan eenieders wensen en noden beantwoordt, eventueel geïnspireerd door internationale praktijken. De huidige internationale archiefstandaard beantwoordt immers niet ten volle aan de vereisten inzake inhoudelijke en formele beschrijvingen. Standaarden dienen zich eveneens aan te passen aan de huidige trend om historische materialen summier en economisch te beschrijven.

Een dergelijke gemeenschappelijke minimale dataset moeten eveneens kansen bieden voor een betere samenwerking tussen de diverse archief- en erfgoedactoren, geconcretiseerd in het eenvoudig uitwisselen van bronnen. Krachtenbundeling kan opnieuw leiden tot een betere zichtbaarheid van de waardevolle collectie en de representatie van de sector in het educatieve, socio-culturele en politieke veld versterken.

De ontwikkeling van een gemeenschappelijk ontsluitingssysteem biedt een vruchtbare bodem voor het ontwikkelen van nieuwe methodes voor (vol)automatische toekenning van

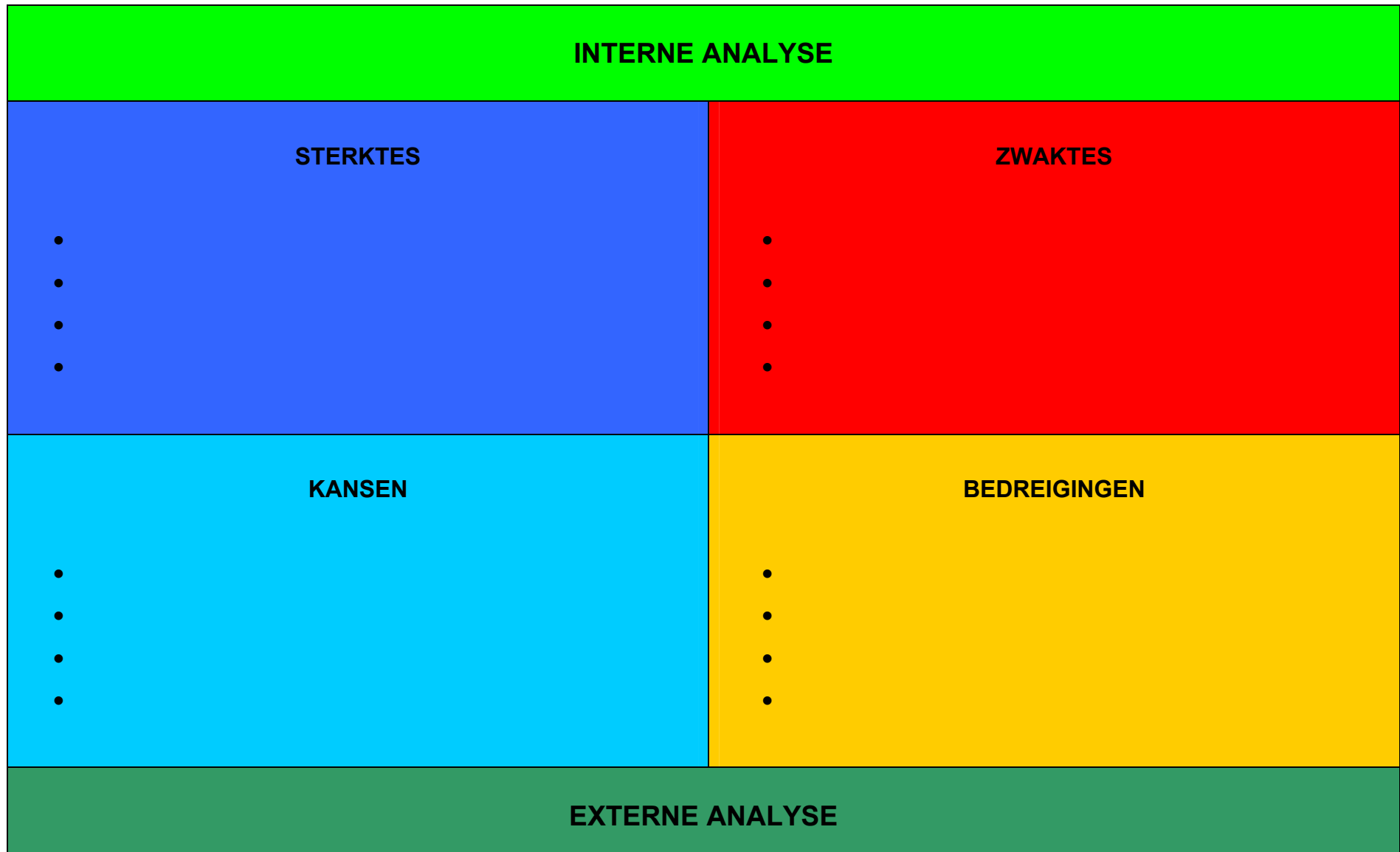
metadata, een probleem waar men binnen de eigen systemen mee geconfronteerd wordt. Door eliminatie van het tijdrovend omzetten van bestandsformaten en dragers kan eveneens de workflow binnen de bevraagde organisaties geoptimaliseerd worden.

Bij de mogelijke **bedreigingen** identificeren we de mogelijke incompatibiliteit van de huidige systemen. Het hanteren van een gemeenschappelijk ontsluitingssysteem impliceert de integratie van de huidige systemen in één globaal model. De afwezigheid van een uniform metadatamodel en de diversiteit van gebruikte bestandsformaten verhoogt het risico op een complex technisch integratieproces.

Ernstige aandacht dienen we ook te besteden aan het behoud van de samenhang van de verschillende historische bronnen en de context waarin ze opgenomen, bewaard en ontsloten zijn. Het reduceren van historische bronnen tot consulteerbare bestanden doet afbreuk aan de historische waarde ervan, die mede beïnvloed wordt door de context en samenhang met andere collectiestukken.

Bovendien dient de exacte rol van de betrokken actoren nauwkeurig vastgelegd te worden en moet de onduidelijkheid over de kostenstructuur van het digitaal archiveren en ontsluiten uitgeklaard worden. Er zijn met andere woorden financiële afspraken vereist met respect voor de relatieve positie van de diverse instellingen in het erfgoedveld.

Een andere techn(olog)ische uitdaging betreft de continue ondersteuning en aanpassing van het gemeenschappelijke systeem. Systemen dienen zich aan te passen aan de veranderende omstandigheden en daardoor bestaat het gevaar op verstarring van het systeem. Het te ontwikkelen systeem moet toekomstgericht zijn en bovenal werkbaar blijven. In het ideale geval kan een zelfstandig onderzoekscentrum of een onderzoekscel behorend tot een bestaande structuur de continue ondersteuning en actualisering van ontsluitingssystemen voor de socio-culturele sector begeleiden.



5 Internationale cases

5.1 *Instituut voor Beeld en Geluid in Nederland*

5.1.1 Historiek

De stichting Nederlands Instituut voor Beeld en Geluid (B&G) is opgericht in 1997. B&G brengt verschillende voormalige archieven samen, zoals het RTV-Archief Publiek Omroepen, het Filmarchief van de Rijksvoorlichtingsdienst, het Omroepmuseum, Film en Wetenschap, Smalfilmmuseum en verschillende particuliere collecties. Beeld en Geluid bezit ca. 700.000 uur beeld en geluid (waarvan een beperkt aandeel gedigitaliseerd), 2 miljoen foto's en 20.000 voorwerpen.

De doelstelling van B&G is vierledig:

- B&G wil het bedrijfsarchief zijn van de Nederlandse omroepen
- B&G wil het audiovisueel cultureel erfgoed van Nederland bewaren, beheren en ontsluiten
- B&G wil dit erfgoed ontsluiten voor het grote publiek via een nieuwe 'Media Experience' (interactief media-museum)
- B&G wil een kennisinstituut zijn inzake archivering van audiovisueel materiaal

B&G is opgericht om het duurzaam behoud van het Nederlandse nationale audiovisuele erfgoed te garanderen en het toegankelijk te maken voor zoveel mogelijk gebruikers: professionals, het onderwijs en het grote publiek.

Voor professionals wordt het digitale archief van B&G toegankelijk gemaakt via IMMix.

5.1.2 Infrastructuur

B&G is momenteel gehuisvest in het Mediapark in Hilversum. Eind 2006 werd de nieuwbouw betrokken, een indrukwekkend gebouw dat moet toelaten alle doelstellingen (zowel qua conservering als qua ontsluiting naar het grote publiek) te realiseren. Het gebouw bevat niet alleen kantoor- en geacclimatiseerde archief ruimte maar ook een interactief bezoekerscentrum (Media Experience) dat jaarlijks tussen 100.000 en 150.000 bezoekers moet lokken.

B&G wil op termijn de audiovisuele collectie migreren naar een digitaal archief. Het instituut beschikt nu reeds over een digitale collectie van 10.000 uur.

B&G, dat in totaal ongeveer 200 mensen tewerkstelt, wordt gesubsidieerd door het Nederlands Ministerie van Onderwijs, Cultuur en Wetenschap. Uit de mediabegroting wordt jaarlijks een structurele subsidie van ruim 10 miljoen euro voorzien, voor specifieke projecten verwerft B&G additionele subsidies. Daarnaast wordt ook eigen inkomsten gegenereerd uit dienstverlening en licenties op eigen collecties.

5.1.3 Gebruikers

De collectie van B&G wordt in de eerste plaats gebruikt door programmamakers om nieuwe programma's te produceren. Daartoe wordt de multimediacatalogus IMMix ontwikkeld.

Verder is de collectie een onderzoeksbron voor studenten en wetenschappers. Ongeveer 3/4 van het bewaarde audiovisueel materiaal in Nederland bevindt zich immers bij B&G.

B&G wil het gebruik van de collectie stimuleren in het Nederlandse onderwijs. Volgens hen kan het audiovisueel materiaal een waardevolle aanvulling zijn op de traditionele lesmethoden. In het bezoekerscentrum zijn ook speciale workshops en schoolprogramma's voorzien.

Ten slotte richt het archief van B&G zich tot het grote publiek. Op de website van B&G kan het publiek bepaalde items uit de collectie bekijken of downloaden. Ook het bezoekerscentrum moet het ruime publiek aanspreken.

5.1.4 Huidige werking

Het archief van B&G heeft enerzijds de bedoeling om het Nederlands audiovisueel erfgoed te conserveren, te beheren en te ontsluiten naar een breed publiek. Anderzijds heeft B&G via de samenwerkingsovereenkomsten met de Publieke Omroepen een belangrijke taak als archief en bron voor audiovisuele producenten.

In het archief van B&G wordt volgend materiaal opgeslagen:

- Televisie: vooral programma's van openbare omroepen, slechts sporadisch ook van commerciële omroepen
- Radio, muziek, geluid (onder meer een boel 'verdwenen geluiden'), polygoon (vooral bioscoopjournaals), foto, objecten, bibliotheek (omroep en media), papieren collecties (persoonsarchieven van radio- en televisiepersoonlijkheden), biografieën

Niet al dit materiaal is reeds gedigitaliseerd. Wanneer het (nog) niet online beschikbaar is, kan het materiaal bij B&G geraadpleegd worden (tegen betaling). Tot de opening van Media Experience kregen professionele klanten voorrang, nu heeft ook het grote publiek vlotte toegang.

Werkwijze ontsluiting (trefwoorden, thesaurus, velden, ... kortom metadata)

In de context van deze state-of-the-art is het vooral belangrijk om het digitaal archief dat in de eerste plaats bedoeld is voor mediaprofessionelen meer in detail te bespreken.

In onderstaande beschrijving wordt dan ook het opzet van dit digitaal archief besproken. Het zal enerzijds gevoed worden via de import van de dagelijkse aanwas van televisie- en radio-uitzendingen vanuit De Digitale Voorziening (DDV). DDV is een systeem dat omroepen toelaat om de productie, distributie, uitzending en archivering op een volledig digitale manier te bewerkstelligen. Tevens wordt het digitaal archief gevoed vanuit de encoding van historische collecties die in het bezit van B&G zijn. iMMix is het ontsluitingssysteem van dit digitaal archief. In iMMix zullen de metadata van DDV geïmporteerd en de metadata van historische collecties geëncodeerd worden. De digitalisering van historische content en metadata zal hoofdzakelijk on-demand gebeuren. De verschillende gebruikers van iMMix zullen ook een verschillende front-end hebben. In eerste instantie richt B&G zich op de professionele gebruiker. In onderstaand schema wordt deze structuur visueel duidelijk gemaakt.



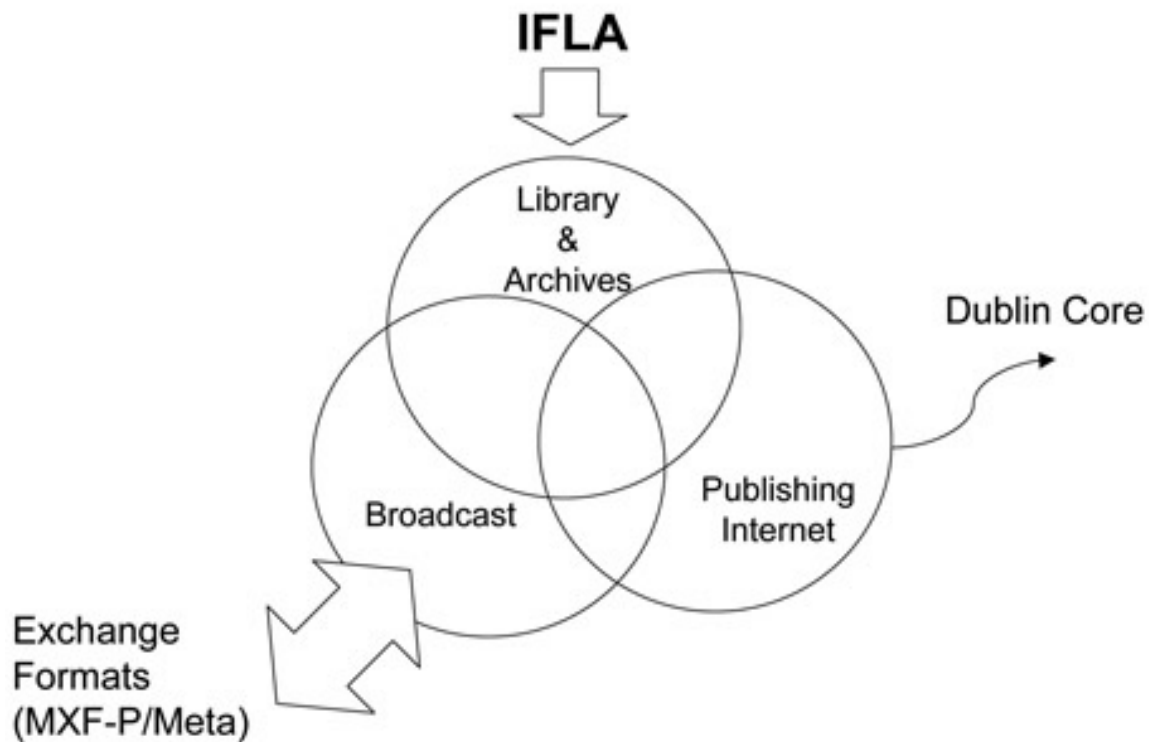
Figuur: Digitale workflow

De front-end voor de professionele gebruiker biedt de mogelijkheid om 'gericht, compleet, uitgebreid en gestructureerd' te zoeken. De front-end zal toelaten om vrij te zoeken (via een zoekmachine) met Booleaanse operatoren en om specifiek te zoeken in bepaalde velden (bijvoorbeeld de titel of de datum), in de index of in de thesaurus. Zoekmogelijkheden zullen variëren van eenvoudig, via standaard, tot geavanceerd (waar bijvoorbeeld de mogelijkheid aanwezig is om de rol van een persoon te definiëren: Mr. X als presentator, of als acteur, of als geïnterviewde, enz.).

De resultatenlijst van een bepaalde zoekopdracht zal volgende onderdelen bevatten: drager of formaat, collectie, genre, titel, uitzenddatum, zendgemachtigde, tijdsduur en trefwoorden. Door op een titel te klikken kan detailinformatie bekomen worden.

Bij het zoekresultaat zullen eveneens keyframes te zien zijn. De keyframes zijn gelinkt aan videofragmenten die via een ingebouwde player kunnen afgespeeld worden. Via iMMix kan ook specifiek op shots en quotes gezocht worden.

IMMix maakt gebruik van een metadatamodel dat gebaseerd is op de volgende componenten:

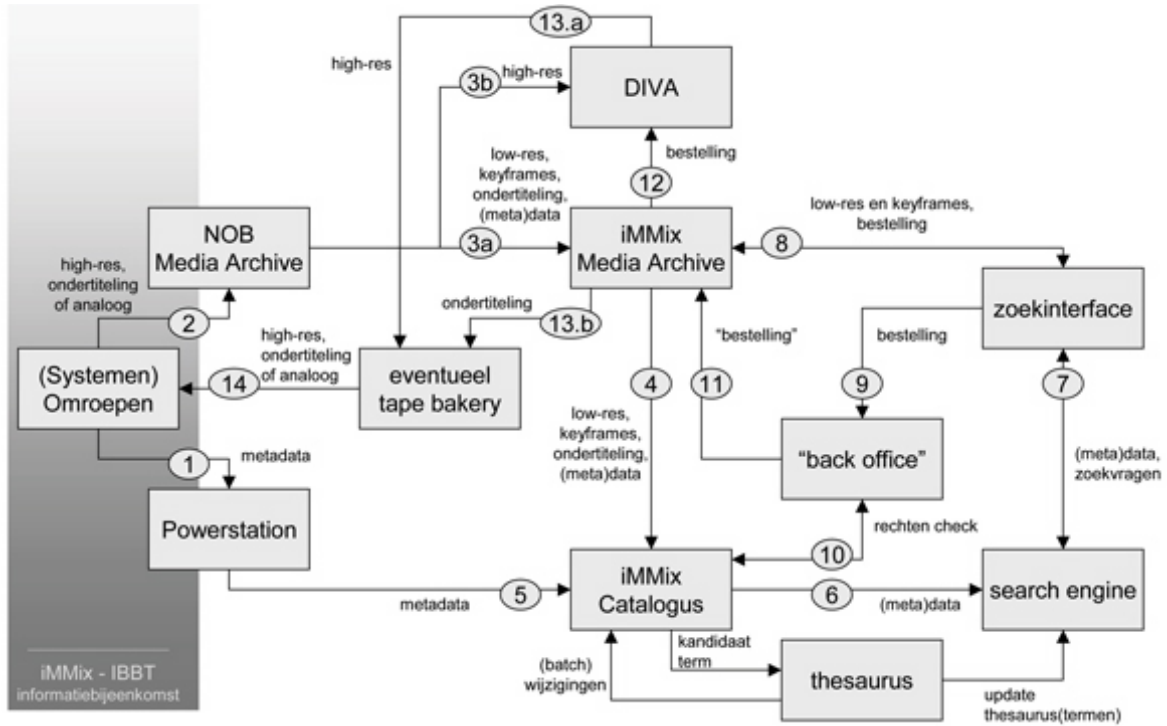


Figuur: Metadatamodel IMMix

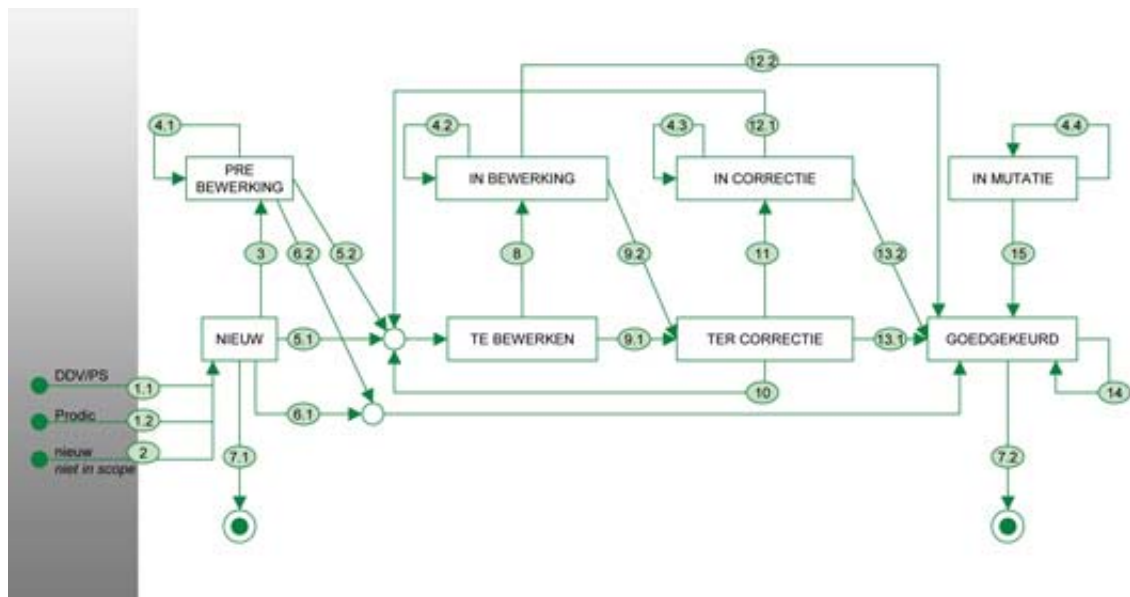
5.1.5 Digitaliseringsplannen

B&G heeft reeds een digitaal archief en zal gedurende de volgende jaren dit digitale archief uitbreiden. B&G voorziet een jaarlijkse instroom van ongeveer 8000 uur TV van publieke omroepen, 2000 uur TV van commerciële omroepen, 3500 uur TV via digitalisering on-demand, 24000 uur radio en een nog onbepaalde hoeveelheid foto's en webpagina's. In totaal gaat het om een aanwas van meer dan 250 TB per jaar.

In onderstaande schema's worden de workflow van DDV, iMMix, en de catalogus visueel uiteengezet.



Figuur: Workflow DDV en IMMIX



Figuur: Workflow catalogus

5.2 CATCH projecten (Continuous Access To Cultural Heritage)

5.2.1 Onderzoeksopzet

Nederlands cultureel erfgoed heeft zowel een enorme symbolische als economische waarde. Het Cultureel erfgoed van een land behoort immers toe aan de hele bevolking en speelt een rol in diverse domeinen (bvb. toerisme, educatie, onderzoek, ...). Om historische redenen zijn objecten en hele collecties echter in tal van verschillende erfgoedinstellingen terechtgekomen, wat hun toegankelijkheid (in tijd en ruimte) sterk vermindert. Digitalisering van al dit materiaal zorgt ervoor dat het altijd toegankelijk is (geen beperkingen in tijd/ruimte) en creëert zo tal van mogelijkheden (bvb. in kader van onderzoek, educatie, ...). Maar ondanks vele investeringen en inspanningen bvb. van overheidswege, zijn er een aantal obstakels, die deze vooruitgang afremmen:

- Het digitaliseringsproces verloopt traag en is bovendien erg duur.
- Er bestaan heel veel gefragmenteerde, onafhankelijke collecties, die elk in een eigen database zitten en op geen enkele manier aan elkaar gelinkt worden.
- Zelfs al moesten deze collecties gelinkt worden, dan zouden er obstakels blijven bestaan omdat het om zeer diverse objecten gaat (en dus zeer diverse schema's en metadatasystemen)
- Er moeten geautomatiseerde manieren gezocht worden om én de data te linken én tegelijkertijd te zorgen dat de combinatie van bestaande data met bestaande kennis, tot nieuwe kennis leidt (kennisverrijking).
- De resultaten van een zoekopdracht moeten op een dusdanige manier gepresenteerd worden, dat ze tegemoet komen aan de precieze noden van de persoon die op zoek is naar informatie (personalisering).

Vanuit de erfgoedinstellingen is er aldus een sterke nood aan nieuwe oplossingen met betrekking tot 'access':

- In welke mate is het mogelijk om nieuwe, innovatieve tools te ontwikkelen die kennis en culturele objecten verbinden, en om de gedigitaliseerde culturele objecten (die vandaag de dag virtueel verspreid zijn) te integreren om zo de toegankelijkheid van en de interactie met het cultureel erfgoed te verbeteren? (zowel naar professionele als naar eindgebruikers toe)
- Kunnen we wetenschappelijk relevante methoden ontwikkelen om nieuwe, fundamentele en toegepaste kennis over deze processen en hun IT-gebaseerde oplossingen te verwerven?

In eerste instantie is CATCH gericht op de specifieke situatie in Nederland. Uiteraard hoopt het onderzoeksprogramma met de ontwikkeling van nieuwe methoden en technieken ook iets bij te kunnen dragen aan het erfgoedonderzoek op internationaal niveau.

5.2.2 Financiering

De financiering gebeurt via het onderzoeksprogramma van NWO (de Nederlandse organisatie voor wetenschappelijk onderzoek) dat

- methoden en technieken wil ontwikkelen waarmee collectiebeheerders hun digitale collecties beter toegankelijk kunnen maken,
- mogelijkheden wil bieden om nieuwe verbanden en relaties tussen verschillende collecties zichtbaar te maken,
- de interactie tussen gebruikers en bronnen wil optimaliseren. (zowel professionele gebruikers als bvb. onderzoekers, journalisten, leraren, het grote publiek, ... > eindgebruikers)

Het volledige CATCH plan is gestart in 2004 met een budget van 7,5 M euro. Het is de ambitie om dit bedrag te verhogen tot 12,5 M euro. Op dit moment lopen er 10 CATCH onderzoeksprojecten. Elk van die projecten werken met kleine, toegewijde teams bestaande uit informaticaonderzoekers en mensen uit culturele erfgoedinstellingen om aldus kennisoverdracht te verbeteren [8].

5.2.3 Realisaties

Thema 1: semantische interoperabiliteit d.m.v. metadata

In de erfgoed- en archiveringsinstituten: groeiend besef m.b.t. de nood aan gestandaardiseerde metadata, afspraken m.b.t. structuur van beschrijvingen, terminologie, ... Op basis van o.a. de grote verschillen in de instituten onderling, de veranderende wensen van de gebruiker, ... gaat men er in CATCH vanuit dat de doelstelling van gestandaardiseerde, niet-ambigue metadata niet enkel kan bereikt worden door menselijke intellectuele arbeid. Binnen dit thema situeert zich dus de uitdaging om intelligente IT systemen te combineren met menselijke expertise en om tools en methoden te ontwikkelen die bronnen automatisch of semiautomatisch documenteren en beschrijven aan de hand van metadata. Deze manier van beschrijven wordt immers aanzien als een eerste voorwaarde om te komen tot semantische interoperabiliteit van de collecties.

Binnen dit thema lopen de volgende onderzoeksprojecten:

- **CHOICE:** CHarting the informatiOn landscape employing ContExt information [9]
Partners: B&G, Telematica Instituut, Max Planck Instituut, Vrije Universiteit Amsterdam
- **MuSeUM:** Multiple-collection Searching Using Metadata [10]
Partners: Universiteit van Amsterdam, Gemeentemuseum Den Haag, Rijksbureau voor Kunsthistorische Documentatie, Stadsarchieven van Rotterdam
- **STITCH:** Semantic Interoperability to Access Cultural Heritage [11]
Partners: Nationale Bibliotheek van Nederland, Vrije Universiteit Amsterdam, Max Planck Instituut

Thema 2: Kennisverrijking d.m.v. geautomatiseerde analyses

Het CATCH programma wil onderzoek doen naar en kennis opdoen over geautomatiseerde kennisverrijkingstools. Enerzijds gaat het om een groep van tools die de experts moet helpen en bijstaan, anderzijds om een groep van tools die volledig geautomatiseerde analyses mogelijk maakt. Men beoogt bovendien een kennisverrijking op niveau van de verschillende mediatypes die opgenomen zijn binnen CATCH (tekst, beeldmateriaal, handgeschreven bronnen, archeologische objecten, ...). De methoden die binnen dit thema ontwikkeld zullen worden, zullen gebaseerd zijn op uitvoerig empirisch onderzoek. Ze zullen beoordeeld en geëvalueerd worden in termen van accuraatheid, snelheid, ...

Binnen dit thema lopen de volgende onderzoeksprojecten:

- **CHoral:** Access to oral history [12]
Partners: Universiteit Twente, Stadsarchieven van Rotterdam, de Erasmus Universiteit van Rotterdam en Radio Rijnmond.
- **MITCH:** Mining for Information in Texts from the Cultural Heritage [13]
Partners: Nationaal Natuurhistorisch Museum, Universiteit van Tilburg.
- **MunCH:** Multimedia aNalysis for Cultural Heritage [14]
Partners: Universiteit van Amsterdam, Vrije Universiteit van Amsterdam, B&G, Vereniging Digitaal Erfgoed Nederland

- **RICH:** Reading Images in the Cultural Heritage [15]
Partners: Rijksdienst voor het Oudheidkundig Bodemonderzoek, Universiteit van Maastricht
- **SCRATCH:** SCRipt Analysis for the Cultural Heritage [16]
Partners: Nationaal Archief, Universiteit van Groningen

Thema 3: personalisering d.m.v. presentatie

In het kader van dit onderzoeksthema beoogt CATCH nieuwe kennis te vergaren m.b.t. de selectie van informatie, de automatische generatie van presentaties en de mogelijkheden tot adaptatie en personalisering. Men gaat er immers van uit dat nieuwe ontwikkelingen in het IT-domein tal van mogelijkheden bieden op het vlak van de personalisering van presentaties, maar dat deze mogelijkheden tot op vandaag nauwelijks benut worden.

Binnen dit thema lopen de volgende onderzoeksprojecten:

- **CHIP:** Cultural Heritage Information Personalization [17]
Partners: Rijksmuseum, Technische Universiteit Eindhoven, Telematica Instituut
- **WITCHCRAFT:** What is Topical in Cultural Heritage: Content-based Retrieval Among Folksong Tunes [18]
Partners: Universiteit Utrecht, Meertens Instituut, Theater Instituut Nederland

5.2.4 CHOICE (uitgebreid)

Vanuit thema 1 (semantische interoperabiliteit d.m.v. metadata) ligt de nadruk hier op het semi-automatisch semantisch annoteren en het uitbuiten van contextuele informatie. Semantische annotatie impliceert de annotatie van gearchiveerde objecten -zoals audio, video, beelden en boeken- door middel van semantische categorieën gekozen uit een gestandaardiseerde metadata verzameling (domein specifieke thesauri en/of ontologieën). Het gebruik van dergelijke semantische annotatie laat immers verbeterde zoekmethodes toe binnenin een collectie. Zo kan bijvoorbeeld een foto geannoteerd worden met de semantische categorie "schip" (met de betekenis "deel van een kerk") uit een specifieke domeinontologie waardoor bij het zoeken naar "kerkschip" geen foto's van vaartuigen zullen gevonden worden. De meeste thesauri en ontologieën laten ook 'broader'/'narrower' relaties toe waardoor het mogelijk wordt om meer generieke of gespecialiseerde semantische

zoekopdrachten uit te voeren: bijvoorbeeld foto's van "zijbeuken", een 'narrower' semantische categorie van "kerkschepen".

De gebruikte use case in dit project is het videoarchief van Beeld & Geluid. Het objectief is aan te tonen dat het linken van een aantal divergente thesauri/ontologieën aan een video annotatie applicatie zowel het manuele semantische annotatieproces als het daaropvolgende zoekproces kan verbeteren. Om het manuele annotatieproces te versnellen zal hiervoor gebruik gemaakt worden van 'natuurlijke taal' procestechnieken om semantische categorieekandidaten te genereren die op hun beurt zullen mappen op de geselecteerde thesauri, gegenereerd uit de (tekstuele) context informatie. Binnen audiovisuele objecten, is de bijkomende contextinformatie bijvoorbeeld: transcripties, externe reviews, uitzendschema's, kijkerwaardering en filmprijzen.

De proprietaire thesaurus van Beeld & Geluid wordt daarvoor volledig herwerkt en uitgebreid tot een semantische ontologie. Verder wordt een facet gebaseerde thesaurus browser ontworpen om het manuele annotatieproces te vereenvoudigen. Het intelligent uitwerken van de indexeringsmodule door middel van 'natuurlijke taal' procestechnieken gebeurt aan de hand van het verder ontwikkelen en integreren van een aantal Open Source bibliotheken: de Apolda plugin (voor keyword extractie) en het CARROT algoritme (voor keyword ranking). De (vernieuwde) GTAA thesaurus bezit volgende facetten: Onderwerp, Persoon, Naam, Locatie, Genre en Producent. Volgende standaardconstructies worden binnen de GTAA thesaurus ondersteund: broader term/narrower term, related term, use/use for, scope note, qualifiers. Er is ook nog een niet-standaard constructie aanwezig: Categorie (binnenin het Onderwerp facet). De GTAA thesaurus werd aldus volledig beschreven als een SKOS-model [19]. Het ranking algoritme CARROT is ontworpen als een onafhankelijke, tunable service gebaseerd op een Open Source RDF store en een RDF query language. De volledige architectuur wordt opgebouwd uit Open Source componenten: de basisarchitectuur, het GATE-raamwerk [20], wordt met volgende plugins uitgebreid:

- Apolda (semantic annotation, plugin extra ontologies) [21]
- Lucene (search engine) [22]
- Weka (data mining) [23]
- Treetagger (lemma) [24]

Ook de backend RDF triple store is een Open Source implementatie: Sesame [25]

5.2.5 CHORAL (uitgebreid)

Vanuit thema 2 (kennisverrijking d.m.v. geautomatiseerde analyses) ligt de nadruk hier op ontsluiting van mondelinge historische collecties door middel van ‘Spoken Document Retrieval’. De gebruikte use case in dit project zijn auditieve bronnen met getuigenissen en verklaringen die historische gebeurtenissen uit de 2^{de} wereldoorlog weergeven. Dit audio mining project bestudeert en integreert automatische technieken voor spraakherkenning met informatie retrieval technieken.

Door middel van spraakherkenning worden transcripties gegenereerd van gesproken documenten (audio en video). Binnenin de transcripties worden woorden geassocieerd met tijdstempels die worden gebruikt om de index te bouwen zodat het mogelijk wordt om deze audiobestanden te doorzoeken tot op fragment niveau. Als core dataset wordt het uitzendarchief van het regionale radiostation Radio Rijnmond gebruikt. In combinatie met deze uitgezonden programma’s worden ook andere opnames en achtergrondteksten uit het stadsarchief van Rotterdam gebruikt als metadata. De huidige spraakherkenning software voor nieuwsuitzendingen wordt uitgebreid met adaptieve spraak- en taalmodellen, alsook met metadata extractie technologieën. Verder wordt ook de koppeling onderzocht tussen de auditieve data en de gerelateerde tekstuele records.

Als initiële demonstrator werd ‘Radio Oranje’ uitgewerkt. Tijdens de Tweede Wereldoorlog werden toespraken van H.M. Koningin Wilhelmina, gericht aan het Nederlandse volk, uitgezonden op Radio Oranje. Van deze toespraken zijn indertijd volledige transcripties gemaakt, welke tesamen met de originele geluidsopnamen bewaard zijn gebleven. In het Radio Oranje project is een interface gemaakt die een gebruiker in staat stelt om deze transcripties online te doorzoeken en direct de bijbehorende audiofragmenten te beluisteren. Dit is mogelijk gemaakt met behulp van de ontwikkelde automatische technieken voor spraakherkenning. Hierdoor kan tevens ondertiteling synchroon met het geluid in beeld worden gebracht, wat het makkelijker maakt de historische toespraken te verstaan.

5.2.6 Bemerkingen

CHOICE:

- Op dit ogenblik is er nog geen “live” link met de huidige thesaurus. De omzetting van die thesaurus naar zijn semantische equivalent gebeurt “op aanvraag” als een nieuwe versie van de thesaurus aanwezig is. De omzetting naar RDF-triples duurt vooralsnog een paar uren. Er bevinden zich dan immers meer dan vijfhonderdduizend triples in de triple store.

- Er wordt gewerkt aan een integratie met 2 andere CATCH-projecten, CHoral [12] & MunCH [14], maar daarvoor moet er een GTAA API voorzien worden aangevuld met een aantal extra services. Verder zal er ook nog een repository (plus bijhorende mappings) voorzien worden voor alle lexicons.

CHORAL:

- De Nederlandstalige spraakherkenningssoftware heeft heel veel trainingsdata (+ 50 uur auditieve data) nodig om een goed werkend akoestisch model af te leveren. Er moet verder ook een uitspraakcatalogus en een taalmodel aangemaakt worden.
- Er wordt ook veel onderzoek gedaan in dit project naar vernieuwende manieren om de auditieve content voor te stellen en te doorbladeren: een gewone tijdslijn versus een interactieve tijdslijn, 'karaoke'-like ondertiteling.
- Er zal ook nog nagegaan worden hoe de essence (audio/video) kan gelinkt worden aan verschillende krantenartikels bijvoorbeeld om aldus een betere context en een completer beeld van de gebeurtenissen te schetsen aan de eindgebruiker. Door verschillende perspectieven met elkaar te combineren kan men tot nieuwe, betere inzichten komen en aan beter vergelijkend onderzoek doen.

5.3 *British Library (Sound Archive) in Verenigd Koninkrijk*

5.3.1 Historiek

Opricht in 1955 als het British Institute of Recorded Sound werd het in 1983 onderdeel van de British Library (Londen), met meer dan 150 miljoen items de belangrijkste bibliotheek in de wereld. De British Library werd in 1972 opgericht als nationale bibliotheek van het Verenigd Koninkrijk via de British Library Act.

Het British Sound Library Archive staat te boek als één van de grootste archieven van geluidsopnames in de wereld en huisvest in totaal meer dan drieëneenhalf miljoen items met inbegrip van een miljoen discs, 185.000 tapes en vele andere geluid- en video-opnames op 19^e-eeuwse cilinders, goed voor meer dan 550.000 uur geluidsmateriaal.

5.3.2 Infrastructuur

The Sound Archive is gesitueerd in de British Library in Londen, die in 2006 2140 mensen (in FTE's) tewerkstelde. De Library beschikte in 2006 over een budget van 136,9 miljoen pond waarvan 97,6 miljoen (ca. 71%) afkomstig van het Britse Departement voor Cultuur, Media en Sport (DCMS).

5.3.3 Gebruikers

De British Library stelt zich tot doel het Britse erfgoed zo ruim mogelijk te ontsluiten naar industrie, onderzoekers, studenten... toe. Elk jaar bezoeken ongeveer 400.000 geïnteresseerden de leeszaal en genereert de online catalogus ruim zes miljoen zoekopdrachten. Het Sound Archive biedt publieke toegang tot een ruime waaier van specialistische publicaties zoals boeken, vaktijdschriften en wetenschappelijke literatuur over elk aspect van opgenomen geluid.

De catalogus is online doorzoekbaar, slechts een beperkt aantal geluidsopnames is ook meteen beluisterbaar in .wma-formaat. De 'Listening and Viewing Service' biedt echter de mogelijkheid om alle collectie-items ter plaatse te consulteren. Tot slot kan de gebruiker (particulier, commercieel of academisch) ook intekenen voor een aantal betalende diensten zoals het ontvangen van geluidskopieën.

5.3.4 Huidige werking

The Sound Archive verzamelt commerciële uitgaven uit het Verenigd Koninkrijk aangevuld met commerciële uitgaven van andere landen, radio-uitzendingen en private opnames zoals orale bronnen. Het archief fungeert als legaal depot voor non-print materiaal in het Verenigd Koninkrijk en dient het Britse muzikale erfgoed zo ruim mogelijk te verspreiden en te conserveren naar de toekomst toe.

Binnen het geluidsarchief onderscheiden we zes grote collectiedomeinen:

- Klassieke muziek: opnames uit de concerttraditie, opera, liturgische muziek en het prebarokke repertoire uit West-Europa
- Toneel en literatuur: registraties van live performances van toneelvoorstellingen en poëzie- en literatuurvoordrachten
- Mondelinge geschiedenis: audio- en video-opnames van interviews en story-telling
- Populaire muziek: exemplaren van commerciële uitgaven van diverse stijlen vanaf het begin van de 20^e eeuw (music hall) tot vandaag (rock, dance...), inclusief live registraties van concerten en festivals
- Natuurgeluiden: geluidsregistraties van alle dierenklassen uit elke zoogeografische regio
- Wereld- en traditionele muziek: opnames van traditionele ethnomuziek sinds het ontstaan van de geluidsopname

5.3.5 Digitaliseringsplannen

De British Library beschikt over een extensief en snel groeiend aantal van digitale items zoals gedigitaliseerde versies van origineel analoog materiaal (boeken, kranten, geluidsopnames...) maar ook origineel digitaal materiaal neemt een steeds belangrijkere plaats in. Zo bedraagt het groeiritme van het Sound Archive liefst 15TB per jaar... De Library stelt zich daarom tot doel in 2010 over 500TB opslagcapaciteit te beschikken.

Om deze digitale plannen te realiseren, opereert de British Library in sneltempo in een aantal projecten om de huidige analoge collectie te digitaliseren. Het ziet in de huidige technologische ontwikkelingen op het gebied van digitalisering mogelijkheden om [27]:

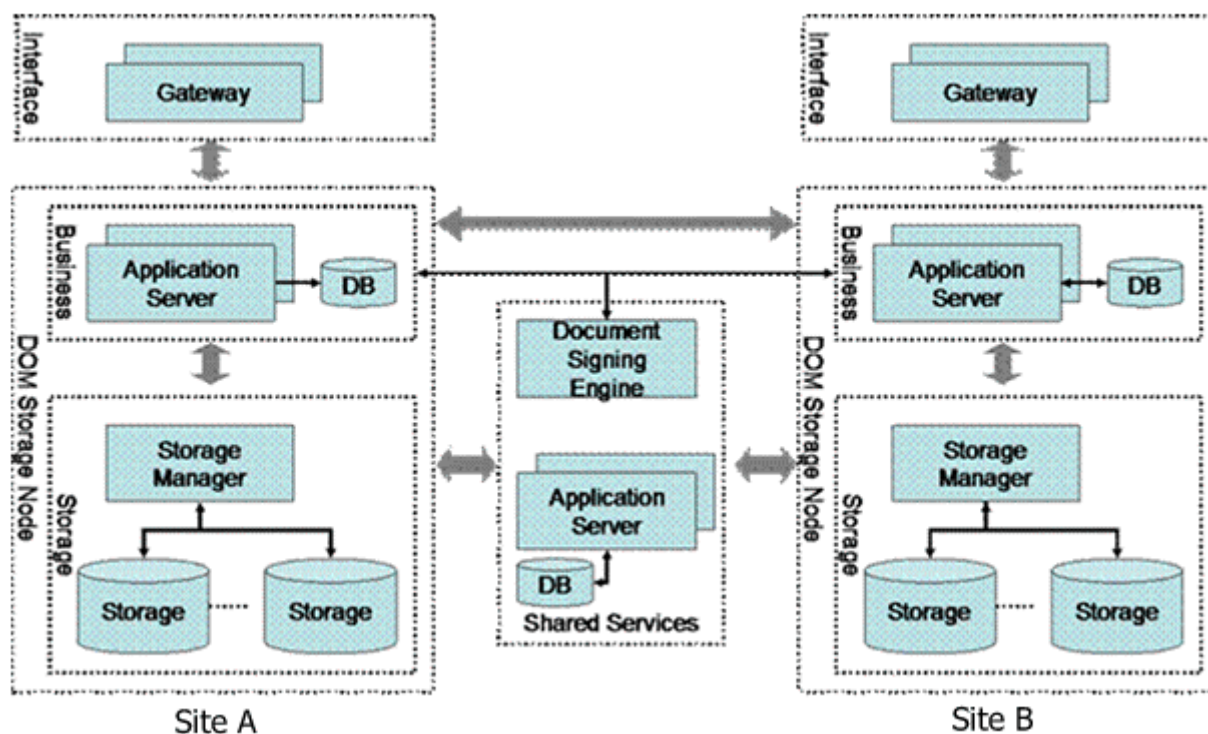
- de consultatie van de collectie te vergemakkelijken door een verbeterde netwerktoegang
- de (inter)nationale prioriteit op het gebied van een ruimere toegang tot geïntegreerde collectie voor cultureel, educatief en wetenschappelijk gebruik te weerspiegelen
- bij te dragen tot de conservatie van originele analoge materialen voor toekomstige gebruikers door middel van digitale surrogaten
- de bewaring van de analoge collectie geluidsopnames te verzekeren
- additionele inkomsten te verwerven uit producten die commercieel kunnen geëxploiteerd worden door partners of door de British Library zelf

5.3.6 Digital Object Management Programme

De Library zette in 2004 het Digital Object Management Programme (DOM) [28] op om technologische antwoorden te zoeken op de uitdagingen die digitale objecten stellen aan de archiefbeheerder. De bedoeling daarvan is een managementsysteem te ontwikkelen dat elk digitaal collectietype opslaat en bewaart, de gebruiker toegang tot de digitale collectie verschaft en deze bovendien gemakkelijk doorzoekbaar maakt.

Het technologische antwoord bestaat erin een meerschallig en flexibel subsysteem te ontwikkelen dat de integriteit en authenticiteit van de digitale objecten garandeert. Dat opslagsysteem dient echter aan twee sleutelprincipes te voldoen:

- onderscheid tussen de opslagdienst en de fysieke opslag zodat de fysieke opslag onafhankelijk van de rest van het systeem verloopt
- twee evenwaardige clusters in plaats van één systeem om de nodige flexibiliteit in te bouwen



Figuur: Storage subsystem

De betrokken onderzoekers kozen voor Microsoft.NET-software, in het bijzonder BizTalk 2004 en C#, omwille van de duidelijke productiviteits- en kostenvoordelen die aan deze tools verbonden blijken te zijn. DSpace en Fedora werden om voornoemde redenen niet weerhouden.

5.3.7 Archival Sound Recordings 1 & 2

5.3.7.1.1 Onderzoeksopzet

Deze projecten stellen zich tot doel applicaties te ontwikkelen die het mogelijk maken audiofragmenten te downloaden, op te slaan en te manipuleren met de bedoeling deze in een educatieve context te gebruiken. Dit baanbrekend project moet het mede mogelijk maken een online volledig doorzoekbare digitale audiobibliotheek te creëren. Indachtig de specifieke uitdaging van de British Library Sound Archive betreffende het preserveren van origineel analog geluidsmateriaal, werden passende materialen uit het rijke archief van de instelling geselecteerd.

Archival Sound Recording 1 (ASR1) [29] betreft een project waarbij ca. 12.000 audio-items (goed voor ongeveer 3900 uur) gedigitaliseerd en samengebracht worden in een centrale

database, toegankelijk via het internet. Het vervolgproject ASR2 zal daarbovenop ongeveer 4200 uur audiomateriaal online toegankelijk maken.

Beide projecten putten uit het rijke en diverse archief van de British Library Sound Archive en stellen zich tot doel elk negen collecties te digitaliseren zoals klassieke repertoires (van o.a. Bach, Mozart, Haydn, Beethoven en Brahms), discussies over kunst en literatuur (met o.a. Salman Rushdie, Richard Dawkins...), interviews met Holocaustoverlevenden, traditionele muziek uit Engeland en Uganda, interviews met Britse jazzartiesten... De digitale collecties bevatten met andere woorden materialen uit mondelinge geschiedenis, populaire en klassieke muziek, *soundscales* en *spoken word* opnames. Deze collecties worden tot slot verrijkt met fotomateriaal.

Bedoeling is dus academici, medewerkers en leraars uit colleges en universiteiten gratis toegang te verlenen zodat deze de geluidsfragmenten kunnen afspelen, downloaden en bewerken. Particuliere gebruikers kunnen echter enkel doorheen het archief bladeren, wegens auteursrechtelijke beperkingen is het afspelen van geluidsmateriaal niet mogelijk. Deze beperking geldt ook voor gebruikers buiten het Verenigd Koninkrijk.

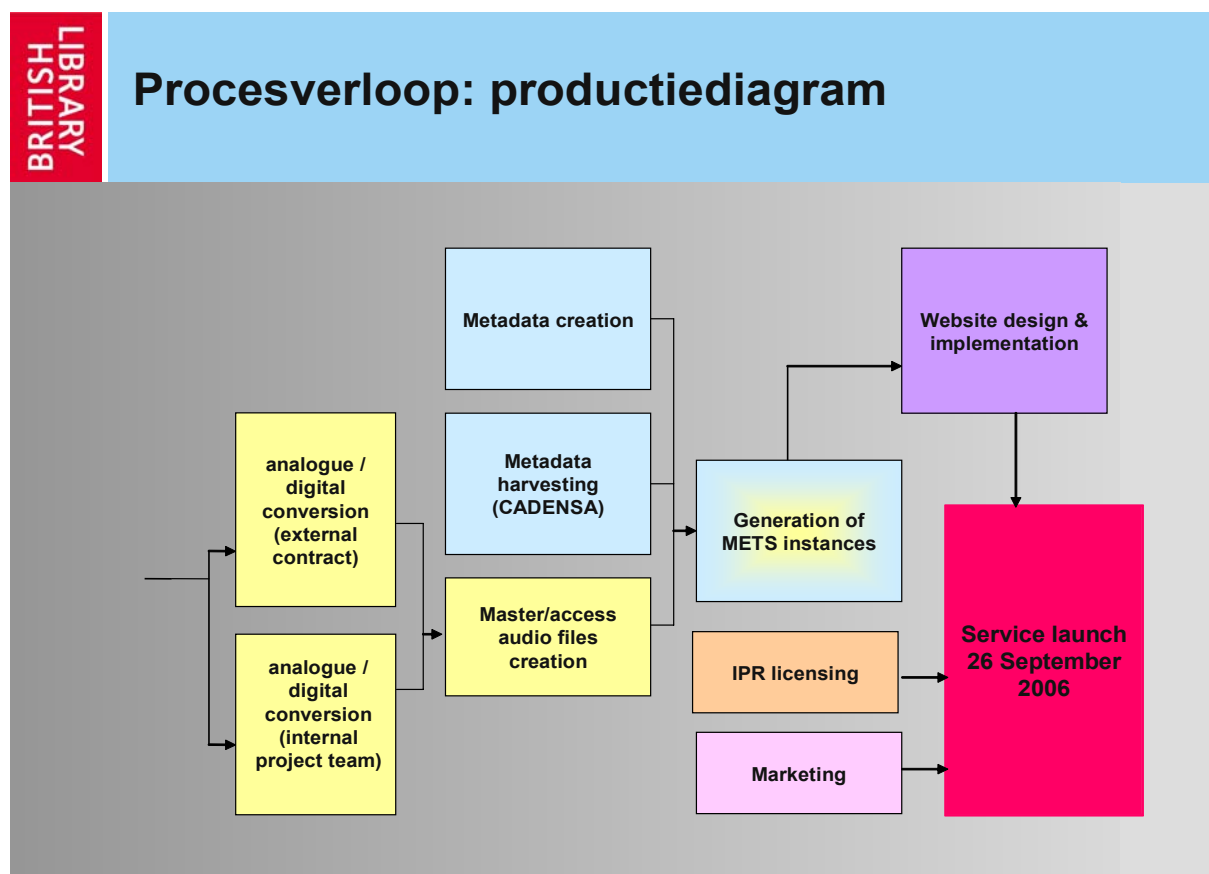
5.3.7.1.2 Financiering

Archival Sound Recording 1 en 2 is een samenwerking tussen British Library Learning (de educatieve dienst van de nationale bibliotheek) en JISC (Joint Information Systems Committee), dat het hoger onderwijs ondersteunt door strategische sturing en advies te verlenen op vlak van het implementeren van ICT.

ASR1 (looptijd: 01/04/2004 – 30/09/2006) werd mogelijk gemaakt door een investering van 1,1 miljoen pond vanwege JISC en past binnen een ruimere investering van 10 miljoen pond voor de digitalisering van geluid, bewegende beelden, kranten, parlementaire verslagen... Het vervolgproject ASR2 (looptijd: 01/01/2007 – 01/01/2009) is eveneens onderdeel van een groter digitaliseringprogramma (12 miljoen pond) waarbij in totaal 16 gelijkaardige projecten bij betrokken zijn.

5.3.7.1.3 Realisaties

Onderstaande figuur geeft overzichtelijk het productieverloop van beide projecten weer. Na AD-conversie (zowel in- als out-house) werd de metadata gecreëerd en aan de geluidsbestanden toegevoegd.



Figuur: Digitale workflow [26]

De ASR-website biedt ruime en gedetailleerde informatie over de geluidsopnames. De metadata (in XML) is conform met Open Archives Initiative (OAI) standaarden en werd door middel van Metadata Encoding and Transmission Standard (METS) aan de bestanden toegevoegd. Bij de projecten werd voor minimale metadata geopteerd, zowel technisch, beschrijvend als administratief.

Het clearen van de intellectuele rechten van geluidsopnames vormt vaak een complexe aangelegenheid, zeker voor het downloaden van bestanden is de muziekindustrie zeer beducht. Naast de vele onderhandelingen met auteursmaatschappijen zijn eveneens kosten verbonden aan het clearen.

The Archival Sound Recordings Service [30] bevat de resultaten van het tweejarig project om de toegang tot het British Library Sound Archive te verhogen. Na een *pilot demonstrator* in de herfst van 2005 werd de definitieve dienst in september 2006 gelanceerd, die veelvuldig werd getest op gebruiksvriendelijkheid via een testpanel. De gebruiker kan kosteloos doorheen een ruim en gediversifieerd aanbod van 12.000 geluidsmaterialen browsen en krijgt de keuze tussen twee digitale formaten voor consultatie. Voor streaming wordt geopteerd voor .wma (28kbs, 56kbs en 128kbs), voor snelle downloads worden MP3-bestanden aangeboden (128kbs) omdat bepaalde fragmenten meer dan 40 minuten beslaan. Bijgevoegde foto's zoals hoesafbeeldingen zijn in JPEG consulteerbaar.

5.3.7.1.4 Bemerkingen

ASR1 heeft aan het licht gebracht dat het grote probleem niet zozeer de digitalisering op zich is, dan wel het curatorschap en management van de gedigitaliseerde collecties naderhand. Het bepalen van de juiste metadata en op welke manier die het best bewaard kan worden zodat tijdsgebaseerde media door middel van labels efficiënt terug gevonden kan worden, was en blijft nog steeds één der grootste uitdagingen. De METS-container werd verkozen boven MPEG-21 omdat het op dat ogenblik al verder stond in het standaardisatieproces. De authenticiteit van een object kan in METS immers op een excellente manier gegarandeerd worden, wat heel belangrijk is voor langdurige bewaring volgens een archivaris. Het uiteindelijk ontwikkelde METS-schema wordt binnenkort ter disseminatie gepubliceerd. Door de verscheidenheid aan te ontsluiten collecties is gebleken dat interoperabiliteit van metadata een potentieel probleem vormt. Een "loosely coupled" metadata systeem uitgaand van Dublin Core is waarschijnlijk een goed uitgangspunt. In ASR2 wil men ook web 2.0 functionaliteiten, zoals tagging, inbouwen. Vanuit METS voorziet men hier geen fundamentele problemen, aangezien het gemakkelijk kan gerelateerd worden naar RDF. Een conversie van Dublin Core naar RDFa wordt immers voorzien.

Een ander heikel punt is het klaren van de rechten. Onderhandelingen over copyrights, opneemrechten en bijdragerechten waren/zijn tijdrovend en duur. Het is dan ook aan te raden om dit te centraliseren en over lange periodes goede betrekkingen aan te knopen met de grote spelers, zoals de muziekindustrie, aangezien er vroeg of laat toch terug zal moeten onderhandeld worden over alle rechten voor bepaalde werken. Dit alles indachtig voorziet men best een vast (berekend) budget om al deze verschillende kosten te dekken. Huidige en toekomstige technische oplossingen om bepaalde rechten te kunnen afdwingen maken ontegensprekelijk de discussies nog ingewikkelder. In METS is het tot nader order niet mogelijk om een "rights expression language" in te bedden.

Er is vooralsnog ook geen koppeling tussen het Content Management Systeem, de webomgeving en ASR 1 omgeving. Ieder object krijgt op dit ogenblik enkel een eigen directory toegewezen op de UKERNA (backbone van de UK academici) server. Daarin bevinden zich de metadata METS-file, op dit ogenblik nog steeds handmatig aangemaakt, en de verschillende digitale versies van het bronmateriaal, de ongecomprimeerde en verliesloze WAV en de afgeleide versies voor gebruik op het internet MP3 en WMA. Voor wat betreft beschrijvende metadata werd gekozen voor een “extended Dublin Core” (nieuwe rollen - machineries, animals, environment- werden toegevoegd) profiel wat de interoperabiliteit niet ten goede komt. Verder wordt nog bijna geen data geïndexeerd wat op dit ogenblik bij het opzoeken een grote negatieve invloed heeft.

5.4 National Library in Australië

5.4.1 Historiek

De National Library werd in 1960 opgericht. Het is de grootste referentiebibliotheek van Australië. Haar collectie bestaat uit boeken, dagbladen, afbeeldingen, kaarten, manuscripten, geluidsopnames, microfilm en andere media. In totaal worden er meer dan vijf miljoen items bewaard. Het overgrote deel van de collecties wordt ter plaatse bewaard, andere stukken vindt men onder meer op het Hume Annexe en in andere opslagplaatsen.

5.4.2 Infrastructuur

De National Library is gelegen in Canberra. Het telt ruim 400 werknemers en beschikt over een jaarlijks budget van meer dan 75 miljoen Australische dollar.

5.4.3 Gebruikers

Rechtstreekse toegang tot de werken in de Library is niet toegestaan voor externe gebruikers. Gebruikers kunnen wel gebruik maken van de Libraries Australia Search service. Libraries Australia is een resource-sharing service die gecoördineerd wordt door de National Library of Australia voor Australische bibliotheken en hun gebruikers. De service wordt onder andere gebruikt voor referentie, verzamelingontwikkeling, catalogeren en interbibliotheecair leenverkeer. Het hart van Libraries Australia is de Australian National Bibliographic Database (ANBD) die de locatiedetails van de 40 miljoen items opneemt die in de meeste Australische academische, onderzoeks-, nationale, staats-, publieke en speciale bibliotheken bewaard worden. De bibliotheek verleent eveneens Cataloguing-in-Publication (CiP) details, ISSN's en ISMN's voor de Australische uitgeverijen.

5.4.4 Huidige werking

Het doel van de National Library is het verzamelen, bewaren en toegankelijk maken van documentatiebronnen van nationaal belang, gerelateerd aan Australië en het Australische volk, evenals significant niet-Australisch materiaal. Dit gebeurt ook door collaboratieve overeenkomsten met andere bibliotheken en informatieproviders.

In het archief onderscheiden we vijf collectiedomeinen:

- Boeken: de bibliotheek bezit meer dan 2.7 miljoen boeken (inclusief op microfilm)
- Periodieke publicaties: dit zijn publicaties die op regelmatige of onregelmatige periodes uitgebracht worden. Hieronder vallen onder andere tijdschriften, nieuwsbrieven, jaarlijkse rapporten en kranten
- Kaarten: topografische kaarten en thematische kaarten waaronder aarde, geologie en bevolking
- Muziek: zoals uitgebrachte en onuitgebrachte partituren en opgeslagen geluidsopnames van folk muziek en traditionele dans in Australië
- Ephemere verschijnselen: verzameling van vb. sociale gewoontes, populaire cultuur en nationale evenementen

5.4.5 Digitaliseringsplannen

De bibliotheek heeft meer dan 105.000 items van zijn collectie gedigitaliseerd en levert, indien mogelijk, directe toegang tot deze items via het internet. De Library is ook een wereldleider in digitale preservatietechnieken. Om de digitalisering verder te realiseren, engageert de Library zich in een aantal activiteiten die hieronder besproken worden.

5.4.5.1.1 Digitalisering

Net zoals vele andere culturele en onderzoeksinstellingen begon de Library met het aanmaken en afleveren van digitale verzamelingen om zo onderzoek en educatie te ondersteunen. De digitalisering kende voor de Library een start in 1996 met het PANDORA Project. Hierbij werden kopieën gemaakt en bewaard van significante Australische websites. Even later begon de Library ook alle hun afbeeldingen te digitaliseren. Vervolgens voerde de Library ook digitaliseringstechnieken in voor ander materiaal zoals kaarten. Digitalisatie wordt binnenshuis uitgevoerd en er wordt gebruik gemaakt van technologieën en bestandsformaten die geschikt zijn voor dat type materiaal. In het geval van afbeeldingen worden masterafbeeldingen in hoge resolutie (300 ppi), ongecomprimeerd in het TIFF formaat bewaard, terwijl lage resolutie afgeleiden voor de weergave op websites bewaard worden in 72 ppi JPEG formaat.

5.4.5.1.2 *Digitale archivering*

Naast de eigenlijke digitalisering focust de Library zich ook op de ontwikkeling van principes en toepassingen op het gebied van selecteren, verzamelen, catalogeren, (duurzaam) bewaren, persistent identificeren en toegang verlenen tot online publicaties. De focus van de digitale archiveringsactiviteiten is gericht op het begrijpen van de complexe en evoluerende technische kwesties die het bewaren van digitale informatie met zich meebrengen. Verder probeert de National Library zowel interne als nationale strategieën te ontwikkelen om toekomstige toegang tot de informatie te verzekeren. De Library werkt hiervoor nauw samen met andere nationale bibliotheken en belangrijke culturele instellingen.

De belangrijkste initiatieven op het gebied van digitale archivering zijn:

- **PANDORA:** *Preserving and Accessing Networked Documentary Resources of Australia.*

Dit archief, dat ook gekend is als Australia's Web Archive, is een selectief archief van Australische online publicaties, zoals websites. Het archief werd ontwikkeld in samenwerking met andere bibliotheken en culturele instellingen. Via de PANDORA website krijgt men toegang tot de inhoud van het archief, maar ook tot informatie over de geschiedenis, de ontwikkeling, het beleid, de statistieken over de grootte en de groei van de databank en de ondersteunde technische infrastructuur [31].

- **Digital Collections:** Om gebruikers een eenvoudige toegang tot Australië's meest significant cultureel materiaal te verlenen, heeft de National Library de Digital Collections webpagina ontwikkeld [32]. Op deze webpagina kunnen de gebruiker gedigitaliseerde afbeeldingen, mappen, manuscripten, boeken, partituren en audio opvragen. De Library's Picture Catalogue, Online Exhibitions en Libraries Australia zoekdiensten zoals PictureAustralia, AustraliaDancing en MusicAustralia leveren alternatieve toegangsmanieren.

- **PADI:** *Preserving Access to Digital Information.*

Dit initiatief, dat opgericht werd door de National Library of Australia, richt zich op het definiëren en uitwerken van mechanismen die helpen bij het beheren van digitale informatie, met het oog op bewaring en toekomstig gebruik. Er werden vijf doelen vooropgesteld binnen dit initiatief. Het eerste doel was het vereenvoudigen van strategieën en richtlijnen voor de preservatie van toegang tot digitale informatie. Daarnaast werd er ook een website ontwikkeld voor het bewaren van informatie en promotiedoeleinden. Verder werd er ook geprobeerd relevante activiteiten te identificeren en te promoten. Er werd ook

een forum opgezet om samenwerking tussen verschillende sectoren te realiseren. Tot slot kan er ook een discussielijst voor de uitwisseling van nieuws en ideeën over digitale bewaringskwesties teruggevonden worden op de PADI website [33], die een subject gateway vormt voor digitale bewaringsresources.

5.4.5.1.3 Digitale bewaring

Naast het verzamelen van digitaal materiaal en het beheren ervan, verbindt de Library zich tot het garanderen van de toegang op lange termijn tot die digitale verzamelingen waarvoor het bewaringsverantwoordelijkheid heeft. De National Library maakt hiervoor gebruik van belangrijke beheersinformatie over zijn digitale verzameling. De Library heeft een Digital Preservation Policy uitgewerkt, die de principes van digitale bewaringsactiviteiten uiteenzet en een risk assessment uitvoert. De Library heeft ook een Digital Preservation Action Plan ontwikkeld om prioriteiten toe te kennen over hoe de digitale bewaring te leiden. Naast dit plan loopt er ook een langdurig programma (door middel van projecten) om tools, procedures en infrastructuren te ontwikkelen om de bewaring van de digitale verzamelingen te beheren.

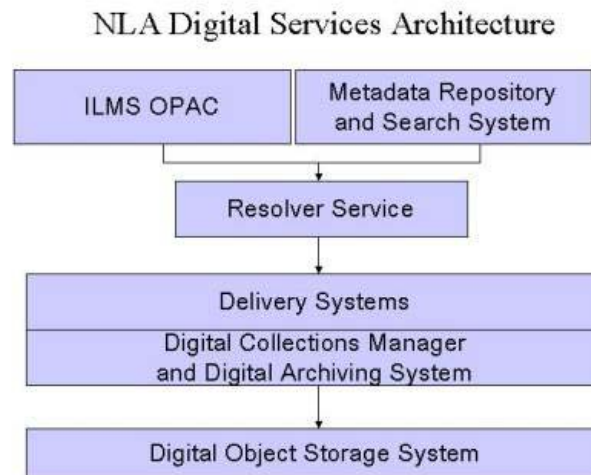
5.4.6 Digital Services Architecture

De National Library of Australia heeft een architectuur ontwikkeld voor het beheren, opzoeken en afleveren van zijn digitale collecties. De componenten van deze architectuur werden ontwikkeld in nauwe samenwerking met standaardisatie comités die actief zijn binnen het digitale bibliotheekwezen. De Library gebruikt deze architectuur om toegang te verlenen tot zijn meer dan 100.000 digitale objecten, waaronder dus gearchiveerde websites, gedigitaliseerde foto's, kaarten, muziekbladen, manuscripten en audio verzamelingen behoren.

De hoofdconcepten van de architectuur zijn:

- een Digital Object Storage Systeem
- een Digital Archiving Systeem dat het personeel van de Library tools aanbiedt om de digitalisering van de verzameling te ondersteunen en te beheren
- een Metadata Repository and Search System
- een Persistent Identifier Resolver Service zodat alle digitale objecten een persistente identifier toegekend kunnen krijgen en met behulp van deze identifier kunnen benaderd worden via het web
- Delivery Systems voor complexe gedigitaliseerde items zoals manuscripten, kaarten,...

Hieronder worden deze componenten toegelicht met de nadruk op de onderliggende technische standaarden die interoperabiliteit, zoeken, beheer en aflevering van digitale verzamelingen ondersteunen.



Figuur: Digital Services Architectuur

5.4.6.1.1 Digital Object Storage System (DOSS)

In juni 2001 verkreeg de Library een server en tapebibliotheek voor de opslag van zijn digitale collecties. De initiële capaciteit van het systeem was 2 terabyte aan schijfopslag en 4 terabyte aan tape-opslag. Deze capaciteit wordt continu uitgebreid en bereikt nu respectievelijk 4 en 12 terabyte.

Er werd een gestructureerd bestandssysteem ontwikkeld om beelden en geluidsbestanden te bewaren. Het bestandssysteem bevat gescheiden gebieden voor digitale masters, afgeleiden voor weergave op het web en een werkgebied waar bestanden tijdelijk opgeslagen worden tijdens creatie, verwerking en verplaatsing. Digitale masterbestanden worden aldus op tape bewaard terwijl webafgeleiden op schijf bewaard worden. Bestanden die op tape bewaard worden lijken voor de user online te zijn (eigenlijk nearline) en worden automatisch in enkele minuten opgehaald wanneer de gebruiker ze opvraagt.

5.4.6.1.2 Resource discovery

Gebruikers doorzoeken digitale verzamelingen doorgaans door middel van een aantal entry points. Voor de Library is het hoofd entry point de catalogus, aangezien de digitale collectie een integraal onderdeel is van de totale collectie. Bijgevolg zorgt de Library ervoor dat alle items in de digitale collecties in de catalogus vertegenwoordigd zijn. Verder zorgt het er dus ook voor dat de links uit de catalogusrecords gekoppeld zijn aan het afleveringssysteem door gebruik te maken van het Library's persistent identifier raamwerk, dat verder beschreven wordt. Een sterk punt van deze architectuur is dat ze niet afhankelijk is van één enkel Metadata Repository and Search Systeem. Tussen 1999 en 2002 was het platform van de Library's Metadata Repository and Search Systeem een Metastar Enterprise product van Blue Angel Technologies. De Library stapte daarna over op het Teratext Database System van Inquirion Pty Ltd, een spin-off van de RMIT universiteit.

5.4.6.1.3 Persistent identifiers en de resolver service

Een belangrijke eigenschap van de architectuur die gebruikt wordt door de Library is, zoals eerder vermeld, de implementatie van een persistent identifier raamwerk. Het gebruik van persistente identifiers garandeert dat elk item in de digitale verzameling van de National Library op een persistente manier toegankelijk is. De identifier van een object zal steeds terug leiden tot de locatie van het object op het web. De Library kent een persistente identifier toe aan elk digitaal item en aan elk afleverbaar deel van deze digitale items. De Library heeft ook een eenvoudige resolver service ontwikkeld die de gebruiker doorverwijst naar de huidige locatie van het object. De service baseert zich hiervoor op de persistente identifier. De resolver is een Java servletapplicatie dat eerst de persistente identifier ontleedt en daarna, gebaseerd op een aantal regels, het daarna bewaart in een XML configuratiebestand en uiteindelijk de weblocatie bepaalt waarheen de gebruiker moet doorverwezen worden. Indien de Library later de wijze van aflevering voor een bepaald object wijzigt, dan hoeft enkel dit configuratiebestand aangepast te worden.

5.4.6.1.4 Digital object management

Binnen de architectuur van de National Library werden de frontend en backend lagen geïmplementeerd met behulp van reeds op de markt beschikbare producten. Dit bleek echter niet mogelijk te zijn voor de middelste laag, de digital object management laag. Hier was de Library dus verplicht om zelf twee significante softwarecomponenten te ontwikkelen. Een van deze componenten is het Digital Archiving System, dat medewerkers van de National Library in staat stelt om items te verzamelen en te beheren. Dit systeem, gekend als PANDAS (PANDORA Digital Archiving System), wordt momenteel gebruikt door een aantal

bibliotheken die samenwerken met de National Library om het archief op te bouwen. Dit systeem dat de digitale afgeleiden beheert, wordt de Digital Collections Manager genoemd. De Library deed een grote inspanning om deze te baseren op een “digital asset management system” dat beschikbaar is op de markt. Toch was het niet mogelijk om een systeem te vinden dat zowel betaalbaar was, als dicht bij de vereisten van de Library aansloot. Het hoofdprobleem was dat systemen op de markt onvoldoende het hiërarchisch digitaal bibliotheekmodel en de digitalisatie workflow van de Library ondersteunden. Bovendien was ook de vereiste scheiding in de architectuur tussen de digitale opslag, het beheer, het zoeken en de aflevering onvoldoende. De Digital Collections Manager (DCM) ondersteunt het digitaliseringsproces, de aanmaak van afgeleide objecten voor webweergave en de opslag van technische metadata en digitale bestanden.

5.4.6.1.5 Webafleveringssystemen

Alle afleveringssystemen zijn gebaseerd op beschrijvende en structurele metadata afkomstig van de Digital Collections Manager. De generieke webafleveringssystemen geven een digitaal object terug in een contextuele display binnenin de webbrowser van de gebruiker. De gebruikersinterface laat de gebruiker toe om te navigeren doorheen de pagina's van het object, in te zoomen op details van elke pagina, en de metadata te bekijken die deel uitmaakt van het object. Het afleveringssysteem wordt aangesproken door middel van de resolver service.

5.4.6.1.6 Besluiten

De National Library heeft een technische architectuur ontwikkeld die de strategische doelen van de National Library ondersteunt. Hierbij was een ondersteuning belangrijk voor zowel de geïntegreerde toegang tot de afgeleide digitale bronnen, alsook voor de fysieke, originele afgedrukte bronnen belangrijk. Ook de garantie dat naar elk item van de digitale verzameling op een persistente manier kan verwezen worden, was van belang. Tot slot moest er een hiërarchisch digitaal bibliotheek datamodel opgemaakt worden dat de complexe aard van sommige objectverzamelingen weerspiegelt en moest er gegarandeerd worden dat de Library's Integrated Library Management System (ILMS) een centrale rol heeft in de architectuur. Dit vereiste een architectuur die de digitale opslag, het beheer, de zoekmogelijkheden en de aflevering in verschillende componenten scheidde.

6 Algemene vereistenanalyse

Deze sectie dient als leidraad bij het opstellen van een specifiek lastenboek voor het te ontwerpen ontsluitingssysteem voor mondelinge historische bronnen. Het beschrijft in algemene termen welke functionele blokken best aanwezig zouden zijn en hoe deze in een use case zouden kunnen vervat worden. Het doel is niet zozeer een volledige, technische specificatie van het te bouwen systeem te geven, dan wel het geven van een holistisch beeld aan de verantwoordelijken die zullen instaan voor het gebruik, de aankoop en/of de ontwikkeling van het te ontwerpen systeem. De opsomming hieronder is dan ook niet exhaustief.

6.1 Functionele vereisten

- Content management (submitatie, acquisitie, ingest, creatie, update, verwijdering)
 - De applicatie **moet** toelaten om nieuwe objecten en eventuele nieuwe metadata te creëren, deze in bulk in te laden en alle integriteitsregels te checken en de initiële identiteit te vrijwaren
 - De applicatie **moet** nieuwe objecten een unieke identiteit geven
 - De applicatie **zou moeten** verschillende standaard formaattypes en structuren ondersteunen
 - De applicatie **mag** overweg kunnen met complexe objecten, in zoverre dat de structurele integriteit van de onderliggende componenten steeds bewaard wordt
 - De applicatie **moet** bij creatie, update en verwijdering een transactiemechanisme voorzien waarbij een reeks acties volledig lukt (commit) of volledig ongedaan wordt gemaakt wanneer er iets mislukt (rollback) zodanig dat het onderliggende datamodel zich steeds in een consistente staat bevindt.
 - De applicatie **moet** relaties kunnen beschrijven tussen verschillende objecten onderling en tussen de verschillende versies van objecten (al dan niet in verschillende formaten)
 - De applicatie **zou moeten** op automatische wijze metadata (beschrijvende, contextuele en technische) kunnen extraheren tijdens het creatie of ingest proces
 - De applicatie **moet** toelaten om manueel metadata (beschrijvende, contextuele en technische) toe te voegen, te updaten of te verwijderen
 - De applicatie **zou** metadata technisch **moeten** kunnen valideren op juistheid (semantische regels)

- De applicatie **moet** er voor zorgen dat data en bijhorende metadata steeds persistent gelinkt zijn
- Objecten die geen operationele waarde meer hebben **zouden moeten** verwijderd worden (of verplaatst worden naar een archiefsectie)
- De applicatie **zou moeten** metadata kunnen bijhouden van objecten die verwijderd zijn (of verplaatst werden naar een archiefsectie)
- De applicatie **zou moeten** versioning van objecten ondersteunen
- Rechten management
 - De applicatie **moet** toelaten dat rechten (hier IPR-related) aangemaakt, aangepast en verwijderd kunnen worden. Deze wijzigingen moeten dan onmiddellijk repercussie hebben op de toestand van het object binnen de applicatie
- Gebruikers management
 - De applicatie **moet** exclusieve verantwoordelijkheden & rechten (bijvoorbeeld via rollen) toekennen op basis van de autoriteit van een gebruiker (bijvoorbeeld via login & paswoord) voor het creëren, aanpassen, annoteren, verplaatsen en verwijderen van objecten
 - De applicatie **zou moeten** in een aangepaste user interface voorzien voor elk toegangsniveau
 - De applicatie **moet** zoekresultaten vernauwen tot objecten waar men de toegangsrechten toe heeft
- Zoeken & indexering
 - De applicatie **moet** toelaten om (verschillende) classificatieschema's te definiëren, te creëren, aan te passen en te verwijderen
 - De applicatie **moet** er voor zorgen dat classificatieschema's uniek identificeerbaar zijn
 - De applicatie **moet** verbieden dat klassen die objecten bevatten kunnen verwijderd worden uit een classificatieschema
 - De applicatie **zou moeten** naamruimtes ondersteunen voor classificatietaxonomieën en classificatie ontologieën
 - De applicatie **mag** gecontroleerde lexicons aanbieden om zelf naamruimtes aan te maken
 - De applicatie **moet** nieuwe objecten bij creatie steeds een classificatie toewijzen
 - De applicatie **moet** objecten steeds een andere classificatie kunnen toekennen

- Classificatie schema's **zouden moeten** de intrinsieke inhoud van de business processen reflecteren
- De applicatie **moet** er voor zorgen dat alle data en metadata steeds kan gevonden worden
- De applicatie **zou moeten** in een geïntegreerde zoekfunctie voorzien
- De applicatie **zou moeten** in een interface voorzien zodanig dat ook externe zoekmachines kunnen ingeplugd worden
- De applicatie **zou moeten** full-text search ondersteunen voor zowel data als metadata
- De applicatie **zou moeten** metadata search ondersteunen door middel van een gecontroleerd vocabularium uit een thesaurus
- De applicatie **zou moeten** Booleaanse & relationele zoekoperatoren ondersteunen
- De applicatie **zou moeten** wildcards & patronen bij het zoeken ondersteunen
- De applicatie **zou moeten** iteratief verfijnd zoeken ondersteunen (faceted search)
- Archivering (lange termijn preservatie)
 - De applicatie **zou moeten** administrators waarschuwen dat objecten (bijvoorbeeld omwille van IPR-regelingen) naar de archiefsectie zullen verplaatst worden
 - De applicatie **moet** archiefrijpe objecten integriteitsvol kunnen verplaatsen naar de archiefsectie
 - De applicatie **zou moeten** objecten in bulk kunnen verplaatsen naar de archiefsectie
- Rapportering & Backup
 - Procedures **zouden moeten** aanwezig zijn om (automatisch) backups te nemen van data, metadata en configuratiebestanden
 - Backups **zouden moeten** kunnen bewaard worden op geografisch verschillende plaatsen
 - De applicatie **zou moeten** een inconsistente versie van het systeem door middel van een backup kunnen vervangen door een consistente versie
 - De applicatie **zou moeten** veiligheidslekken kunnen opsporen, melden en dichten
 - De applicatie **zou moeten** procedures voorzien om op gezette tijdstippen veiligheid audits en login logs te inspecteren

- Interfaces
 - De applicatie ***zou moeten*** op een gestandaardiseerde manier kunnen communiceren met externe systemen (bijvoorbeeld via webservices)
 - De applicatie ***zou moeten*** een API voorzien zodat andere business applicaties geïntegreerd kunnen worden

Bijgevoegd vind je hoe belangrijk de respondenten het toevoegen van specifieke kenmerken in het te ontwikkelen systeem vinden:

Hoe belangrijk zou u volgende features vinden binnen het te ontwerpen systeem:	Helemaal niet belangrijk	Niet belangrijk	Eerder niet belangrijk	Eerder belangrijk	Belangrijk	Zeer belangrijk
Eenvoudig zoeken (free text zoekveld, cfr. Google search)					X	
Doorzoeken van de collectie aan de hand van bepaalde metavelden (vb. titel, naam geïnterviewde, ...) (zoekwoord vrij in te geven)						X
Voor een bepaald veld zoeken adhv een thesaurus (vb. bij thema kan je kiezen uit de vaste lijst: WOII, verkiezingen, de mijnbouw, sport, ...)					X	
Online bekijken of beluisteren van fragmenten (streaming)					X	
Online raadplegen van transcripties				X		
Mogelijkheid tot gratis downloaden van de mondelinge bronnen			X			
Mogelijkheid tot downloaden van de mondelinge bronnen tegen betaling			X			
Mogelijkheid tot het aanvragen van kopieën op CD of DVD				X		
Verfijnen van een zoekopdracht (zonder een volledig nieuwe zoekopdracht te lanceren)					X	
Suggesties op basis van gebruikers met gelijke interesse			X			
Suggesties op basis van categorieën (andere bronnen die in gelijkaardige categorieën vallen als het resultaat van uw oorspronkelijke zoekopdracht)				X		

Hoe belangrijk zou u volgende features vinden binnen het te ontwerpen systeem:	Helemaal niet belangrijk	Niet belangrijk	Eerder niet belangrijk	Eerder belangrijk	Belangrijk	Zeer belangrijk
Suggesties op basis van eigen ingevoerd profiel			X			
Mogelijkheid tot opzoeking per instelling				X		
Bewaren van de eigen selecties				X		
Mogelijkheid tot aanmaken van een eigen gebruikersprofiel			X			
Materiaal opzoeken aan de hand van gelijkaardige beelden (picture similarity)			X			
Een zoekopdracht opslaan voor later gebruik				X		
De mogelijkheid om te kiezen welke velden (metadata) en visuele informatie (keyframes, storyboard, ...) worden weergegeven in de resultatenlijst				X		
Online toevoegen van commentaar door gebruikers				X		
Online toevoegen van een score door de gebruikers		X				
Online toevoegen van trefwoorden door gebruikers			X			
Online omgeving voor discussie tussen eindgebruikers over bepaalde thema's			X			
Omgeving waar de eindgebruiker zelf content kan bewerken		X				
Online doorzoeken van de audiobestanden (op klanken)			X			
Bestanden raadplegen of downloaden in lage resolutie				X		
Bestanden raadplegen of downloaden in hoge resolutie				X		

6.2 Niet-functionele vereisten

- Digitalisatie
 - De applicatie **zou moeten** (transcoding) services aanbieden waardoor bestaande, externe objecten (in een aantal voorgedefinieerde standaard formaten) kunnen ge-ingest worden
- Interoperabiliteit (openheid systemen & support voor standaarden)
 - De applicatie **zou moeten** draaien op de meeste operating systems en andere core infrastructuur software
 - De applicatie **zou moeten** alle objecten kunnen tonen in een open export formaat
- Huidig gebruik in het veld van standaarden
 - De applicatie **zou moeten** gebruik maken van standaarden binnen het applicatiedomein reeds een grote gebruiksbasis kennen.
- Schaalbaarheid & onderhoud
 - De applicatie **zou moeten** ontworpen zijn rond een flexibele architectuur die kan aangepast worden naargelang de noden veranderen
 - De applicatie **mag** gedistribueerde depots met multi-site services ondersteunen
 - De applicatie **zou moeten** ontworpen zijn dat er gemakkelijk aan load balancing kan worden gedaan
 - De applicatie **zou moeten** backwards compatible zijn (toch op zijn minst één versie)
 - De applicatie **zou moeten** ontwikkeld worden op 3 verschillende niveaus: ontwikkelomgeving, testomgeving en productieomgeving
 - Configuratieparameters **moeten** gemakkelijk aanpasbaar zijn
 - Routine onderhoudstaken **moeten** geautomatiseerd zijn
 - Transitie van ontwikkelomgeving naar testomgeving of van testomgeving naar productieomgeving **moet** geautomatiseerd en gemonitord worden
 - Downtime **moet** geminimaliseerd worden
 - Het aantal simultane gebruikers **moet** gemaximaliseerd worden
- Levensduur van de data
 - De applicatie **zou moeten** data en metadata in standaardformaten opslaan die gemakkelijke conversie (bijvoorbeeld transcoding) toelaat naar standaarden die in de toekomst aan de oppervlakte zullen komen

- Performantie
 - De applicatie **moet** garanderen dat primaire business functies steeds aanspreekbaar zijn binnen een aanvaardbare tijdspanne
 - Stress testen **zouden moeten** ontworpen worden om het op te leveren systeem op voorhand te testen
- Kosteffectiviteit (“In huis” ontwikkeling versus uitbesteding)
 - Er **zou moeten** een kosten/baten analyse gebeuren omtrent wie best de eigenlijke implementatie van het systeem zal doen. Veel zal afhangen van de beschikbaarheid en de bekwaamheid van de eigen IT-staff
- Inpasbaarheid legacy systemen
 - De applicatie **moet** API’s voorzien zodanig dat legacy systemen kunnen ingeplugd worden (bijvoorbeeld via webservices)
- Opgelegde beperkingen van het consortium
 - De applicatie **moet** rekening houden met de eventuele beperkingen die het consortium oplegt (bijvoorbeeld: ontwikkeltijd, te verkiezen OS of Db-vendor, etc)
- Stabiliteit & reputatie van HW & SW vendor
 - Het consortium **zou moeten** rekening houden bij de keuze van het HW en SW platform met de stabiliteit van het product en de reputatie van de vendor
- Authenticatie/authorisatie & security
 - De applicatie **moet** het aanmaken en aanpassen van veiligheidsprofielen voorzien
 - De applicatie **moet** toelaten om gebruikers een ander veiligheidsprofiel te geven
 - De applicatie **moet** services voorzien om beveiligde authenticaties te gebruiken
 - De applicatie **moet** gebruikers authenticeren vooraleer services aan te bieden
- Usability (administrator & eindgebruiker)
 - De user inface **moet** gemakkelijk in gebruik zijn
 - De user interface **zou moeten** een consistente look&feel hebben
 - De user interface **zou moeten** betekenisvolle foutboodschappen geven als er een fout optreedt en eventueel een oplossing proberen voor te stellen
 - Alle zoekresultaten **moeten** kunnen getoond worden (eventueel via een filter)
 - Gebruikers **zouden moeten** zoek query’s kunnen bewaren en hergebruiken
 - De applicatie **moet** alle componenten (data en metadata) van een object op een logische manier tonen en dit zowel op een scherm als op een afdruk

- De applicatie **moet** de content, structuur, context en functionaliteit van een object kunnen tonen

6.3 Algemeen plan van aanpak

- Voer behoeftanalyse uit bij de belanghebbende partijen
- Identificeer de use cases en definieer de te ontwikkelen services
- Vervolledig de vereistenanalyse (iteratief)
- Stel een uitvoerings- en beheersplan op
- Identificeer de benodigde resources en bijhorende kosten
- Kies een ontwikkelraamwerk en installeer dat platform
- Ontwikkel de nodige software componenten en services (iteratief)
- Stel een onderhoudsplan op
- Stel een uitrolplan op
- Kies een marketing strategie

7 Haalbaarheidsstudie metadata

Digitale bronnen zullen slechts maximaal benut kunnen worden wanneer ze door verschillende gebruikers in verschillende omstandigheden gemakkelijk gebruikt kunnen worden. Dit zal enkel mogelijk zijn wanneer de collectiebeheerders zich rigoureuus houden aan een afgesproken set standaarden wanneer ze hun digitale bronnen aanmaken en beheren. Standaardisatie is hierbij ook de sleutel om volledige interoperabiliteit te waarborgen tussen de verschillende bronnen en datasets. Het is eveneens belangrijk voor het borgen van de onderhoudbaarheid en de preservatie van het digitaal materiaal. Snelle technologische evoluties hebben ervoor gezorgd dat er al een grote waaier standaarden aanwezig is, die op zich ook nog steeds kunnen veranderen en verder uitgewerkt worden. Het doel van deze haalbaarheidsstudie is de verschillende collectiebeheerders een houvast te geven in het doolhof der standaarden en een model naar voor te brengen waardoor het steeds mogelijk blijft om alle digitale bronnen te blijven raadplegen en converteren.

Vooreerst moet er beslist worden welke technische standaarden de culturele instellingen best volgen wanneer ze hun mondelinge historische bronnen digitaliseren. Deze standaarden bepalen de codering van karakters en tekst, alsook de codering van stilstaande en bewegende beelden, digitale audio, video en virtuele realiteit. Er zijn ook standaarden die ervoor zorgen dat digitale bronnen kunnen uitgewisseld worden tussen applicaties. Deze standaarden bepalen de graad van interoperabiliteit en zullen in de toekomst nog aan belang winnen aangezien het volume aan beschikbaar digitaal materiaal alleen nog zal toenemen.

Beschrijvende “culturele” standaarden bepalen de gedetailleerde beschrijvingen van de bronnen. Sommige instellingen wensen volledige collecties bronnen te kunnen beschrijven, andere culturele huizen willen eerder individuele objecten beschrijven. Het is duidelijk dat er een verschillende aanpak nodig is voor het beschrijven van verschillende types bronnen. Mondelinge historische bronnen, bijvoorbeeld, zullen anders beschreven worden dan archieven of boeken.

In laatste instantie worden er best ook standaarden gevolgd voor het beheer van digitale bronnen, meer specifiek wanneer de bronnen bewaard en gearchiveerd moeten worden zodanig dat ze steeds zullen kunnen gebruikt blijven worden ondanks technologische veranderingen.

7.1 Containerformaten

7.1.1 Audiocontainers

7.1.1.1 *WAV*

WAV, of Waveform Audio Format, is een audio bestandsformaatstandaard voor het opslaan van audio op Pc's. WAV is een bestandsformaat dat standaard beschikbaar is op Windows-systemen. WAV slaat de audiodata ruw op. Door het verliesloze karakter van ruwe audio kunnen WAV-bestanden echter heel groot worden.

Een WAV-bestand wordt opgebouwd uit zogenaamde chunks [34]. Deze chunks geven informatie over het geluid of bevatten het geluid zelf. Naast deze chunks bevat een WAV-bestand ook een header met onder ander informatie over de gebruikte formaatstructuur voor het bestand.

De maximale grootte van een WAV-bestand bedraagt 4GiB wat overeenkomt met ongeveer 405 minuten geluid in Cd-kwaliteit (44.1kHz, 16 bit, stereo) en 62 minuten in DVD-Audio-kwaliteit (tot 192kHz, tot 24 bit, stereo). Om deze beperkingen weg te werken werd later het W64-formaat ontworpen dat de grootte van het bestand in 64 bits in plaats van in 32 bits beschrijft [35]. De EBU heeft om dezelfde reden het RF64-formaat ontwikkeld [36]. Dit formaat voegt verder ook nog onder andere de ondersteuning toe voor maximaal 18 surround kanalen. Naast ruwe audio ondersteunt de WAV-container ook andere codecs zoals GSM, ADPCM en MPEG Layer-3.

7.1.1.2 *AIFF*

AIFF, of Audio Interchange File Format, is de Apple Macintosh tegenhanger van WAV. Dit formaat komt grotendeels overeen met het WAV-formaat van Microsoft. Het grote verschil is dat waar bij WAV de samples in een little-endian byte volgorde worden opgeslagen, dit bij AIFF in big-endian byte volgorde word gedaan. Sinds Mac OS X heeft Apple echter een nieuw type AIFF gecreëerd dat in little-endian byte volgorde wordt opgeslagen. Dit door de overgang naar Intel-processoren die little-endian byte volgorde gebruiken.

AIFF is ook opgebouwd uit een header en zogenaamde chunks die informatie over en het geluid zelf kunnen bevatten [37].

7.1.1.1.3 XMF

XMF, of eXtensible Music Format, is een familie van muziekgerelateerde formaten ontworpen door de MIDI Manufacturer's Association [38] [39] [40]. XMF heeft tot doel één of meerdere bestanden, in bestaande formaten zoals MIDI en WAV, samen te voegen.

XMF bestaat uit twee delen: het XMF Meta-File Format en een reeks XMF File Types, die gebruik maken van het XMF Meta-File Format. Tot dusver zijn XMF Type 0, XMF Type 1 en Mobile XMF gedefinieerd. Deze zijn echter allemaal op MIDI gericht.

Een XMF Meta-bestand is samengesteld uit verschillende nodes die hiërarchisch gegroepeerd zijn zoals een bestandssysteem met folders en bestanden. Bij XMF worden hiervoor respectievelijk de term containers en resources gebruikt. Een node kan ofwel een container ofwel een resource zijn. Een resource node bevat dan ofwel een verwijzing naar een intern bestand ofwel een URL die verwijst naar een extern bestand.

7.1.2 Multimediacontainers

7.1.2.1.1 OGM/OGG

OGM, of OGG Media, is een containerformaat dat een uitbreiding vormt op het OGG-containerformaat van Xiph.org [41]. OGM voegt aan OGG onder andere de ondersteuning voor andere codecs toe dan die ontworpen door Xiph.org (Speex, Theora en Ogg Vorbis) [42] [43] [44]. OGM biedt namelijk ook ondersteuning voor videocodecs die gebruik maken van Vfw en audiocodecs die ACM gebruiken. Bij de audiobitstromen wordt, net als bij videobitstromen, ondersteuning geboden voor een variabele bitrate. Algemeen wordt OGM gezien als een tussenstap tot andere containerformaten zoals Matroska volgroeid zijn en dezelfde mogelijkheden bieden. Tot deze mogelijkheden behoren onder andere ondersteuning voor hoofdstukken, meerdere ondertitels en meerdere audiokanalen.

7.1.2.1.2 Matroska (MKV/MKA)

Matroska is een openstandaard multimedia containerformaat wat gebaseerd is op EBML (Extensible Binary Meta Language) [45]. Dit is een binair bytegealigneerd formaat gebaseerd op de principes van XML [46].

Een Matroska bestand bestaat uit een header met informatie over de gebruikte EBML versie en het bestandstype, in dit geval dus een Matroska bestand. De header wordt gevolgd door de Metaseek-sectie die de plaats aanduidt van de verschillende andere secties binnen het bestand. Dit is nodig omdat elke sectie in principe overal in het bestand kan voorkomen en

men dus het hele bestand zou moeten parsen bij het zoeken naar informatie. Er zijn secties voor o.a. kanaalinformatie, hoofdstukinformatie en tags [47].

Matroska kent twee onderverdelingen: MKV, dat zowel video als audio kan bevatten, en MKA, dat enkel bedoeld is voor audio. Er is ondersteuning voor bijna alle video- en audioformaten zoals MPEG-1, MPEG-2, MPEG-4, Quicktime, Real, Theora voor video en MP1, MP2, MP3, PCM, AC3, FLAC, AAC voor audio. Hierbij worden zowel variabele audio bitrate als variabele framerate ondersteund. Verder biedt Matroska ook de mogelijkheid bestanden van om het even welk type toe te voegen. Dit maakt het mogelijk om bijvoorbeeld transcripties toe te voegen aan het bestand.

Matroska kan een ongelimiteerd aantal videostreamen, audiostreamen, afbeeldingen en ondertitels bevatten en laat ook toe lettertypes toe te voegen voor bijvoorbeeld de ondertitels. Matroska biedt verder ook een robuuste ondersteuning voor streaming, hoofdstukken en DVD-achtige menu's.

7.1.2.1.3 MXF

MXF, of Material eXchange Format, is standaard voor professionele video en audio en wordt gevormd door een set SMPTE standaarden [48]. Een MXF-containerbestand bestaat uit een header, de audio- en videostreamen en een EDL of Edit Decision List. Een EDL bevat de gegevens zoals gebruikt bij audiovisuele content editing systemen en dient als een soort tijdlijn.

MXF ondersteunt alle mogelijke video- en audioformaten en laat ook toe dat willekeurige bestanden worden toegevoegd. Dit laat toe transcripties, beelden, enz. toe te voegen.

MXF ondersteunt ook het toevoegen van metadata en enkele professionele functies zoals een volledige timecode en platformafhankelijkheid. MXF laat ook streaming toe.

7.1.2.1.4 MP4

MP4 of MPEG-4 part 14 is een multimedia containerformaat dat een onderdeel vormt van de MPEG-4 standaard [49]. MP4 kan zowel audio- als videostreamen bevatten. Hierbij ondersteunt MP4 als videoformaten de standaarden MPEG-1, MPEG-2, MPEG-4 en MPEG-4 AVC. Voor audio worden de standaarden (HE)-AAC, MP3, MP2, MP1, CELP, TwinVQ, Vorbis en Apple Lossless ondersteund. Wanneer een MP4-container enkel audio bevat wordt deze vaak van de extensie M4A voorzien. Deze extensie wordt onder andere gebruikt voor podcasts.

Naast de gewone audio- en videostromen kent MP4 ook zogenaamde private stromen. Deze private stromen kunnen om het even welke gegevens bevatten. Zo gebruikt Nero deze stromen om ondertitels in Dvd-formaat toe te voegen.

MP4 ondersteunt verder ook afbeeldingen, hyperlinks, ondertitels, hoofdstukken, variabele audio bitrate en variabele framerate.

7.1.2.1.5 3GP

3GP, of 3G Protocol, is een multimedia containerformaat ontworpen door de Third Generation Partnership Project (3GPP) voor gebruik met 3G mobiele telefoons [50]. 3GP is een vereenvoudigde versie van het MP4 containerformaat en is ontworpen met het doel de opslag- en bandbreedtevereisten te verminderen. 3GP ondersteunt zowel MPEG-4 Part 2, H.264/AVC als H.263 voor video en AMR-NB, AMR-WB, AMR-WB+ en (HE)-AAC-LC voor audio. 3GP biedt ook ondersteuning voor variabele audio bitrates, variabele framerate's en ondertitels. 3GP bestanden kunnen zowel gestreamd als gedownload worden (denk maar aan MMS-berichten).

7.1.2.1.6 ASF

ASF, of Advanced Systems Format, is een propriëtair containerformaat ontworpen door Microsoft als onderdeel van het Windows Media Framework [51]. De vroegere naam Advanced Streaming Format geeft het hoofddoel weer van het containerformaat, namelijk streaming. ASF kent twee versies. Versie 1.0 is veruit de meest gebruikte versie maar is gesloten, de opbouw is dus op enkele details na ongekend. Versie 2.0 is open maar wordt weinig tot niet gebruikt.

ASF ondersteunt bijna alle video- en audioformaten die werken via VFW en ACM maar wordt meestal gebruikt in combinatie met Microsofts eigen formaten. Verder ondersteunt ASF ook metadata zoals artiest en titel, variabele audio bitrate, variabele framerate, hoofdstukken en ondertitels. ASF biedt ook foutcorrigerende technieken en een digital rights management framework.

7.1.2.1.7 MOV

MOV is een multimedia containerformaat ontworpen door Apple en diende als basis voor het MP4 containerformaat [52]. Deze container kan zowel video, audio, hoofdstukken als bevatten en ondersteunt variabele audio bitrate en variabele framerate. MOV ondersteunt alle formaten die de Quicktime codec manager ondersteunt, zoals MPEG-4 en de Sorensen codec. Voor audio wordt alles ondersteund wat de sound manager en coreaudio ondersteunen zoals AIFF, WAV en MP3.

In een MOV-container kan elk kanaal voorgesteld worden door ofwel de mediastroom zelf ofwel een verwijzing naar de mediastroom in een ander bestand. Binnen de MOV-container worden de kanalen in een hiërarchische structuur van atomen geplaatst. Deze atomen kunnen ofwel “ouder” zijn van andere atomen ofwel zelf media of data bevatten.

MOV-containers bevatten een tijdslijn die los staat van de mediastromen. Hierdoor kunnen MOV-containers eenvoudig worden aangepast zonder dat de mediastromen moeten worden gekopieerd.

7.1.2.1.8 AVI

AVI, of Audio-Video Interleaved, is een multimedia containerformaat ontworpen door Microsoft [6.1.2.8-1]. AVI-containers kunnen meerdere audio- en videokanalen bevatten. Een AVI-container bestaat uit een header met informatie over de video, zoals breedte, hoogte en framerate, de eigenlijke data zelf. Verder kan een container ook een index bevatten die toelaat te navigeren binnen de container. AVI-containers ondersteunen bijna alle audio- en videoformaten die beschikbaar zijn via DMO, ACM en Vfw. AVI ondersteunt variabele audio bitrates zij het met beperkingen (niet via ACM) en variabele framerates. Ondertitels en hoofdstukken worden ook ondersteund via modificaties (zij het buiten Microsoft om).

7.1.2.1.9 FLV

FLV, of Flash Video, is een propriëitair containerformaat ontworpen door Adobe en wordt gebruikt door bijvoorbeeld Google Video en YouTube [54]. FLV kan slechts één video en één audiostream bevatten per bestand [55]. Verder kan een container ook Flashcontent bevatten. FLV ondersteunt de videoformaten Sorensen, VP6 en Screen Video en de videoformaten MP3, Nellymoser, ADPCM en PCM. Een FLV-container kan op verschillende manieren bij de eindgebruiker terecht komen. Via download, embedded in een flash animatie of door streaming via het RTMP-protocol.

In de nieuwe versie van FLV wordt ook ondersteuning voor H.264/AVC en HE-AAC aangeboden.

7.1.2.1.10 RealMedia

RealMedia is een multimedia containerformaat ontworpen door RealNetworks. Realmedia is een populair formaat voor het streamen van audio en video via het internet. RealMedia ondersteunt als videoformaten RealVideo 8-9-10 en als audioformaten HE-AAC, Cook, Vorbis en RealAudio Lossless. Verder biedt Realmedia ook ondersteuning voor variabele framerate, ondertitels en met behulp van de RMVB-extensie ook variabele bitrates.

7.2 *Metadatastandaarden*

Metadata kent verscheidene definities maar algemeen wordt metadata beschreven als “bits about bits” of vrij vertaald “data over data”. Metadata biedt dus (gestructureerde) informatie over een resource. Onder resources verstaan we alle mogelijke objecten/subjecten/... waarover informatie kan worden opgeslagen zoals bvb. tekst, fysieke objecten, software, maar ook bvb. personen, gebeurtenissen en diensten.

Metadata kan worden opgedeeld in verschillende types volgens de informatie die erin beschreven staat. Deze types gaan van administratieve metadata (zoals rechten plaats) over beschrijvende metadata en bewaringsinformatie (zoals toestand en verhuizingen) tot pure technische informatie (zoals formaat en encryptie) en gebruik.

Een meer geavanceerde vorm van metadata bevat ook onderlinge relaties zoals bvb. creatie en publicatie. Hierbij worden de persoon “Interviewer” en de resource “mondelijke historische bron” verbonden door de relatie “creatie”.

Metadatastandaarden kunnen worden ingedeeld op zoekmogelijkheden. Naast gewone metadatastandaarden (zoals MARC/MARC21) bestaan er immers ook semantische standaarden die “intelligentere” zoekmethoden ondersteunen waarbij rekening wordt gehouden met de betekenis van zoektermen of die gebruik maken van thesauri. Een hierbij aansluitend probleem bij metadata bestaat er in dat “computers” gegevens als “H. Claus” en “Hugo Claus” zullen zien als 2 verschillende personen, dit in tegenstelling tot mensen. Om zulke problemen te voorkomen worden vaak “woordenboeken” gebruikt waarin termen worden afgesproken die moeten worden gebruikt. Ook thesauri kunnen hierbij helpen door zoekopdrachten uit te breiden met termen met gelijkaardige betekenis.

7.2.1 MARC/MARC21

MARC is een acroniem voor Machine-Readable Cataloging. MARC is een standaard voor het representeren en de communicatie van bibliografische en aanverwante informatie en dit in een machine-leesbare vorm, aangevuld met aanverwante documentatie. De standaard wordt onderhouden door de Amerikaanse Library of Congress en vindt zijn oorsprong in de jaren 1960 als een digitale vorm van bibliotheekfiches. De hoofdfunctie van de standaard was dan ook het vereenvoudigen en versnellen van het terugvinden van boeken in de bibliotheek. De MARC data-elementen vormen dan ook de basis van de meeste bibliotheekcatalogi. Verder is er geen enkel alternatief met een gelijkaardige graad van gegranuleerdheid [56].

MARC ondersteunt acht soorten materiaal waaronder het type “Sound recordings” wat alle geluid omvat, uitgezonderd muziek. Hieronder vallen dus ook mondelinge historische bronnen. Verder bevat MARC ook zeven types records waaronder “Computer file”, zoals een gedigitaliseerde versie van een mondelinge historische bron, en “Manuscript (textual) language material”, zoals de transcriptie van een mondelinge historische bron.

Een MARC (bibliografisch) record bestaat uit meerdere velden [57]. Er zijn velden voor auteur, titel informatie, enz. Deze velden kunnen verder worden opgedeeld in subvelden.

De tekstuele namen van de velden (zoals auteur en onderwerp) worden vervangen door tags bestaande uit een 3-cijferige code. Deze code beschrijft dus welke gegevens in het veld staan.

Subvelden worden gescheiden door middel van een karakter (bvb \$), aangevuld met een subveldcode die aangeeft welke gegevens volgen.

Sommige velden worden verder gedefinieerd door indicatoren. Dit zijn 2 posities die een karakter tussen 0 en 9 kunnen bevatten. Het 2^e karakter kan bijvoorbeeld aangeven dat een aantal volgende karakters door de computer moeten worden genegeerd bij het sorteren. Dit kan bijvoorbeeld gebruikt worden bij auteurs met een familienaam die met “van” begint.

Eenvoudig voorbeeld van een MARC Entry:

245 10 \$aInterview met een oudstrijder**\$h**[sound recording].
260 ## \$aKortrijk**\$b**Vereniging voor oudstrijders**\$c**1999.
300 ## \$a1 minidisc**\$b**digital, ATRAC, stereo.
500 ## \$aInterview met een oudstrijder 1940-1944.
500 ## \$atranscriptie beschikbaar.
511 0# \$aInterview afgenomen door X

De eerste regel bevat een veld met de code 245 wat wijst op een "Title Statement". De indicatoren hebben de waarde 1 en 0 en het veld bevat de subvelden \$a, de eigenlijke titel, en \$h het medium.

Om de records overzichtelijker te maken en bewerking van de records te vereenvoudigen is later de MARC XML-standaard ontworpen die MARC-records in XML-bestand voorstelt [58].

Het voorbeeld toont de hoge mate van gegranuleerdheid die het MARC-formaat biedt en de daarmee gepaard gaande complexiteit. Het formaat is desondanks compact. Veldnamen zoals "plaats van publicatie" worden immers vervangen door een korte code. Bovendien zit er een logica in de veldcodes wat de complexiteit iets vermindert (bvb 6XX betekent veld met informatie over het onderwerp, X00 betekent een naam).

MARC kent geen semantische zoekfunctie, wat wil zeggen dat gezocht wordt naar de gegeven sleutelwoorden in de verschillende velden maar geen rekening wordt gehouden met de betekenis of het concept van de sleutelwoorden.

Naast de bibliografische records die de karakteristieken van resources bespreken zijn er nog andere types records die bvb. een classificatie beschrijven of informatie geven over namen, onderwerpen, enz.

Nadelen MARC/MARC21:

- Complex
- Geen hiërarchische opbouw
- Geen semantiek
- Niet geschikt voor "leken"

Voordelen MARC/MARC21:

- Een hele hoge gegranuleerdheid
- Wijdverbreid
- Kunnen alsnog in XML worden weergegeven

7.2.2 Dublin Core

Dublin Core is een sectoroverschrijdende metadatastandaard [59] [60] [61] [62] [63]. Deze is niet ontworpen met het doel een verfijning en complexiteit van bijvoorbeeld het MARC-formaat te evenaren. Bij Dublin Core wordt namelijk getracht een grootste gemene deler te vormen tussen verschillende sectoren. Dit met het oog op de vereenvoudiging van onderlinge informatie-uitwisseling en zoekopdrachten. Bij het invullen van metadata moet hierdoor rekening gehouden worden dat sommige elementen in verschillende sectoren een verschillende semantiek kennen.

Binnen Dublin Core wordt gesproken over *Resources*, *Elementen*, *Qualifiers* en *Schemes*. *Resources* zijn de te beschrijven objecten en worden beschreven met behulp van 15 verschillende *elementen* gaande van creator tot rechten. Onder deze elementen bevindt zich ook het element type dat toelaat de aard van het object weer te geven. Dit element kent onder andere de categorie Sound waaronder mondelinge historische bronnen vallen. Deze 15 elementen vormen het zogenaamde Dublin Core Simple profiel.

Het Dublin Core Qualified voegt aan Dublin Core Simple 3 extra elementen toe (zoals bvb. doelpubliek) en vult het profiel ook verder aan met het toevoegen van *Qualifiers* en *Schemes*. *Qualifiers* worden gebruikt om elementen te verfijnen (kan bvb. bij het element creator vastleggen of dat een fotograaf is, een auteur, enz.). Deze *Qualifiers* zijn niet aan regels gebonden. Dit heeft dan ook tot gevolg dat niet elke software eventueel toegevoegde qualifiers zal begrijpen. Software die niet weet wat de term “fotograaf” uit het voorbeeld is zal deze dan interpreteren als een creator. Qualifiers bieden dus extra informatie aan aan die software die er mee overweg kan zonder aan compatibiliteit in te boeten.

De andere toevoeging Scheme laat toe aan te geven op welke manier elementen moeten worden ingevuld. Zo kan bvb. worden aangegeven dat het subject een trefwoord is uit een bepaalde thesaurus en niet een willekeurig gekozen woord. In het attribuut Scheme wordt dan vastgelegd welke thesaurus er wordt gebruikt.

Naast de standaardset van elementen kunnen ook andere elementen worden toegevoegd. Hierbij wordt aangeraden elementen te kiezen die in andere metadatastandaarden wel worden gebruikt. Zo een nieuwe set elementen vormt een “application profile”.

Voorbeeld van een Dublin Core Entry:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://media.example.com/audio/guide.ra">
    <dc:creator>X</dc:creator>
    <dc:title>Interview met een oudstrijder</dc:title>
    <dc:description>Interview met een oudstrijder 1940-
      1944</dc:description>
    <dc:date>1999</dc:date>
  </rdf:Description>
</rdf:RDF>
```

Nadelen Dublin Core:

- Dublin Core beperkt zich tot het beschrijven van resources zoals boeken en geluidsfragmenten en ondersteunt niet het beschrijven van personen en instellingen.
- Dublin Core beschrijft voor een belangrijk deel het voorwerp zelf en slechts in beperkte mate het uitgebeelde/beschreven onderwerp.
- Verschillende interpretaties van eenzelfde element kan leiden tot “vertaalproblemen”, al geldt dit voor alle metadatastandaarden.

Voordelen Dublin Core:

- Dublin Core maakt het aaneensluiten van metadata met andere instellingen die gebruik maken van Dublin Core heel eenvoudig
- Vele nadelen van de beperkte hoeveelheid elementen kunnen worden weggewerkt door het gebruik van qualifiers.
- De standaard ondersteunt RDF-gebaseerde opslag

7.2.3 EAD

EAD is een acroniem voor Encoded Archival Description en is een metadatastandaard die is ontwikkeld door de bibliotheek van de University of Berkeley in Californië [64] [65]. Deze zou meer informatie kunnen invoeren dan die voorzien bij MARC records.

Tot de vereisten behoorden:

- Mogelijkheid tot het weergeven van uitgebreide en intergerelateerde beschrijvende informatie
- Mogelijkheid tot het bewaren van de hiërarchische relaties tussen verschillende niveaus van beschrijving
- Mogelijkheid tot het weergeven van beschrijvende informatie die geërfd wordt over de hiërarchische niveaus heen
- Mogelijkheid te navigeren binnen een hiërarchische informatiestructuur
- Ondersteuning voor elementspecifieke indexering en navigatie.

EAD is SGML-gebaseerd maar ondersteunt ook XML. De elementen die mogen gebruikt worden om een manuscript collectie te beschrijven en de ordening van deze elementen (bvb. welke elementen nodig zijn, welke elementen toegelaten zijn binnen andere elementen) worden gespecificeerd in de EAD Document Type Definition (DTD). De tag set hierin gespecificeerd bestaat uit 146 elementen en wordt gebruikt zowel voor het beschrijven van een collectie in zijn geheel als voor de encoding van een gedetailleerde multi-level inventaris van de collectie. Vele EAD elementen zijn, of kunnen worden, afgebeeld op andere standaarden zoals MARC en Dublin Core, wat de flexibiliteit en interoperabiliteit van de data verhoogd.

Een EAD bestaat uit verschillende onderdelen:

- De EAD-header bevat de titel en gedetailleerde informatie over de collectie en het document. De elementen in de header worden vaak ook gemapt naar Dublin Core-elementen
- De archiefbeschrijving bestaat uit de Data Item Description (DID), aangevuld met eventuele extra beschrijvingen en vervolgens, het grootste deel, de volledige inventaris van de collectie

De did bevat een beschrijving van de collectie in zijn geheel, inclusief de beheerder (persoon of organisatie), taal, korte beschrijving,... Deze DID kan gevolgd worden door verschillende extra elementen:

- Een biografische beschrijving van de persoon of organisatie
- Een uitgebreide beschrijving van de collectie
- Beschrijving van objecten gerelateerd aan de collectie
- Objecten die tot de collectie behoren maar die gescheiden zijn van de collectie (bv voor speciale behandeling, specifieke opslagbehoeften,...)
- Een lijst van onderwerpen of trefwoorden voor de collectie
- Beperkingen op het materiaal in de collectie.

De inventaris van de collectie wordt progressief opgedeeld in kleinere stukken met steeds “fijnere” informatie. Dit laat toe om bij zoekopdrachten en inventariseren de gewenste informatiediepte te bepalen.

Verder biedt The Research Libraries Group een “coördinatiecentrum” aan. Leden kunnen hun informatie doorgeven aan deze groep en deze zal dan de gegevens indexeren en een zoekinterface genereren voor deze index. Dit laat onderzoekers toe met één enkele query te zoeken in honderden collecties.

Een voorbeeld van een EAD-bestand:

```
<filedesc>
  <titlestmt>
    <titleproper> Interview met een oudstrijder
      <date>1999</date>
    </titleproper>
    <author>Vereniging voor oudstrijders</author>
  </titlestmt>
  <notestmt>
    <note>
      <p>
        <subject>Wereldoorlog II</subject>
      </p>
    </note>
  </notestmt>
</filedesc>
```

Voordelen EAD:

- Ondersteunt hiërarchie
- Kan worden vertaald naar MARC en Dublin Core

Nadelen EAD:

- SGML is minder gebruiksvriendelijk

7.2.4 METS

De Metadata Encoding en Transmission standaard, kortweg METS, is een specificatie voor het beschrijven en uitwisselen van digitale objecten en hun eigenschappen [66]. METS is een open, niet-proprietaire standaard die werd ontworpen door de bibliotheekgemeenschap.

METS is XML-gebaseerd en biedt de middelen om metadata op te slaan voor zowel het beheren als het uitwisselen van digitale objecten. Door de XML-basis kent METS een hiërarchische structuur en kan het de hiërarchie uitdrukken van digitale objecten. Een METS-document wordt opgebouwd uit verschillende METS-elementen [67]. Deze elementen worden opgebouwd uit meerdere secties.

```
<mets>
  <dmdSec/>
  <amdSec/>
  <fileSec/>
  <structMap/>
  <structLink/>
  <behaviorSec/>
</mets>
```

Deze secties voorzien mogelijkheden voor het uitdrukken van de verschillende types metadata (zoals administratieve en beschrijvende) en informatie.

De secties dmdSec (Descriptive Metadata Section) en amdSec (Administrative Metadata Section) dienen als een soort wrappers waarin elementen van andere schema's kunnen worden geplugd. Deze wrappers zorgen er dus voor dat METS uitbreidbaar en modulair is. Voor de inhoud van deze wrappers kent METS geen eigen woordenschat en syntax. Deze worden verzorgd door de standaard die binnen de wrappers worden gebruikt. In de praktijk bestaan er reeds extensie-schema's voor bvb. Dublin Core en MARCXML die gebruik maken van deze techniek. De data in deze wrappers hoeft echter niet strikt tekstueel te zijn, ook binaire formaten zoals MARC21 kunnen hierin worden opgeslagen.

Voorbeeld van een dmdSec en een amdSec:

```
<mets:dmdSec ID="DMD1">
  <mets:mdWrap MIMETYPE="text/xml" MDTYPE="MODS">
    <mets:xmlData>
      <mods:mods version="3.1">
        <mods:titleInfo>
          <mods:title>Interview met een oudstrijder</mods:title>
        </mods:titleInfo>
        <mods:name type="personal">
```

```

      <mods:namePart>Jan De Smedt</mods:namePart>
    </mods:name>
    <mods:typeOfResource>audio</mods:typeOfResource>
  </mods:mods>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>

<mets:amdSec>
  <mets:rightsMD ID="ADMRTS1">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="METSRights">
      <mets:xmlData>
        <rts:RightsDeclarationMD RIGHTSCATEGORY="PUBLIC DOMAIN">
          <rts:Context CONTEXTCLASS="GENERAL PUBLIC">
            <rts:Constraints CONSTRAINTTYPE="RE-USE">
              <rts:ConstraintDescription>
                Het verdelen en/of kopiëren van dit
                object is enkel toegelaten mits
                toestemming van de rechthebbenden.
              </rts:ConstraintDescription>
            </rts:Constraints>
          </rts:Context>
        </rts:RightsDeclarationMD>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:rightsMD>
</mets:amdSec>

```

Na de dmdSec sectie en de amdSec sectie volgt de fileSec sectie. Deze sectie houdt bij welke bestanden tot het beschreven object behoren. Dit kan gebeuren door het digitale object zelf in te voegen of door een link naar dit bestand op te slaan.

```

<mets:fileSec>
  <mets:fileGrp USE="archive image">
    <mets:file ID="epi01m" MIMETYPE="audio/wav" ADMID="TECHWAV01">
      <mets:FLocat xlink:href=http://www.xxxx.com/01.wav
DOCTYPE="URL"/>
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>

```

Na de fileSec sectie volgt de structMap sectie. In de structMap struct wordt de hiërarchische structuur weergegeven van het digitale object. Dit laat toe de opbouw van het digitale object weer te geven. De structMap sectie laat toe meerdere hiërarchische structuren weer te geven per object. Zo kan men bvb. zowel een logische als een fysieke hiërarchie beschrijven. Een interview kan bijvoorbeeld in één bestand zijn opgeslagen (fysieke hiërarchie) maar meerdere “onderwerpen” bevatten (logische hiërarchie). Het weergeven van de hiërarchie gebeurt met behulp van divisies.

```
<mets:structMap TYPE="physical">
  <mets:div TYPE="book" LABEL="Het leven tijdens WOII" DMDID="DMD1">
    <mets:div TYPE="page" LABEL="Blank page"/>
    <mets:div TYPE="page" LABEL="Page i: Main title page"/>
    <mets:div TYPE="page" LABEL="Page ii: Blank page"/>
    <mets:div TYPE="page" LABEL="Page iii: Title page"/>
  </mets:div>
</mets:structMap>
```

Tot slot is er de structLink-sectie. Deze zorgt voor het weergeven van hyperlinks tussen de verschillende componenten van een METS-structuur die beschreven zijn in de structMap.

Een minder gebruikte sectie is de zogenaamde behaviorSec sectie. Deze voorziet METS van de middelen om digitale objecten te verbinden met toepassingen of programma code die in combinatie met andere informatie binnen het METS-document worden gebruikt voor het renderen of weergeven van het digitale object.

METS biedt ook verschillende profielen. Deze dienen als hulp voor het creëren van METS-documenten. Profielen bieden hiertoe een beschrijving van een klasse van METS-documenten in voldoende detail. Voor profielen is een schema beschikbaar. Deze profielen helpen ook programmeurs bij het creëren van software voor het gebruik en de processing van METS-documenten. Verder helpen ze ook bij de interoperabiliteit van digitale bibliotheken.

Een profiel bestaat uit een 13-tal componenten gaande van de titel, een abstract over extension schema's tot een voorbeelddocument.

Voordelen:

- Uitbreidbaar en modulair dankzij de wrapper-secties

Nadelen:

- Mogelijke veiligheidsproblemen bij het invoegen van programmacode in de behaviorSec-sectie
- Kleine community en userbase

7.2.5 LOM

LOM, of Learning Objects Metadata Standard, is een IEEE-standaard ontworpen om zogenaamde leerobjecten te kunnen beschrijven [68]. Dit kan bijvoorbeeld multimedia content zijn, educatieve content, leerobjectieven, enz. Deze standaard is ontworpen met het oog op het verkrijgen van een minimale set attributen die nodig zijn om de leerobjecten te beheren, lokaliseren en evalueren. De standaard ondersteunt onder andere security, privacy en evaluatie.

LOM definieert een basisschema dat de hiërarchie van data-elementen voor leerobjecten definieert. Op het hoogste niveau bestaan er negen categorieën:

- “General” die algemene informatie bevat over het leerobject in zijn geheel
- “Lifecycle” die informatie bevat over het verleden en de huidige staat van een leerobject, samen met wat het leerobject heeft beïnvloedt tijdens zijn evolutie
- “Meta-Metadatas” die informatie bevat over de metadata zelf
- “Technical” die informatie bevat over de technische eisen en karakteristieken van het leerobject
- “Educational” die informatie bevat over het educatieve en pedagogische karakter van het leerobject
- “Rights” die informatie bevat over de intellectuele eigendomsrechten
- “Relation” die de mogelijkheid biedt de relatie met verschillende leerobjecten weer te geven
- “Annotation” die commentaren kan bevatten over het educatieve gebruik van het leerobject en wanneer en door wie deze commentaren zijn toegevoegd
- “Classification” die het leerobject beschrijft in relatie tot een specifiek classificatiesysteem

Voor elk element specificeert LOM een naam, een uitleg, een grootte, een voorbeeldwaarde, een datatype en nog enkele andere basisdetails. Een voorbeeld van zo een element is “Technical.Location”. Dit is een element “Location” binnen het element “Technical”. Dit element geeft informatie over de plaats van het leerobject, bv. een URL. Sommige elementen kennen een beperkte woordenschat. Dit is een lijst van toegelaten waarden. Andere waarden worden evenwel toegelaten, dit ten koste van een lagere semantische interoperabiliteit. LOM laat verder ook toe data-elementen uit te breiden. Deze data-elementen mogen echter geen LOM-elementen vervangen met het oog op semantische

interoperabiliteit. Een voorbeeld hiervan is een element "Naam" toevoegen daar dit kan verward worden met het data-element "General.Title".

Voor LOM zijn er reeds bindingen ontwikkeld naar RDF [69] en XML [70]. Een LOM-element zou er in XML-vorm dan als volgt kunnen uitzien:

```
<lom xmlns="http://ltsc.ieee.org/xsd/LOMv1p0">
  <general>
    <title>
      <string xml:lang="nl">Interview met een oudstrijder</string>
    </title>
    <language>nl</language>
  </general>
  <technical>
    <location type="URI">
      http://www.interviews.org/oudstrijderx3242.mp3
    </location>
  </technical>
</lom>
```

Tot slot voorziet de LOM-standaard ook in een mapping naar Unqualified Dublin Core voorzien.

Voordelen:

- Heel flexibel en uitbreidbaar
- Mapping voorzien naar Dublin Core en binding met RDF
- Uitgebreid softwareaanbod [71]

Nadelen:

- Geen voorzieningen voor ontologieën

7.2.6 BibTeX

BibTeX is een bestandsformaat geschikt voor het representeren van lijsten van bibliografische objecten zoals artikels, boeken en thesissen [72]. De standaard kent echter een heel beperkt aantal velden want hem limiteert voor ander gebruik.

7.2.7 MPEG-7

MPEG-7 werd ontworpen door de Motion Pictures Expert Group (MPEG) [73]. Deze werkgroep is vooral bekend voor zijn standaarden voor het coderen van video en audio. MPEG-7 focust echter op het representeren van informatie over de content in plaats van de content zelf. MPEG-7 wil een rijke verzameling aan gestandaardiseerde hulpmiddelen bieden voor het beschrijven van multimediale content. De beschrijving van content moet kunnen ongeacht de wijze van opslag, de codering, de technologie, enz. Zo kan een beschrijving zowel handelen over een geprinte foto als over een interview in een digitaal audioformaat.

MPEG-7 bestaat uit Descriptors, Multimedia Description Schema's, Description Definition Language en hulpmiddelen die de binarisatie, de synchronisatie, het transport en de opslag van de descriptors voor hun rekening nemen.

Een Descriptor is de voorstelling van een kenmerk. Deze voorstelling is zowel syntactisch als semantisch vastgelegd. Een object heeft natuurlijk meerdere kenmerken en logischerwijs kan een uniek object door meerdere descriptors worden beschreven. Multimedia Description Schema's zorgen voor het weergeven van de structuur en de semantiek van de relaties tussen de verschillende descriptors, maar ook tussen andere Description Schema's.

Voor het definiëren van de structurele relaties tussen de descripties wordt gebruik gemaakt van een XML-gebaseerde taal, de Description Definition Language. Hiermee kunnen beschrijvende schema's worden gecreëerd, aangepast en gecreëerd.

MPEG-7 beschrijvingen laten toe verschillende dieptes van detail weer te geven. Hierdoor is het mogelijk bepaalde informatie weg te laten, of juist verder te verfijnen. Deze verfijningen verschillen natuurlijk per toepassingsgebied. Voor historische audiobronnen zou een beschrijving op hoog niveau kunnen zijn: "Interview met een oudstrijder". Op lagere niveaus kan dan gedetailleerdere informatie worden meegegeven. Zo kan de beschrijving op lager niveau uitgebreid worden met de naam van de oudstrijder, informatie over welke oorlog, enz.

Naast de beschrijvingen over de inhoud van een object kan ook extra informatie toegevoegd worden:

- Informatie over de creatie en de productie processen van de content
- Informatie over het gebruik van de content zoals copyright informatie en raadplegingen in het verleden
- Informatie over het opslagformaat van de content
- Informatie over collecties, interactie van de gebruiker met de content, enz.

Voordelen:

- MPEG-standaard

Nadelen:

- Voorlopig weinig industriële interesse
- Geen ontologie
- Nog steeds worden wijzigingen voorgesteld, specifiek voor het Query Format

MPEG-7 voorbeeld [74] :

<Mpeg7>

```
<Description xsi:type="CreationDescriptionType">
  <CreationInformation id="track4">
    <Creation>
      <Title type="songTitle">Interview met oudstrijder</Title>
      <Abstract>
        <FreeTextAnnotation>Interview over het leven van een
          oudstrijder</FreeTextAnnotation>
      </Abstract>
      <Creator>
        <Agent xsi:type="PersonType">
          <Name>
            <FamilyName>De Smedt</FamilyName>
            <GivenName>Jan</GivenName>
          </Name>
        </Agent>
        <CreationCoordinates>
          <Date><TimePoint>1999</TimePoint></Date>
        </CreationCoordinates>
      </Creator>
    </Creation>
  </CreationInformation>
</Description>
</Mpeg7>
```

7.2.8 Ondersteunende standaarden

Naast de eerder vermelde standaarden zijn er ook enkele standaarden ontwikkeld die archieven willen helpen bij het digitaliseren van hun collectie.

SPECTRUM is een open standaard die wordt onderhouden en gestuurd door MDA [75]. Deze standaard die procedures beschrijft voor het documenteren, de behandeling en de identificatie van objecten. Verder wordt er ook aandacht besteed aan bvb. rechtenbeheer, uitleenbeheer en risicobeheer. SPECTRUM evolueert continu en biedt dan ook de mogelijkheid tot groei en uitbreiding.

SPECTRUM is een Britse standaard die werd ontwikkeld met de hulp van de ervaring en het inzicht van honderden personen uit de museumbranche. Deze standaard wordt dan ook gezien als de “industriestandaard” voor documenteren.

Een tweede standaard is ISAD(G) of General International Standard Archival Description [76]. Deze standaard moet helpen bij het opstellen van beschrijvingen van collecties en objecten. De standaard bestaat uit verschillende regels die moeten gevolgd worden. Voor multi-level beschrijvingen (zoals mogelijk bij EAD) wordt bvb. aangeraden de beschrijvingen van algemeen naar bijzonder in te vullen en de plaats van een beschrijvingseenheid in de hiërarchie duidelijk te maken.

Op gelijkaardige manier worden ook regels gegeven voor het invullen van referenties, titels, datering, enz.

Een standaard afgeleid van ISAD(G) is SEPIADES (SEPIA Data Element Set) [77]. Deze standaard is gericht op het beschrijven en beheren van fotografische collecties en bevat 21 “Core”-elementen, aangevuld met meer dan 400 andere elementen. SEPIADES gebruikt een multi-level aanpak gelijkaardig aan ISAD(G). De standaard voorziet niet in een eigen encoding en is dus eerder een “handleiding” voor het beschrijven van collecties. Voor het opslaan van records wordt Dublin Core aangeraden.

Naast SPECTRUM en ISAD(G) is ook het ABC-model het vermelden waard [78]. Dit model is het resultaat van “The Harmony Project” en is ontworpen om in een gemeenschappelijk conceptueel model te voorzien dat de interoperabiliteit tussen verschillende metadata ontologieën van verschillende domeinen moet vergemakkelijken [79]. Hij is niet bedoeld als een metadatawoordenschat maar als een model dat als basis kan dienen bij het ontwerpen van specifieke ontologieën.

De kernbedoeling bij het ontwerpen van dit model was het voorzien van de mogelijkheid om het hele traject van een object weer te geven (naast de traditionele metadata). Dit maakt het

mogelijk de creatie, de evolutie en de overgangen die objecten meemaken te beschrijven. Zo kan worden beschreven waar en wanneer het interview is afgenomen, wie het afnam, transfers naar andere media, enz. Op deze manier kan de gehele levenscyclus van een object worden opgevraagd en dankzij de bidirectionele relaties kan ook informatie opgevraagd worden over elk object dat verbonden is met een object gedurende zijn levenscyclus.

Het ABC-model voorziet ook in een hiërarchie voor objecten en voor eigenschappen. Zo kunnen objecten worden georganiseerd in een hiërarchie waarbij elke subklasse extra informatie levert over het object. Deze hiërarchie vergemakkelijkt de interoperabiliteit tussen verschillende standaarden dankzij "partial understanding". Beschikt een bepaald object niet over een tegenhanger in de andere standaard dan kan men op zoek gaan naar een tegenhanger op een hoger niveau.

De hiërarchie voor eigenschappen laat eveneens toe deze te verfijnen met zogenaamde subeigenschappen. Hierdoor kan men niet alleen een gewenst informatieniveau bepalen (internen krijgen toegang tot alle informatie, externen enkel tot een bepaald niveau) maar ook de zoeknauwkeurigheid.

De hiërarchieën, levenscycli en bidirectionele relaties laten verder ook toe om van eenzelfde object/item verschillende representaties ter beschikking te hebben.

Tot slot is ook MarcOnt het vermelden waard [80] [81]. De MARC Ontologie is een ontologie die conform is met MARC21, Dublin Core en BibTeX metadata. MarcOnt wil echter ook meer bieden zoals tools die vertaling toelaten tussen de verschillende standaarden, dit samen met technologieën voor efficiënt doorzoeken en een efficiënte redentatie van de kennisdatabank.

Verder wil ze ook het probleem van communicatie tussen verschillende digitale bibliotheken wegnemen. Hierbij zou bijvoorbeeld een functionaliteit kunnen voorzien worden die bij communicatie tussen twee databanken uitzoekt welke functionaliteiten door beide databanken worden ondersteund en enkel deze functionaliteiten toelaten bij de communicatie. Op die manier kunnen meer geavanceerde bibliotheken alsnog communiceren met eenvoudigere bibliotheken. Deze functionaliteit zou een meer efficiënt gedistribueerd zoeken moeten mogelijk maken. Er wordt immers steeds maximaal gebruik gemaakt van de mogelijkheden.

MarcOnt probeert verder ook compatibiliteit te bereiken met al bestaande communicatiestandaarden zoals Open Archive Initiative.

Bovendien is MarcOnt “sociaal”. Het is gebaseerd op onderhandelingen tussen bibliothecarissen en software ontwerpers. Hiertoe is een website met stemmogelijkheden ontwikkeld. Voor wijzigingen worden toegepast vindt er een stemming plaats met behulp van een gewogen stemming. Hierbij wordt rekening gehouden met de “vriendschapsafstand” tussen de stemmer en de voorsteller, de verdiensten van het lid dat stemt en het aantal positieve/negatieve meningen van de andere gebruikers.

Deze standaard is echter nog in volle ontwikkeling en wordt pas verwacht tegen juni 2008.

7.2.9 MPEG 21

Het MPEG-21 Multimedia Framework is een open raamwerk dat tot doel heeft de levering van multimediale data en de consumptie ervan te definiëren voor alle spelers binnen de leverings- en verbruiksketen [82]. MPEG-21 is gebaseerd op 2 essentiële concepten. Het digitale item, een fundamentele unit voor distributie en transactie, en het concept van de gebruikers die interageren met deze digitale items. Samengevat kan men zeggen dat het hoofddoel van MPEG-21 er in bestaat een technologie te definiëren die gebruikers ondersteunt bij het uitwisselen, de toegang tot, het verbruiken, verhandelen of manipuleren van digitale items.

De gebruiker is een entiteit die binnen de MPEG-21 omgeving interageert met een andere gebruiker of gebruik maakt van een digitaal item. Zulke gebruikers kunnen individuen, verbruikers, gemeenschappen, organisaties, bedrijven, consortia, regeringen, enz. zijn. Gebruikers worden geïdentificeerd door hun relatie tot een andere gebruiker voor een bepaalde interactie. Puur technisch maakt MPEG-21 geen onderscheid tussen verbruiker en provider, beide zijn gebruikers. Een gebruiker kan gebruik maken van content op vele verschillende manieren (publiceren, verbruiken, enz.). Een gebruiker kan desondanks specifieke of zelfs unieke rechten en verantwoordelijkheden aannemen overeenkomstig zijn interactie met andere gebruikers binnen MPEG-21.

MPEG-21 biedt, in de basis, een raamwerk binnen welke een gebruiker interageert met een andere gebruiker met als onderwerp van interactie een digitaal item. Een interactie kan onder andere het creëren van content, archiveren van content en afleveren van content zijn. MPEG-21 bestaat uit 9 delen. DID of Digital Item Declaration beschrijft een set abstracte termen en concepten die een model vormen voor het beschrijven van digitale items. Deze termen en concepten zijn zo algemeen mogelijk en dienen als steunpunt bij het beschrijven van digitale items.

```
<dia:container>
  <dia:item>
    <dia:component>
      <dia:descriptor/>
      <dia:resource/>
    <dia:component>
  </dia:item>
</dia:container>
```

DII of Digital Item Identification neemt de identificatie van digitale items voor zijn rekening. DII ondersteunt onder andere het uniek identificeren van digitale items, het uniek identificeren van beschrijvende schema's en het identificeren van verschillende types digitale items.

```
<Statement>
  <di:Identifier>
    myID:1234
  </di:Identifier>
</Statement>
```

REL of Rights Expression Language heeft tot doel om rechten en toelatingen weer te geven in een machine-leesbare vorm. Het uitdrukken van een recht bestaat uit 4 entiteiten en de relaties ertussen: de gebruiker aan wie de rechten zijn toegekend, de rechten zelf, het object waarop de rechten van toepassing zijn en de voorwaarden voor het uitoefenen van de rechten.

Tot slot is er nog DIA of Digital Item Adaptation. Dit deel laat toe de karakteristieken bij raadpleging van data uit een archief weer te geven. Dit gaande van de mogelijkheden die de terminal of het netwerk bieden tot de omgevingsvariabelen en de voorkeuren van de gebruiker.

7.2.10 Overzicht

Titel	MARC/MARC21 Machine-Readable Cataloguing
Auteur	Network Development and MARC Standards Office of the Library of Congress
Uitgever	Library of Congress
Date	2002 (update)
Website	http://www.loc.gov/marc/bibliographic/ebcdhome.html
Rechten	Open Standaard
Beschrijving	Standaard voor het representeren en communiceren van bibliografische informatie in een machinaal leesbare vorm

Titel	Dublin Core The Dublin Core Element Set Version 1.1
Auteur	Dublin Core Metadata Initiative
Uitgever	Dublin Core Metadata Initiative
Date	1999
Website	http://dublincore.org/documents/1999/07/02/dces/
Rechten	Open Standaard
Beschrijving	The Dublin Core is een eenvoudig metadata element set met de bedoeling de inzage van elektronische bronnen te vergemakkelijken. Elementen kunnen gegroepeerd worden in elementen die gegevens bevatten over: inhoud - beschrijving, type, relatie, bron, onderwerp, titel; intellectuele eigendom – datum, formaat, taal, identificatiecode. Het gebruik van deze standaard is opgelegd door meerdere regeringen in Europa en verspreid over de wereld

Titel	EAD Encoded Archival Description
Auteur	University of California, Berkeley, Library
Uitgever	University of California, Berkeley, Library
Date	2002 (update)
Website	http://www.loc.gov/ead/index.html
Rechten	Niet-proprietaire standaard
Beschrijving	Standaard voor het beschrijven van collecties, vergelijkbaar met de MARC standaarden. Laat toe het detail van de informatie dat wordt weergegeven te bepalen/aan te passen

Titel	METS
Auteur	Library of Congress
Uitgever	Library of Congress
Date	2001
Website	http://www.loc.gov/standards/mets/
Rechten	Niet-proprietaire standaard
Beschrijving	Standaard voor het beschrijven van digitale objecten en hun eigenschappen. Ook het uitwisselen van deze gegevens wordt gespecificeerd.

Titel	LOM Learning Object Metadata
Auteur	IEEE
Uitgever	IEEE
Date	2002
Website	http://ltsc.ieee.org/wg12/
Rechten	
Beschrijving	Standaard die gebruikt wordt voor de beschrijving van leerobjecten en gelijkaardige digitale hulpbronnen gebruikt als ondersteuning bij leerprocessen

Titel	MPEG-7
Auteur	Motion Pictures Expert Group
Uitgever	Motion Pictures Expert Group
Date	2000
Website	http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
Rechten	
Beschrijving	MPEG-7 is MPEGs' visie op metadata.

Titel	BibTeX
Auteur	Oren Patashnik and Leslie Lamport
Uitgever	Oren Patashnik and Leslie Lamport
Date	1985
Website	http://www.bibtex.org/
Rechten	
Beschrijving	Standaard gebruikt voor het beschrijven en verwerken van referentielijsten, meestal in combinatie met LaTeX documenten

8 Haalbaarheidsstudie software systemen

Organisaties en instellingen worden geconfronteerd met de steeds grotere vraag om rijke digitale content te kunnen aanleveren. Een vlugge zoekopdracht op het web wijst uit dat complexe multimedia content zowel tekst, afbeeldingen, audio, als video combineert. Het aanleveren van zulke rijke digitale content wordt reeds mogelijk gemaakt door een overvloed van technologieën. Nu is het aanleveren op zich natuurlijk maar één aspect binnenin een suite van content management taken. Bronnen moeten ook aangemaakt, geïngest en bewaard worden. Ze moeten ook geaggregeerd en georganiseerd worden in collecties. Ze moeten verder ook nog door middel van metadata verder beschreven worden. Alle data moet verder beschikbaar zijn om te hergebruiken en/of te herwerken en uiteindelijk ook om bewaard te worden. Aldus moeten content managers kunnen gebruik maken van een flexibel digitaal depot waar ze op een uniforme manier bronnen kunnen opslaan, beheren en ontsluiten zodat al hun content (legacy, huidige en toekomstige) ook nieuwe vormen die onvermijdbaar zullen opduiken in de al dan niet verre toekomst zal kunnen blijven ondersteunen.

8.1 Open Source Media Asset Management raamwerken

8.1.1 Fedora

8.1.1.1 *Introductie*

Fedora (Flexible Extensible Digital Object and Repository Architecture) [83] is een open source raamwerk dat instellingen flexibele oplossingen aanbiedt om hun digitale bronnen te beheren en te ontsluiten. De kern bestaat uit een krachtig digitaal object model dat toelaat ingewikkelde relaties tussen de digitale objecten te beschrijven en dat het mogelijk maakt om verschillende voorstellingen van eenzelfde object te maken. Fedora's digitale objecten kunnen zowel lokaal als gedistribueerd bewaard en beheerd worden. Alle functies binnen Fedora, zowel op het niveau van de objecten als op het niveau van het depot zelf, worden als webservices aangeboden. Deze extreem flexibele functies kunnen beschermd worden door "fine-grained" access control policies. Deze unieke combinatie van kenmerken maken dat Fedora kan gebruikt worden in een verscheidenheid aan scenario's.

8.1.1.1.2 Projectsamenvatting

Fedora is begonnen als een DARPA onderzoeksproject in 1997 aan de Cornell University. De eerste referentie implementatie was CORBA gebaseerd. Van in het begin was het uitgangspunt al dat interoperabiliteit en uitbreidbaarheid best bekomen worden wanneer men een heldere en modulaire architectuur bouwt door data, interfaces en logica mooi van elkaar te scheiden. In 1999 heeft een team van de University of Virginia Library dit initiële raamwerk herwerkt naar een prototype die gebruik maakt van een relationele databank en een Java servlet engine om het ontsluitingsgedeelte van het depot te implementeren. Dit prototype heeft ertoe geleid dat beide universiteiten ervan overtuigd raakten dat de initiële architectuur een goede fundering zou bieden aan een praktisch, schaalbaar digitaal depot raamwerk. In september 2001 werden de handen in elkaar geslagen en door middel van een 1M\$ beurs werd begonnen aan het huidige digitale objecten repository raamwerk. In mei 2003 was Fedora v1.0 een feit. De huidige versie 2.2 heeft nog steeds tot doel om een open source asset management systeem en gerelateerde services aan te bieden om een verscheidenheid aan informatie management systemen. De Fedora software is vrij te downloaden en is beschikbaar onder de Educational Community License 1.0 (ECL) [84]. Al meer dan 50 grote onderzoeks- en culturele instellingen gebruiken dit raamwerk in een productieomgeving [85]. Deze software is heel matuur, goed getest en wordt gesteund door een grote actieve gemeenschap gebruikers. Ook in de toekomst zal dit raamwerk verder nog uitgewerkt en onderhouden worden [86].

8.1.1.1.3 Technologische overwegingen

Digitaal Object Model

Gezien vanuit het standpunt van een eindgebruiker hebben alle digitale objecten in een Fedora depot dezelfde karakteristieken als om het even welke web content. Ze zijn bereikbaar via standaard URLs [87] die mime-type [88] getypeerde stromen terug geven. Aldus is alles steeds te bereiken via standaard browsers of gebruikerstoepassingen die REST [89] URLs of SOAP [90] calls aankunnen. Uniforme toegang tot de representaties van de digitale objecten via URLs houdt de onderlinge structuur van deze objecten verborgen. Het eenvoudigste digitale object is niet meer dan een verzameling content items, waarvan elk content item een eigen representatie bezit. Het Fedora Object model definieert een component als een datastroom om een content item te representeren. Zo een datastroom component bevat de content intern als een bytestroom of bevat een externe referentie naar die content. In elk geval kan de te bevatten content elk media type aannemen. Bovenop deze rechtstreekse representaties -directe omzetting van datastromen dus-, laat het Fedora Object Model ook toe om virtuele representaties van een object aan te maken. Een virtuele

of afgeleide representatie is een view die bekomen wordt door één van de datastromen van het eigenlijke object te voeden aan een externe Fedora service. Aldus verkrijgt men terug een andere, dynamische representatie van het eigenlijk digitale object. Voor de eindgebruiker blijft dit, zoals al gezegd, volkomen transparant, aangezien deze alles benadert door middel van opgevraagde URLs. Deze manier van informatieafscherming maakt het mogelijk om onafhankelijk van de eindgebruiker veranderingen aan te brengen aan de software.

Het Fedora Object Model voorziet ook metadata die relaties van het object met andere objecten beschrijft. Daarvoor wordt binnen het raamwerk gebruik gemaakt van state-of-the-art semantische web technologieën. Relaties worden bewaard in een speciale datastroom als beweringen geëncodeerd in RDF (Resource Description Framework) XML formaat [91]. Deze relaties kunnen afgeleid worden van om het even welke ontologie, die op zijn beurt inplugbaar is in Fedora. Alle relationele metadata wordt door Fedora automatisch volledig geïndexeerd en bijgehouden in een aparte databank. Deze databank kan bevraagd worden door middel van semantische query talen, die ook als service kan aangeroepen worden via de Fedora API.

Services raamwerk

Vanaf versie 2.1 (februari 2006) gebeurt de verdere ontwikkeling ook binnen het Fedora service raamwerk. Dit extra raamwerk voorziet een structuur waarbij nieuwe services aangemaakt kunnen worden onafhankelijk van de eigenlijke Fedora kern [92]. Daarbij kan nieuwe functionaliteit ontwikkeld worden als een afzonderlijke stand-alone service die naadloos gekoppeld kan worden met de core Fedora repository service. Daardoor kunnen nieuwe functionaliteiten op een flexibele en modulaire manier ontwikkeld en ingepast worden en blijft de Fedora core intact. Waarschijnlijk zal deze extra service laag het belangrijkste instrument worden bij de verdere ontwikkeling van het gehele Fedora raamwerk. De twee belangrijkste services die tot nog toe ontwikkeld waren, zijn de Directory Ingest service en de OAI Provider service.

Preservation services working group

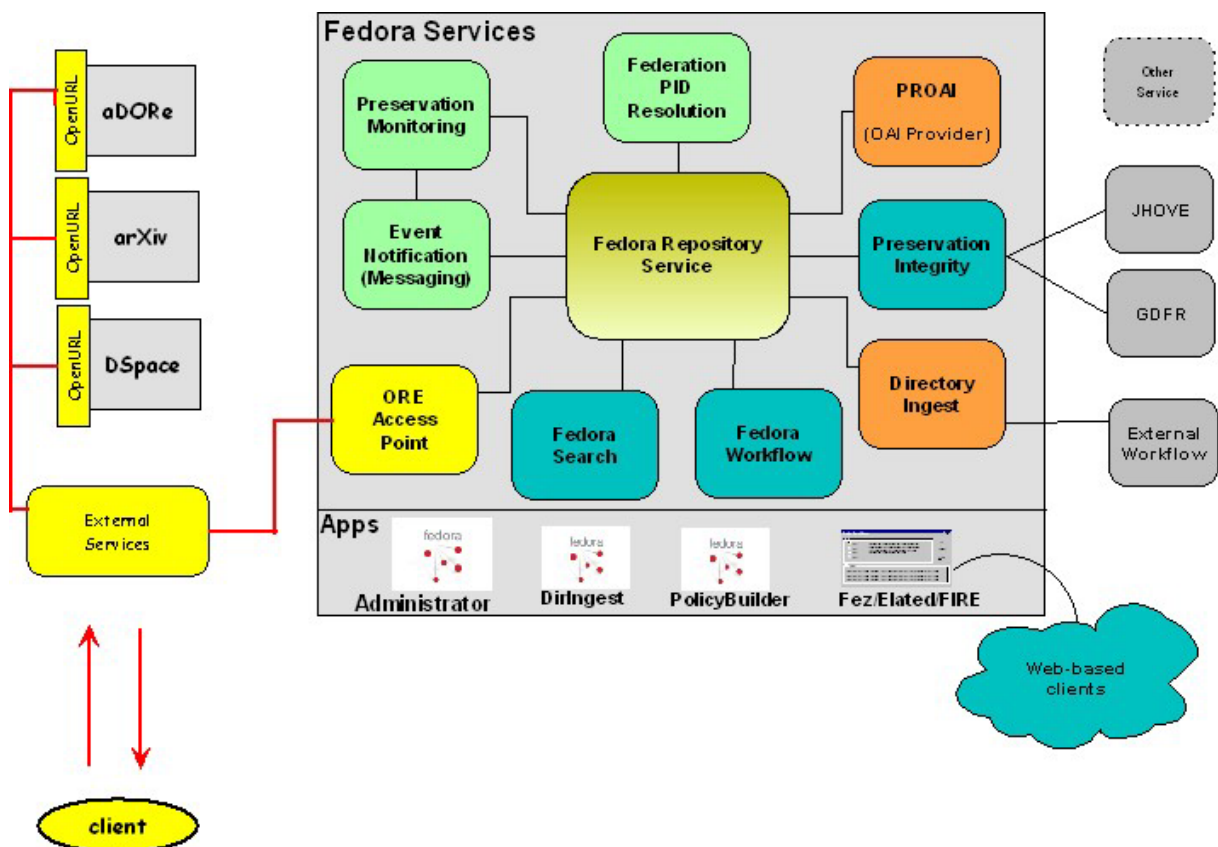
Deze specifieke werkgroep [93] onderzoekt en ontwikkelt services die specifiek archiveringsactiviteiten ondersteunen. Op dit ogenblik focussen ze zich op het ontwerpen van een messaging service die alle (toekomstige) archiveringsservices zal ondersteunen. Het is een generieke oplossing om boodschappen te zenden naar digitale depots over

archiveringsgerelateerde events. Verder biedt deze werkgroep ook al services aan betreffende formaattransformatie, formaatvalidatie, integriteitscontrole en depot history.

Workflow services working group

Deze specifieke werkgroep [94] onderzoekt en ontwikkelt services die specifiek archiveringsactiviteiten ondersteunen. Op dit ogenblik focussen ze zich op het ontwerpen van een prototype set business process en workflow orchestration services waardoor elektronische records gemakkelijker op een geautomatiseerde manier in Fedora kunnen geladen worden. Een dergelijke automatisatiestap is nodig om de mankracht te verminderen bij het processen van grote hoeveelheden elektronische records. Er zijn verschillende kandidaat standaarden om business processen te beschrijven in XML, maar BPEL (Business Process Execution Language) maakt wel de grootste kans om gekozen te worden.

Architectuur



8.1.1.1.4 Gebruik en institutioneel beleid

Security architectuur

Fedora voorziet ook een inplugbare authenticatie module. Deze kan gebruik maken van de methoden voorhanden binnen Tomcat - een standaard module dat gebruik maakt van een file waar gebruikers- en rolleninformatie in staat (tomcat-users.xml) of van een LDAP module om aan de gebruikers informatie van de LDAP-server te raken - of van de nieuwe access controle module dat gebruik maakt van de nieuwe, opkomende standaard XACML (eXtensible Access Control Markup Language).

Policies handhaven

Instellingen die voor de XAMCL-aanpak kiezen kunnen al hun XML-gecodeerde access control policies bewaren in Fedora, in plaats van het in eigen, legacy databanken of file formaten bij te houden. Daarenboven is XACML heel flexibel en laat het de specificatie toe van extreem fijnkorrelige policies. Fedora laat zowel policies toe die gelden voor het gehele digitale, als policies die inwerken op één enkel object. Policies gebaseerd op gebruikersattributen, attributen van digitale objecten of omgevingsvariabelen (bv. datum of tijd) kunnen aangemaakt worden die al dan niet toegang geven tot gelijk welke Fedora API acties. Ook kunnen evengoed fijnkorrelige policies aangemaakt worden die bv. de controle op één specifiek object als geheel of op één welbepaalde datastroom daarvan of één van zijn afgeleiden bewaakt. Binnen één specifieke workflow zouden administrators bv. ook permissies/policies veranderen naargelang het digitaal object zich doorheen die workflow beweegt.

8.1.1.1.5 Metadata en access

Het krachtige digitale object model kan elke combinatie qua aantal en verscheidenheid van datastromen aan. Deze datastromen kunnen lokaal aangeroepen worden of zich eender waar op het internet bevinden. Aangezien data en metadata uniform behandeld worden in het digitale object model, kan ook gelijk welk aantal en verscheidenheid van metadata formaten, naast de content zelf, bewaard worden als datastromen binnen het digitaal object. Dit maakt het metadata beheer zeer uitbreidbaar. Fedora's digitale objecten bevatten ook metadata die allerhande relaties tussen objecten kunnen uitdrukken. Deze expressieve inter-object relationele metadata, uitgedrukt in RDF-triples, is geïndexeerd en kan opgezocht worden door middel van semantische web query-talen.

Fedora's intelligente manier van web service integratie maakt het mogelijk om zich op 2 manieren in te passen in een n-tier applicatie. Enerzijds kunnen web services geassocieerd worden met om het even welke datastroom van het digitaal object, waardoor deze op een dynamische manier content kan aanleveren. Zo zou er bijvoorbeeld een metadata transformatie service kunnen aangemaakt worden die een MODS metadatastroom dynamisch kan omzetten naar het metadataformaat of metadatastroom van Dublin Core. Anderzijds is het zo dat de Fedora core repository zelf ook als een service runt binnenin een web server. Alle functionaliteit van het digitaal model zijn aldus bereikbaar via goed gedefinieerde REST en SOAP interfaces. Aldus kan een Fedora repository in om het even welke applicatieomgeving met verschillende user interfaces ingebed worden.

Fedora is verder ook volledig conform aan het interoperabiliteitsraamwerk zoals het gedefinieerd werd door het Open Archives Initiative [95] Protocol for Metadata Harvesting [96] (OAI-PMH). De Fedora OAI-PMH service benut aldus Fedora's uitbreidbare metadata management, zodat elke vorm van metadata aanwezig in de digitale objecten geharvest kan worden.

Het Fedora raamwerk heeft ook een paar features die de taak van het digitaal bewaren vereenvoudigt. Intern worden alle digitale objecten in het Fedora file systeem voorgesteld door middel van een open XML formaat. Deze XML-files bevatten data en metadata van de objecten en ook nog relaties naar services en andere objecten. De gehele Fedora structuur kan volledig heropgebouwd worden met behulp van de informatie die zich in deze files bevindt. Verder is Fedora ook volledig conform aan het referentiemodel van het Open Archival Information System (OAIS) [97] aangezien het SIPs (Submission Information Packages) en DIPs (Dissemination Information Packages) kan verwerken in standaard formaat containers zoals METS [98] en MPEG-DIDL [99].

8.1.1.1.6 Sterktes en zwaktes

Zwaktes

Het feit dat Fedora over een geavanceerd, extra services raamwerk beschikt, maakt het mogelijk heel uitbreidbaar. Alles daaromtrent staat of valt met de community die deze nodige services zal moeten ontwikkelen om nieuwe technologieën, technieken, standaarden en metadata schema's mee in te sluiten in het oorspronkelijke raamwerk. Verder zal de Fedora community er ook moeten voor zorgen dat bestaande, beproefde en erkende, externe tools - zoals formaat validatoren - services worden en probleemloos kunnen ingeplugd worden in de huidige Fedora raamwerk versie.

Verder is het voor elk open source project moeilijk om de status van “opstart project” achter zich te laten en een volwaardig professioneel product - met bijhorende support - te worden. Fedora heeft hier alvast een voorsprong op andere projecten door zijn diverse community die allerhande nieuwe services aanbrengt. Het zal wel zaak zijn om daar enige structuur in te krijgen door bijvoorbeeld een soort leiderschapsgraad te introduceren in deze community, waardoor de roadmap met nieuwe vereisten en de prioriteiten voor nieuwe Fedora services steeds naar één punt blijven convergeren.

Sterktes

Niettegenstaande dit voorgaande blijft Fedora in de eerste plaats een zeer open, uitbreidbaar raamwerk dat zijn strepen in de academische en culturele wereld al verdiend heeft. De uitgebreide community die dit product ondersteunt, zorgt dagelijks voor een nieuwe resem tools en services [100], waardoor er een rijke set services ontstaat die iedere Fedora gebruiker toelaat om de gewenste functionaliteiten te kiezen en samen te voegen om aldus zijn eigen unieke oplossing voor zijn eigen specifiek probleem te bouwen zonder zijn eigen tools en services te moeten ontwikkelen.

8.1.1.1.7 Conclusie

Fedora's raamwerk uitgebreid met de juiste policies, tools en services is een perfect betrouwbaar systeem dat als basis kan dienen voor het digitaal depot van elke instelling. Fedora's totale agnostisch karakter van file formaten en object types laat toe om elk datatype te beheren. Het kan door middel van het gebruik van RDF of METS metadata ook objecten met onderling complexe (hiërarchische) relaties behandelen. Fedora kan ook verschillende bitstromen van hetzelfde object beheren, waardoor het mogelijk wordt om zowel het originele digitale object en al zijn afgeleide transformaties te traceren en te bewaren. Versiecontrole en persistentie identificatie is mogelijk voor alle content objecten, hun metadata en hun afgeleiden. Door middel van XACML (eXtensible Access Control Markup Language) kunnen instellingen allerhande policies definiëren om hun objecten te beheren. Aldus is Fedora een gelaagd, transparant systeem waar alle objecten helder beschreven worden in XML (meestal FOXML of METS), waardoor het gemakkelijk wordt om data te migreren naar andere systemen.

Fedora's initiële object georiënteerde architectuur en recentelijk ontwikkelde services raamwerk zorgen ervoor dat nieuwe services ontwikkeld door de Fedora community gemakkelijk en elegant kunnen toegevoegd worden. Deze mogelijkheid om Fedora uit te breiden met totaal nieuwe functionaliteiten zorgt er op termijn voor dat de Fedora kern als

digitaal depot steeds nieuwe opkomende archiveringstechnologieën, technieken, standaarden en metadata schema's zal kunnen integreren.

Fedora is bijgevolg veel meer dan een out-of-the-box digitaal depot, het is een uitbreidbare architectuur waar elke instelling op verschillende manieren zijn repository kan bouwen, waarbij de ware definitie van bewaren en ontsluiten volledig kan uitgewerkt worden: het fysisch en intellectueel beveiligen en het technisch stabiliseren van het elektronisch aangeleverde bronmateriaal en zijn context over ruimte en tijd, zodanig dat er digitale kopieën beschikbaar zijn waarvan men mag aannemen dat deze authentiek zijn.

8.1.2 DSpace

8.1.2.1.1 Introductie

DSpace is een geavanceerd digitaal asset management systeem met als doel het langdurig archiveren en ontsluiten van formaatonafhankelijke (onderzoeks)objecten. Het is een open source, web gebaseerd systeem dat van op afstand beschikbaar is voor beheerders, aanbieders en het grote publiek. Het kan worden gewijzigd volgens de behoeften van de instellingen en is aldus een heel flexibel systeem dat kan gebruikt worden in een verscheidenheid aan scenario's [101].

8.1.2.1.2 Projectsamenvatting

DSpace werd in November 2002 ontwikkeld door de laboratoria van Hewlett-Packard (HP) en het technologisch instituut van Massachusetts (MIT). Het werd vrijgegeven onder de tolerante BSD open source licentie [102]. Daardoor kunnen eindgebruikers de code aanpassen aan hun noden en eventueel verbeteringen aanbrengen die naderhand door de DSpace projectbeheerders in een nieuwe distributie kunnen geïntegreerd worden. De laatste versie 1.4.2 van dit raamwerk is vrij te downloaden van de servers van de open source gemeenschap Sourceforge [103]. Al meer dan 250 onderzoeks- en culturele instellingen gebruiken dit raamwerk, waardoor we kunnen stellen dat deze software matuur is, goed getest is, gesteund wordt door een grote gemeenschap gebruikers en nog steeds verder evolueert [104]. Aangezien de software echter vrijgegeven is onder een open source licentie, leveren noch MIT, noch HP officiële support voor dit raamwerk.

Het raamwerk werd initieel ontworpen als een centrale opslagplaats waar instellingen (voor MIT in de eerste plaats de universiteit in al zijn geledingen: faculteiten, departementen, disciplines, laboratoria en onderzoekscentra) hun (onderzoeks- en gepubliceerde) bronnen konden ontsluiten en langdurig archiveren. De software is zo gebouwd dat elke mogelijke digitale opslagplaats als een productiedienst gebouwd kan worden. Daartoe werd het systeem “multidisciplinair” ontwikkeld rondom het concept “community” dat eigen workflows en collecties kan beheren. Dit digitaal depot kan alle soorten digitaal materiaal aan. Tekst, afbeeldingen, audio en video zijn de meest voor de hand liggende dataformaten, maar het raamwerk kan eigenlijk overweg met elk dataformaat, zij het data sets, complexe computermodellen, simulaties of zelfs ruwe binaire data [101]. DSpace is aldus een robuuste, flexibele implementatie van een digitaal depot waar gelijk welke instelling gelijk welk bronmateriaal in kwijt kan.

8.1.2.1.3 Technologische overwegingen

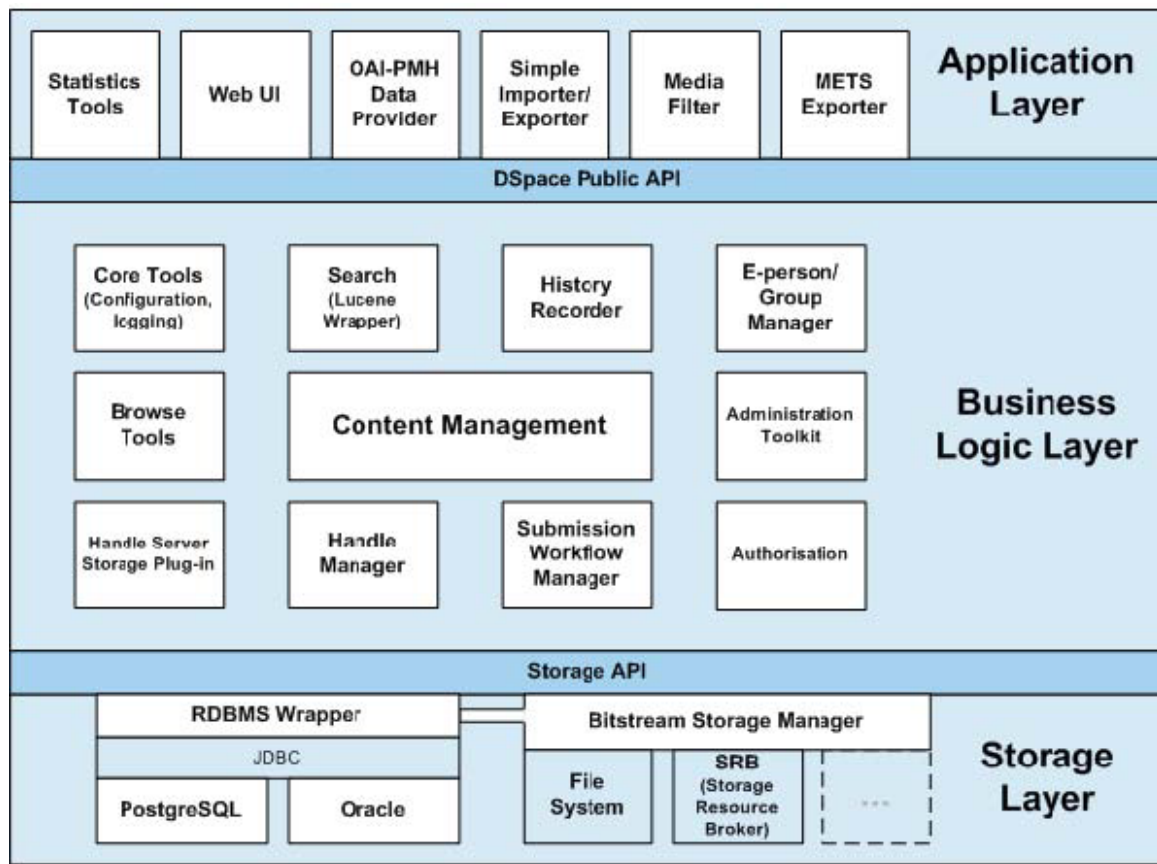
Vereisten

DSpace kan gebruikt worden bovenop de meeste gangbare besturingssystemen (Windows en UNIX), al is het initieel ontworpen bovenop een standaard UNIX platform. Het systeem zelf is verder opgebouwd uit een standaard open source gegevensbank (PostgreSQL) en een open source web server (Apache en Tomcat). De backend service is volledig Java gebaseerd. Iemand met een gedegen UNIX en Java kennis kan een basissysteem snel en gemakkelijk opzetten [105].

Support

Hoewel noch MIT, noch HP support aanbieden voor dit raamwerk, is er wel een heel actieve community die zowel rond het huidige raamwerk als de verdere ontwikkeling ervan in verscheidene vormen support aanbiedt. Behalve een DSpace Wiki [106] die zowel technische als niet-technische secties bevat, zijn er ook een aantal algemene en technische mailinglijsten [107] (zowel voor gebruikers als voor ontwikkelaars) waar heel actief op geparticipeerd wordt. Verder worden software fouten ook actief uit het systeem gehaald via een publieke “bug tracker” [104]. Het blijft natuurlijk zo dat “support” in een open source community “actiever” acties vereist van de instellingen die dit raamwerk gebruiken dan wanneer men kan terugvallen op de betalende support van een commercieel product. Een gedegen IT-professional in eigen rangen is voldoende om die serieuze meerprijs bij commerciële support te counteren.

Architectuur



8.1.2.1.4 Gebruik en institutioneel beleid

Content provisie

Na installatie kan men toegang krijgen tot het systeem via een drietal web interfaces: één voor de eindgebruikers, één voor de zij die nieuwe bronnen aanbrengen en één voor de beheerders. Materiaal aanbrengen gebeurt in ruwweg zeven stappen [108]: drie stappen die het materiaal beschrijven, het opladen van het materiaal, een verificatiestap, een licentiestap en een finaliseerstap. Deze stappen worden overzichtelijk voorgesteld door een progressbar. Men kan heel intuïtief tussen deze verschillende stappen blijven navigeren en technische kennis is geen vereiste. Getuige daarvan het feit dat men zelfs geen notie moet hebben van het bronformaat dat men wenst op te laden in het digitaal depot. DSpace analyseert zelf alle opgeladen bestanden (en dus ook het type) vooraleer het een definitieve bestemming te geven in het systeem. Formaten die ondersteund worden door de webbrowser, worden automatisch geladen bij bevraging, alle andere formaten worden gedownload en kunnen geraadpleegd worden doordat de juiste software (al dan niet automatisch -en dat hangt af van de instellingen van je browser-) geactiveerd wordt. Op het laatste kan men dan een licentie, al dan niet open source door middel van Creative Commons, verbinden met het opgeladen materiaal.

Communities

Het toevoegproces is in grote mate afhankelijk van de “policies” die voor een bepaalde “community” (groep) van toepassing zijn zoals deze door DSpace gedefinieerd worden. Een “community” heeft een variabele grootte (van een klein lab tot een volledige, grote instelling) en heeft bepaalde rechten in verband met submissions en ontsluiting die door deze interne “policies” afgedwongen worden. Wie iets wenst te deponeren in het digitaal depot hoeft niet specifiek verbonden te zijn aan één enkele “community”, maar ze moeten wel duidelijk één “community” opgeven waartoe dit opgeladen materiaal ondubbelzinnig zal toe behoren.

Policies

Waar het de “policy” van een “community” zou kunnen zijn om iedereen digitale mondelinge historische bronnen te laten opladen die automatisch ontsloten en gearchiveerd worden, is een meer complex en realistischer voorbeeld dat een groep bepaalde reviewers heeft, alsook iemand die verantwoordelijk is om alle metadata aan te vullen en een eindverantwoordelijke die alles goedkeurt. Elk stukje bronmateriaal zou op die manier steeds gereviewd, geëditeerd en goedgekeurd kunnen worden volgens de groepsregels vooraleer het ontsloten en gearchiveerd wordt. Elk individu met een bepaalde rol in het proces kan inloggen in het systeem om te zien in welke fase een stukje bronmateriaal zich bevindt en welke acties er nog moeten ondernomen worden door de verschillende actoren binnen de groep. DSpace noemt een dergelijke opeenvolging van handelingen een “workflow”. Het systeem is zo ontworpen dat allerhande “workflows” flexibel geïmplementeerd kunnen worden.

8.1.2.1.5 Metadata en access

Metadata

DSpace houdt drie soorten metadata bij over het gearchiveerde materiaal: beschrijvende metadata, administratieve metadata en structurele metadata.

DSpace ondersteunt verschillende “platte” metadata schema’s voor bronmateriaal te beschrijven. Een gekwalificeerd Dublin Core [109] metadata schema, gedeeltelijk gebaseerd op het “Library Application” profiel [110], wordt “by default”voorzien. Het is echter mogelijk om verschillende schema’s te configureren en een aantal metadata velden te kiezen uit een mix van geconfigureerde schema’s om je items te beschrijven. Andere beschrijvende metadata, bijvoorbeeld beschreven in hiërarchische schema’s, kan bijgehouden worden in gesimaliseerde bitstromen. Verder hebben “communities” en “collections” eenvoudige, beschrijvende metadata dat in de relationele databank van DSpace wordt bijgehouden.

Onder administratieve metadata bevindt zich bijvoorbeeld metadata over afkomst, preservatie en autorisatie policies. De meeste van deze metadata wordt ook in de databank van DSpace opgeslagen. Enkel metadata over afkomst, bitstream byte groottes en MIME types worden gerepliceerd in Dublin Core records, zodat dat gemakkelijk toegankelijk is van buiten DSpace.

Structurele metadata is bijvoorbeeld informatie over hoe een item aan een eindgebruiker kan getoond worden of over hoe verschillende delen van een item kunnen gerelateerd worden. Binnen DSpace wordt dit op een simpele manier geïmplementeerd: binnen een item worden bitstreams geordend in aparte bundels. Eén enkele bundel kan optioneel zelf ook een primaire bitstream hebben. Een HTML pagina op zich kan een item zijn en dus bestaan uit een aantal subitems. Zo zal DSpace aanduiden welke bitstream in de bundel de eerste HTML file is die naar de browser zal gezonden worden.

Integratie

Het gebruik van standaard metadata schema's laat toe dat DSpace kan integreren met andere digitale opslagsystemen door middel van het ter beschikking stellen van een "Open Archives Initiative" protocol API. Daardoor kan data van in DSpace "geharvest" worden door andere systemen.

Access

Gedigitaliseerde bronnen zijn uniek identificeerbaar via een "handle", waardoor objecten een persistent karakter krijgen. Deze "handles" worden beheerd door een speciale proxy server die bijhoudt welke "handle" er met welk object overeenstemt. Daardoor is het mogelijk dat een item kan veranderen in tijd en plaats zonder dat de eigenlijke URL verandert. Deze permanente URLs vergemakkelijken ook het langdurig archiveringsproces. File formaten en technologie zijn immers onderhevig aan verandering. In dit geval kunnen alle objecten in het archief eventueel vertaald worden naar andere file formaten zonder dat hun URL verandert, waardoor gebruikers in de toekomst dus steeds transparant alle materiaal nog steeds op de zelfde manier zullen kunnen benaderen.

8.1.2.1.6 Sterktes en zwaktes

Zwaktes

Natuurlijk heeft dit raamwerk ook een paar minpunten. Vooreerst is deze software zoals al vermeld open source. Onmiskenbaar heeft dit op zich heel veel voordelen, maar er zijn ook een paar minpunten aan verbonden. Tot op heden is er geen commerciële ondersteuning voor dit product. Culturele instellingen die tot op heden steeds beroep deden op commerciële producten en zelf maar weinig IT-kennis in huis hebben, zullen de nodige resources (tijd, geld en mankracht) daarvoor moeten vrijmaken. Misschien ligt het grootste minpunt wel in de manier waarop DSpace omgaat met metadata. Waar het aantal beschikbare velden in (Qualified) Dublin Core op zich waarschijnlijk wel voldoende is om de meeste (zo niet alle) Use Cases te beschrijven, laat de machtigingscontrole over het invullen van die velden te wensen over. Zonder enige controle over uniforme titels, auteurs of onderwerpen kan het opzoeken en opvragen van materiaal in de toekomst problematisch worden. Dit kan echter opgelost worden door afspraken te maken op institutioneel niveau, of nog beter op sectoraal niveau, waarbij overkoepelende “policies” en richtlijnen, over via welke “workflows” en rollen materiaal binnengebracht dient te worden, de semantiek van de metadata ondubbelzinnig bepalen. Het op één of andere manier kunnen toevoegen van een “controlled vocabulary” zou een meerwaarde betekenen voor een volgende versie van dit raamwerk.

Sterktes

Niettegenstaande deze minpunten heeft DSpace vooral veel voordelen. Het is een zeer flexibel en robuust systeem dat eigenlijk ogenblikkelijk in gebruik genomen kan worden. Het kan alle types digitaal materiaal aan en niet-technische gebruikers kunnen zelfs op een eenvoudige manier gebruik maken van de meer ingewikkelde opties zoals de “Open Archive Initiative” API. De software zelf kan volledig op maat van de instelling gemaakt worden en met de nodige IT-kennis in huis kunnen alle problemen zelf opgelost worden. Zelfs al heeft de instelling geen IT-kennis in huis, dan nog is het gemakkelijk op een “default” installatie uit te voeren. Belangrijk is vooral te weten dat het systeem gratis is en kan geïnstalleerd worden op elke “mainstream”computer.

8.1.2.1.7 Conclusie

De DSpace ontwikkelaars hadden tot doel om het collecteren, archiveren, indexeren en distribueren van digitale content laagdrempelig te maken, zodanig dat gebruikers hun eigen werkmateriaal zouden “zelf-archiveren”. Behalve dat ene feit dat er te weinig semantische controle is over de metadata, zijn ze er wonderwel in geslaagd om hun doelen te realiseren en aldus een voortreffelijk digitaal depot te ontwikkelen dat ondersteund wordt door een

grote, actieve community eindgebruikers en ontwikkelaars. DSpace is uiteindelijk een heel flexibel systeem geworden waar elk type digitale content kan in opgenomen worden en dat elk institutioneel rollenmodel kan implementeren. Deze flexibiliteit wordt nog vergroot door het feit dat open source oplossingen steeds zelf kunnen aangepast worden en dat al deze verbeteringen uiteindelijk terug kunnen vloeien naar de volledige DSpace community waardoor de volgende versie nog meer mogelijkheden zal hebben.

8.1.3 JeromeDL

8.1.3.1.1 Introductie

JeromeDL is een sociale semantische digitale bibliotheek met als doel het publiceren van documenten op het Web te vereenvoudigen samen met het doorzoeken en doorbladeren ervan. De documenten kunnen zowel tekstueel als binair zijn.

JeromeDL biedt sociale en semantische diensten aan die elke gebruiker moeten helpen bij het doorzoeken en doorbladeren van de informatie binnen het systeem. Elke gebruiker kan bookmarks, annotaties en kennis delen over JeromeDL en zo anderen helpen bij hun zoektocht. Ook werden mogelijkheden voor verbindingen met andere digitale bibliotheken voorzien.

8.1.3.1.2 Projectsamenvatting

JeromeDL is een gezamenlijk project van de Technologische Universiteit van Gdansk (GUT) en DERI, de universiteit van Galway. JeromeDL wordt vrijgegeven onder de BSD open source licentie. Daardoor kunnen eindgebruikers de code aanpassen aan hun noden en eventueel verbeteringen aanbrengen die naderhand door de DSpace projectbeheerders in een nieuwe distributie kunnen geïntegreerd worden. Het systeem is gebaseerd op het J2EE-platform, een RDF databank en Semantisch Web technologieën. De laatste versie 2.0.2 van het JeromeDL raamwerk is vrij te downloaden van de servers van de open source gemeenschap Sourceforge [111]. Deze versie wordt als “onderzoekstabiel” beschouwd. Een twintigtal ontwikkelaars helpt aan de verdere ontwikkeling van JeromeDL. Momenteel wordt er in 2 stromen aan JeromeDL gewerkt. Enerzijds zorgt DERI voor industriële stabilisatie en ondersteuning en anderzijds wordt er meer nadruk gelegd op onderzoek binnen de sleutelcomponenten zoals MarcOnt en HyperCuP.

Het JeromeDL raamwerk werd ontworpen vanuit het standpunt dat gewone bibliotheken zich verspreid bevinden over een grote oppervlakte en geen onderlinge connectie hebben en dat bij de huidige digitale bibliotheken de gebruiker alleen staat, geen hulp krijgt bij het vinden van de juiste sleutelwoorden en de verbinding tussen de verschillende digitale bibliotheken met vele moeilijkheden gepaard gaat. Om deze nadelen te overwinnen werd JeromeDL ontworpen als een sociale semantische digitale bibliotheek. Sociaal door zijn interactie tussen gebruikers, het bestaan van communities en het verbeteren van de inhoud door middel van blogs en Wikis. Ook kunnen gegevens over andere gebruikers en hun blader- en zoekgedrag helpen bij zoekopdrachten. JeromeDL is gebaseerd op RDF wat een basis geeft voor de verbinding tussen verschillende digitale bibliotheken.

Verder maakt JeromeDL ook gebruik van zogenaamde ontologieën, welke de semantische kant van de digitale bibliotheek voor hun rekening nemen.

8.1.3.1.3 Technologische overwegingen

Vereisten

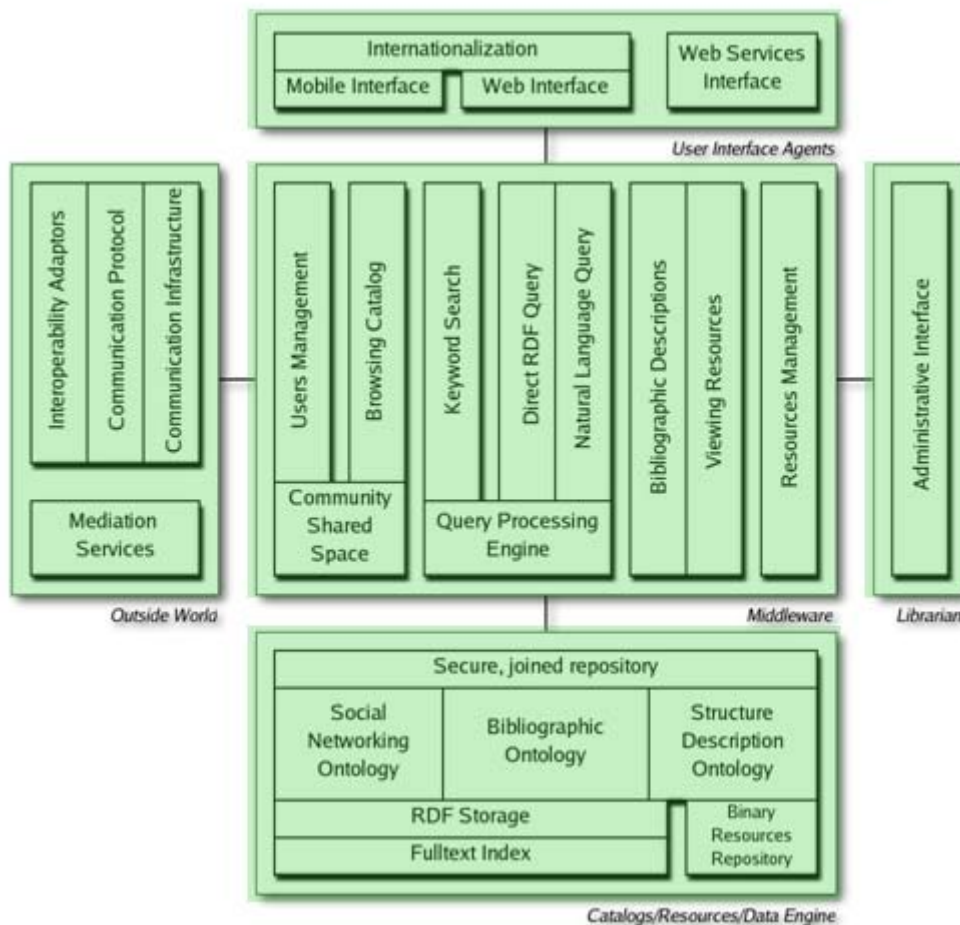
JeromeDL kan gebruikt worden bovenop de meeste besturingssystemen, zoals Mac OS X, Windows 32-bit en BSD- en POSIX-gebaseerde besturingssystemen. Het systeem zelf is verder opgebouwd uit een open source web server (Tomcat) en Sesame, een open-source Java raamwerk voor het opslaan, doorzoeken en werken met RDF. Het kent dus alle eigenschappen van een databank, al kan het ook gebruikt worden als een bibliotheek voor het werken met RDF.

Support

JeromeDL is open source en kent daardoor, zoals bij de meeste open source projecten, geen industriële support (al werkt DERI hier aan). Wel kan JeromeDL steunen op een community die verscheidene vormen van support aanbiedt. JeromeDL kent een Wiki [112], een mailinglist voor ontwikkelaars en één voor gebruikers. Verder is er ook een vrij uitgebreide gebruikersgids beschikbaar, aangevuld met instructievideo's, op de website en dit voor zowel gebruikers als bibliotheekmedewerkers [113]. Verder worden softwarefouten actief uit het systeem gehaald via een bug-tracker [114]. JeromeDL wordt dan ook vrij regelmatig bijgewerkt.

Architectuur

JeromeDL bestaat uit een centrale structuur die verantwoordelijk is voor het beheer van de community gegevens zoals het gebruikersbeheer en informatie over het gedrag van de gebruikers. Daarnaast bevat deze structuur ook de zoekengine, de beschrijvingen van de documenten en het beheer van deze informatie. Verder kent JeromeDL nog vier onderdelen die dienen als interfaces tussen de centrale structuur en de buitenwereld, de gebruikers, de bibliotheekmedewerkers en de verwerkingseenheid voor de verschillende ontologieën.



8.1.3.1.4 Gebruik en institutioneel beleid

Content provisie

Na de installatie kan men toegang krijgen tot de webinterface van JeromeDL. Voor het aanbrengen van materiaal hoeft men geen administrator te zijn. Het aanbrengen gebeurt in verschillende stappen. Eerst moet het bestandsformaat gekozen worden (PDF, RTF, SWF, SCAN, URI), vervolgens wordt de basisinformatie, zoals de titel, auteurs en een abstract, ingevoerd. Daarna kunnen de hoofdstukken of onderverdelingen van het materiaal ingevuld worden. In de volgende stap wordt het materiaal geannoteerd met sleutelwoorden en categorieën. In de laatste stap kunnen andere binaire bronnen toegevoegd worden (zoals eigenschappen uit de MarcOnt ontologie of een MARC21- of BibTeX-beschrijving). Gedurende dit stappenproces kunnen ook de gewenste protecties ingegeven worden zoals protectie voor kopiëren, toegang na betaling, enz.

Policies

Wanneer een lid van een community gegevens toevoegt aan de databank worden deze gegevens niet zomaar toegevoegd. Om de kwaliteit van de bronnen in de bibliotheek te garanderen moet elke bron namelijk goedgekeurd worden door een administrator. Ook bij wijzigingen moet dit gebeuren.

8.1.3.1.5 Metadata en access

Metadata

JeromeDL ondersteunt verschillende types bibliografische metadata. Een bron kan voorzien worden van zowel DublinCore, BibTeX als MARC21 gecodeerde metadata. Dit laat toe om gebruik te maken van al bestaande bronnen van bibliografische beschrijvingen. Eén bron kan meer dan één beschrijvingen tegelijkertijd bevatten.

Elke bron bevat bovendien een beschrijving van diens structuur. Dit laat zowel flexibiliteit als specialisatie toe bij het verwerken van bronnen.

Zoeken en bladeren

Het JeromeDL systeem heeft tot doel een hoge kwaliteit van functionaliteiten voor zoeken en bladeren te voorzien. De verscheidenheid aan metadata gebruikt binnen het JeromeDL systeem toont bijvoorbeeld de intelligentie van het zoekalgoritme aan. Bij het zoeken naar informatie wordt verder met vele gegevens rekening gehouden. Zo wordt er rekening gehouden met het gedrag van de gebruiker (en andere gebruikers) bij het doorbladeren van informatie en de concepten inhoud en betekenis. Verder kan een gebruiker ook informatie, zoals bookmarks, annotaties en eigen kennis, delen met andere gebruikers wat hen kan helpen bij het doorbladeren van informatie. Gebruikers kunnen ook commentaar toevoegen over de inhoud van een bron en antwoorden op commentaar van andere gebruikers. Op deze manier wordt als het ware nieuwe kennis gecreëerd.

8.1.3.1.6 Sterktes en zwaktes

Sterktes

JeromeDL kent een heel uitgebreide ondersteuning voor bibliografische metadata en ondersteunt ook de combinatie van zulke standaarden. Verder kent JeromeDL uitgebreide mogelijkheden voor communicatie tussen gebruikers, zoals een Wiki en een mailinglijst. Bovendien kan elke gebruiker mee helpen het zoeken en doorzoeken van andere gebruikers te vereenvoudigen. Het zoekalgoritme van JeromeDL kan namelijk gebruik maken van informatie die het verzamelt door het zoekgedrag te analyseren van de andere gebruikers.

Zwaktes

De voordelen van JeromeDL brengen ook enkele mogelijke nadelen met zich mee. Allereerst hangen ze voor hun efficiëntie af van de grootte van de gebruikersgemeenschap en deze is momenteel nog heel beperkt. Verder vereist het zoekalgoritme voldoende gegevens over gebruikers. Deze kunnen echter om privacy redenen delen van deze gegevens blokkeren, wat het sociale karakter JeromeDL enigszins kan tegenwerken.

8.1.3.1.7 Conclusie

JeromeDL is een veelbelovend raamwerk met ondersteuning voor vele technieken die het leven van de gebruiker moeten aangenamer maken en de gebruiker moeten helpen bij de zoektocht naar informatie. Het vormt als het ware een mengeling tussen de fysische bibliotheek, waar men iemand om hulp kan vragen wanneer men een term niet meer kent of voor het zoeken naar gelijkaardige informatie maar beperkt is tot één collectie, en een zoekmachine, waar men er alleen voorstaat, geen hulp krijgt bij een term maar wel vele collecties kan doorzoeken. De semantische en sociale technieken binnen JeromeDL kunnen de gebruiker helpen bij zijn zoektocht en JeromeDL biedt ook de mogelijkheid verschillende collecties met elkaar te verbinden. Bovendien is er dankzij de mogelijkheid tot internationalisatie ook een lagere drempel voor gebruik, men kan immers in zijn eigen taal werken.

Voorlopig kent JeromeDL echter nog een heel beperkte gebruikersgemeenschap wat de drempel voor implementatie van dit raamwerk kan verhogen daar meer technische kennis nodig is en minder ondersteuning bestaat.

8.1.4 Bricks

8.1.4.1.1 Introductie

BRICKS is een digitale bibliotheek met als doel het ontwerpen en ontwikkelen van een open, gebruikers- en dienstgeoriënteerde infrastructuur om kennis en bronnen te delen binnen het culturele erfgoed domein. Verder wil BRICKS de gebruikers en instituten voor cultureel erfgoedbeheer de mogelijkheid geven hun bronnen en diensten te delen met andere gebruikers. BRICKS is volledig p2p-gebaseerd en dus gedecentraliseerd. Elk lid van de BRICKS community heeft een zogenaamde BNode die mee het BRICKS netwerk vormt en kan communiceren met andere leden. Een van de hoofddoelen van BRICKS was een zo laag mogelijke kost. Naast de keuze voor P2P werd er dan ook voor gekozen ervoor te zorgen dat leden niets moesten veranderen aan de interne structuur van hun eigen bibliotheken. Ze hoeven enkel een OWL-DL schema te ontwikkelen voor de metadata standaard die ze gebruiken.

8.1.4.1.2 Projectsamenvatting

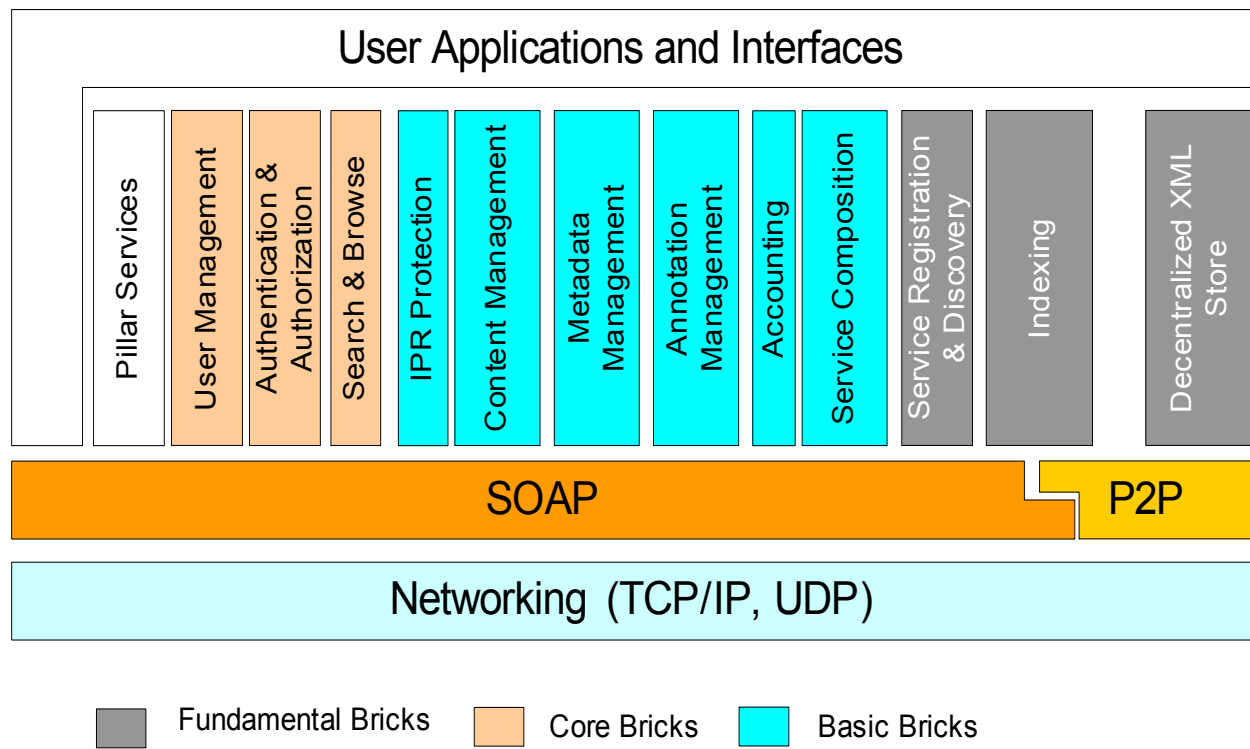
Het BRICKS-project ging van start in januari 2004 met een voorziene duur van 42 maanden en liep af juni 2007 [115]. Het is een “Integrated Project” dat mede werd gefinancierd door het “European Sixth Framework Programme” dat tot doel heeft de bestaande digitale bronnen in een gemeenschappelijke en gedeelde digitale bibliotheek te integreren. Bricks wordt vrijgegeven onder de LGPL licentie [116]. De laatste versie van BRICKS is een 3e prototype (v0.42) en is vrij te downloaden op de officiële BRICKS-website. BRICKS kent 24 partners, 9 research centers en 5 technologische providers. Verder maken al 116 culturele organisaties gebruik van BRICKS.

8.1.4.1.3 Technologische overwegingen

Architectuur

BRICKS is opgebouwd uit verschillende componenten die opgebouwd zijn als web services. Deze componenten kunnen worden opgedeeld in 3 categorieën. Helemaal onderaan bevinden zich de basisblokken. Deze zijn nodig op een BNode voor de werking en het onderhouden van het lidmaatschap binnen de community. Naast de p2p-laag behoort hiertoe ook de gedecentraliseerde XML-opslag, dienstenregistratie en het indexbeheer. Bovenop dit blok bevindt zich het kernblok. Dit blok bevat de componenten nodig om gebruikers toegang te verlenen tot BRICKS. Onder de componenten vinden we het gebruikersbeheer, gebruikersautorisatie en gebruikersauthenticatie maar ook de zoek- en bladerfuncties. De derde categorie zijn de basisblokken die optioneel zijn en enkel worden ingezet indien nodig.

Onder deze categorie bevinden zich onder andere de componenten voor content- en metadatabeheer en het annotatiebeheer.



8.1.4.1.4 Gebruik en institutioneel beleid

BRICKS kent twee verschillende "demo"-interfaces: een offline, desktop interface en een online webinterface. Deze kennen beide een verschillende functionaliteit.

De webinterface, Workspace genaamd, heeft tot doel begrijpbaar en bruikbaar te zijn. De interface laat de gebruikers toe om collecties te beheren, objecten toe te voegen aan collecties of er uit te verwijderen, te zoeken binnen BRICKS, content op te laden en te organiseren en annotaties te beheren. De webinterface wordt gezien als de referentieapplicatie. De webinterface is gericht op software ontwikkelaars en gewone eindgebruikers. Deze gebruikers kunnen indien gewenst ook anoniem gebruik maken van BRICKS maar kennen dan beperkte functionaliteit. Geregistreerde gebruikers kunnen bijvoorbeeld de instellingen personaliseren, content opladen en collecties creëren.

De desktop interface is een uitgebreide client-applicatie die de functionaliteiten biedt om de digitale objecten te organiseren, beheren en becommentariëren. Verder kan deze interface ook nieuwe collecties creëren van zoekresultaten, content en metadata importeren en exporteren, gebruikersrechten beheren, enz. Deze interface is door zijn grotere functionaliteit

vooral gericht op softwareontwikkelaars en deskundige eindgebruikers zoals onderzoekers en onderwijzers.

Beide interfaces zijn overzichtelijk en afgestemd op hun doelgroep. De verschillende functionaliteiten worden overzichtelijk weergegeven. De website van BRICKS biedt bovendien enkele handleidingen in videovorm aan die de gebruiker stap voor stap door de webinterface begeleiden.

Met het oog op verdere ontwikkeling zijn de mogelijkheden voorzien voor het rapporteren van bugs [117]. Verder is er een mailing list voorzien voor gebruikers [118].

8.1.4.1.5 Metadata en access

Metadata

BRICKS ondersteunt verschillende arbitraire metadata standaarden zoals Dublin Core en MARC maar ook andere standaarden waarvan een OWL-DL schema beschikbaar is. Intern worden de gegevens in RDF gerepresenteerd. De component die de metadata beheert is dan ook opgebouwd uit twee lagen. De onderste laag is een RDF-laag waarbinnen de eigenlijke queries worden uitgevoerd. Deze queries kunnen zowel eenvoudige full-text, geavanceerdere veld- als ontologiegebaseerde domeinqueries zijn. De tweede laag is de presentatielaag die zorgt voor het weergeven van de metadata waarbij de RDF graaf wordt gemaskeerd. Vocabularia en thesauri worden ondersteund zolang ze kunnen worden gerepresenteerd in RDF, RDFS of OWL.

Integratie

Het gebruik van standaard metadataschema's in OWL-DL laat toe BRICKS te integreren met andere digitale opslagsystemen. Dit was een van de vereisten voor het raamwerk. Er moet niets gewijzigd worden aan de eigen systemen, enkel een OWL-DL schema moet worden voorzien. Voor het importeren wordt dan gebruik gemaakt van de ter beschikking gestelde "Open Archives Initiative" protocol API. Daardoor kan data uit BRICKS "geharvest" worden door andere systemen.

Access

Binnen BRICKS bevinden de digitale bronnen zich ofwel lokaal bij één van de leden ofwel op een via een URL gedefinieerde plaats. Het opzoeken van informatie gebeurt op een voor P2P typische manier door het aanspreken van buurnodes die op hun beurt weer buurnodes aanspreken. Verder zijn er ook context-aware policies aanwezig via XACML en wordt ook

DRM en watermerken ondersteund voor het beschermen van de content. Voor DRM wordt gebruik gemaakt van de MPEG-21 REL-standaard.

8.1.4.1.6 Sterktes en zwaktes

Sterktes

BRICKS kent als grootste voordeel de lage kosten. Het opstarten van een eigen BNode brengt weinig werk en kosten met zich mee, er hoeft immers aan de interne bibliotheek niets gewijzigd worden. Enkel de verbinding met BRICKS via het OAI-PMH protocol moet worden toegevoegd. Verder is BRICKS P2P-gebaseerd wat ervoor zorgt dat centraal beheer niet nodig is en het systeem daardoor aan betrouwbaarheid wint. BRICKS ondersteunt verder elk type content en metadata en kent een uitgebreide personalisatiemogelijkheid. Verder is BRICKS ook multilinguaal en beschikt het naast een webinterface ook over een GUI-interface.

Zwaktes

Een “zwak” punt van BRICKS is het open source karakter. Hoewel dit vele voordelen met zich meebrengt, zijn er ook enkele nadelen aan verbonden. Zo is er geen commerciële ondersteuning voor dit raamwerk. Culturele instellingen die willen gebruik maken van BRICKS moeten dus steunen op de community. En net deze community is niet groot en niet zo goed georganiseerd als bij sommige andere projecten. De Wiki-website kent nog vele gaten en de bestaande documentatie is niet altijd even begrijpelijk voor niet-ingewijden. Een nadeel lijkt ook dat anonieme gebruikers geen wijzigingen kunnen aanbrengen of objecten kunnen toevoegen. Mits een hiërarchie die wijzigingen pas toepast na controle kan dit een positieve invloed hebben op de compleetheid van informatie en collecties.

8.1.4.1.7 Conclusie

BRICKS is de jongste en kleinste speler van de hier vergeleken systemen. De makers spreken bij versies dan ook nog van een prototypestatus. BRICKS onderscheidt zich van alle andere systemen door zijn kostenefficiënte aanpak. Het raamwerk is compleet P2P-georiënteerd en vereist weinig ingrepen om eigen bibliotheken toe te voegen aan het raamwerk. De gebruikersbasis van het raamwerk is voorlopig nog heel klein want vermoedelijk te maken heeft met het nog maar recent aflopen van het project rond de ontwikkeling ervan (juni 2007), de prototypestatus en de soms beperkte gebruikersinformatie op de website en Wiki. Verder is het de vraag of, nu het project is afgelopen, de ontwikkeling van het raamwerk voort zal lopen.

8.1.5 Samenvatting

8.1.5.1.1 *Generieke aspecten*

	Fedora	DSpace	JeromeDL	Bricks
OS support	Alle gangbare	Alle gangbare	Alle gangbare	Alle gangbare
HW vereisten	500Mb RAM, min 100Gb HD	500Mb RAM, min 100Gb HD	500Mb RAM, min 128Gb HD	500Mb RAM, min 100Gb HD
SW vereisten	Java 1.5, Kowari/Mulgara or MPTStore	Java 1.4, Tomcat 5.5, PostgreSQL 7.3	Java 1.5, Tomcat 5.5, Sesame	Java 1.4/1.5, Jena
Huidige status	Productie v2.2	Productie v1.4.2	Research stabiel, v2.0.2	3de prototype
# installaties	~50 monitored, heel veel downloads	+250 universiteiten en culturele instellingen	12+, >200 downloads	~8
Support model	Open Source	Open Source	Open Source	Open Source

8.1.5.1.2 *Architecturale aspecten*

	Fedora	DSpace	JeromeDL	Bricks
Distributie	Federatie via namerresolver, Alvis P2P gedistribueerde zoekmachine	Niet gedistribueerd	Gedistribueerd zoeken (P2P), geagregeerd browsen (hierarchisch)	Volledig gedecentraliseerd (P2P)
Architectuur granulariteit	Hoog (core repository met configureerbare modules, loosely coupled)	Laag (slechts enkele generieke bouwblokken)	Laag (slechts enkele generieke bouwblokken)	Hoog (veel componenten)
DB Support	Postgres, MySQL, Oracle, McKoi, Kowari/Mulgara	Postgres, Oracle	Alle Sesame-compliant backend systemen	H2, HSQL, Postgres, MySQL, Oracle, SQLServer

8.1.5.1.3 Content & metadata

	Fedora	DSpace	JeromeDL	Bricks
Content types	Alle soorten	Alle soorten	Alle soorten	Alle soorten
Content modellen	# inplugbaar	# inplugbaar	JeromeDL ontologie	# inplugbaar
Metadata schema's	Alle XML & RDF/S & OWL schema's	Qualified Dublin Core	MarcOnt + extensies	Alle RDF/S & OWL schema's
Query types	Full-text (Lucene of Zebra backend), veld gebaseerd zoeken, ontologie gebaseerd (itql, rdql, sparql, spo)	Full-text (Lucene backend), veld gebaseerd zoeken	Full-text, veld gebaseerd zoeken, ontologie gebaseerd, NL query templates	Full-text, veld gebaseerd zoeken, ontologie gebaseerd (sparql)

8.1.5.1.4 Beveiliging en DRM aspecten

	Fedora	Dspace	JeromeDL	Bricks
Beveiligings-model	XACML policy (eXtensible Access Control Markup)	Prioritair, ondersteund LDAP en X.509	FOAFRealm (distributed user profiles management)	RBAC (Role-Based Access Control)
Granulariteit	Object, datastreams, methode	Object, datastreams	Resource	Component, methode, object
DRM model	DRM datastreams	Creative Commons Support	Fair Use & DRM (in ontwikkeling)	MPEG-21 REL
DRM tool support				Watermarking

8.1.5.1.5 Semantische en community aspecten

	Fedora	DSpace	JeromeDL	Bricks
Reasoner	Mulgara-based		Aanbeveling robot (Prolog)	Configureerbare inferentie robot
Tagging	Via middleware (NSDL/NDR, PLoS ONE/Topaz)		Vrije tagging (WordNet gebaseerd)	Annotatie
Taxonomieën	Eender welke		Eender welke (JOnto)	Eender welke
Kennisdeling	Via middleware bovenop Fedora	Via OAI API	SSCF component	Via middleware bovenop Bricks
Communities			SIOC & FOAF compliant	

8.2 Commerciële Media Asset Management Systemen

8.2.1 Pictura

8.2.1.1 Introductie

Pictura bestaat uit twee nauw samenwerkende bedrijven: Pictura Imaginis en Pictura Database Publishing. Pictura richt zich op de erfgoedsector (archieven, musea, bibliotheken, etc.) om op een professionele manier hun collecties te helpen ontsluiten. Een digitale collectie, ontsloten door middel van een backend database, is belangrijk om te komen tot een goed beheer en een duurzaam behoud. Daarnaast biedt een digitale collectie een basis om uitgebreidere services te bieden aan verschillende publieksgroepen. Een voorbeeld hiervan is een internetversie van een beeldbank collectie of een webwinkel waarin reproducties uit deze collecties online te bestellen zijn.

8.2.1.2 Projectsamenvatting

Voor het beheren en raadplegen van de collectie heeft Pictura het pakket Memorix ontwikkeld. In het algemeen kan Memorix gezien worden als een webgebaseerd pakket om omvangrijke digitale databestanden te ontsluiten. De database van het systeem bevat de metadata over de objecten en koppelingen naar deze objecten. Het is mogelijk om elk type object te beschrijven. In het geval van video-objecten kan het object bijvoorbeeld tot op scèneniveau beschreven worden.

Memorix is opgebouwd uit diverse Open Source systemen. Hierdoor bekomt men een maximale controle over het geheel en kan Memorix een onafhankelijke koers blijven volgen ten opzichte van de grote softwarebedrijven. Het feit dat de open source software ondersteund wordt door een grote groep experts op het internet, levert daarbovenop de garantie op dat de continuïteit van Memorix verzekerd blijft. Verder kunnen er ook ad-hoc aanpassingen uitgevoerd worden. De basis van het geheel bestaat uit een MySQL Database Management System (DBMS). De webpagina's worden door een Apache webserver naar de browser gestuurd en een PHP-module zorgt ervoor dat deze pagina's met elkaar kunnen communiceren en dat de gegevens verwerkt kunnen worden. Overigens is het ook mogelijk om Memorix met een ander DBMS zoals Microsoft's SQL server of Oracle te besturen. Deze twee niet open source zorgen natuurlijk voor extra licentiekosten. Voor de webserver kan als alternatief Microsoft Internet Information Server (IIS) gebruikt worden.

8.2.1.3 Technologische overwegingen

8.2.1.3.1 Webbrowser

In tegenstelling tot andere collectiebeheersystemen maakt dit systeem geen gebruik van complexe client-side programmatuur, maar van de webbrowser. De gegevens worden op een dynamische wijze op het beeldscherm gepresenteerd. Het voordeel van het gebruik van een webbrowser is de bekende bediening, de algemene beschikbaarheid en de lage eisen die een webbrowser aan een computer stelt.

8.2.1.3.2 Server

De computer met de webbrowser maakt gebruik van de centraal opgeslagen data op de server. De servers zijn uitgerust met het open source besturingssysteem Linux. Door de keuze voor Linux wordt de gebruiker niet geconfronteerd met dure licenties en nieuwe, al dan niet opnieuw te betalen, softwareversies. Omdat een collectie met digitale beelden al snel een enorme diskopslag vereist, die vaak te omvangrijk wordt voor de server van de klant, biedt Pictura de mogelijkheid om de websites en beeldbanken van de klanten op hun servers te hosten. Het systeembeheer van Pictura zorgt hierbij voor het onderhoud en de beveiliging. Technische bagage van de instantie uit is dus niet nodig.

8.2.1.3.3 Uitbreidbaarheid

Extra informatie toevoegen, buiten de basisinformatie, in de database is mogelijk door in de database zelf vrije tekstvelden te voorzien. De frontend client kan ook eenvoudig in meerdere talen weergegeven worden. Hiertoe dient de klant wel zelf alle velden te vertalen en de metadata in meerdere talen in te geven. Indien nodig kan men ook de volledige interface vertalen. Memorix wordt nog steeds verder ontwikkeld waardoor er continu nieuwe toevoegingen gedaan worden. Op dit moment wordt er ondermeer een IPTC-koppeling en een module om bestanden rechtstreeks van op de eigen desktop in Memorix in te laden gerealiseerd.

8.2.1.3.4 Support

Pictura is een dienstverlenend bedrijf en begeleidt een project van bij het begin tot wanneer het project volledig operationeel is. Pictura adviseert bij de opbouw van de databank en helpt bij de conversie en het inlezen van bestaande data. Bugs worden opgevolgd via het Mantis trackingsysteem waartoe de klant toegang heeft. Voor België en Nederland is er een helpdesk beschikbaar die integraal deel uitmaakt van de dienstverlening. Binnenin Memorix wordt standaard ook een forum ingebouwd. Er bestaat ook een Pictura/Memorix-

gebruikersgroep onder leiding van het Fotomuseum Rotterdam, het Nationaal Archief en het Gemeentearchief Amsterdam.

8.2.1.3.5 Gebruik en institutioneel beleid

Memorix laat toe een digitaal archief te beheren en te publiceren. Verder kunnen ook meerdere archieven met elkaar worden gekoppeld. Voor het bijhouden van de producties en de logistiek daaromtrent gebruikt Pictura een extra webgebaseerd secure-systeem (E-Synergy). Medewerkers van het archief kunnen daarin actief meewerken, commentaar geven en digitaal aftekenen voor ontvangst. Een management rapporteringsmodule is verder ook aanwezig. Collectiebeheer van beeld, audiovisuele bronnen, archief en bibliotheek is op deze manier beschikbaar in een integraal doorzoekbaar systeem. Memorix heeft ook een fijnmazig rechtensysteem waarbij zowel op groepsniveau, als op individueel objectniveau rechten bepaald kunnen worden. Dit omvat ondermeer toegangsrechten op collecties en op functionaliteiten van de software. Lees- en schrijfrechten vallen daar ook onder. Verder is het mogelijk om op individueel objectniveau formulieren aan te passen. De ene gebruiker ziet bijvoorbeeld drie velden terwijl een andere gebruiker er tien kan zien en bewerken.

8.2.1.4 Metadata en access

8.2.1.4.1 Metadata

Nadat de collectie gedigitaliseerd is, zal elke foto, dia, akte, enz. beschreven moeten worden zodat ze geschikt zijn voor publicatie op het web. Door toevoeging van metadata wordt de collectie toegankelijk en doorzoekbaar. Pictura kan de klant, indien gewenst, adviseren bij het beschrijven van de collecties. Er wordt gestreefd om zoveel mogelijk aan te sluiten bij standaarden die in het archiefwezen gebruikt worden, waardoor transparante uitwisseling van gegevens tussen archieven onderling mogelijk wordt. Alle gegevens binnen Memorix worden in XML bewaard. Hoewel Memorix geen beperking oplegt qua metadataformaat wordt er standaard wel een Dublin Core koppeling voorzien. Memorix kan via OAI data harvesten of geharvest worden. Velden kunnen ook gemapt worden op Dublin Core velden. Verder is er ook een speciale EAD-editor beschikbaar.

8.2.1.4.2 Toegang via zowel externe als interne publieke websites

Pictura verzorgt de ontwikkeling van functionele websites die toegang voorzien tot de archieven tot op objectniveau. De bezoeker kan in het archief zoeken, bijvoorbeeld op naam, bron en periode, en selecties maken. Met behulp van zoekopdrachten kan de archivaris het archiefonderzoek verbreden door relaties te leggen naar digitale bronnen van andere archieven. Alle websites worden via een Content Management Systeem (CMS) opgebouwd

en gekoppeld aan een Memorix Archief, zodat die websites op eenvoudige wijze beheerd kunnen worden.

8.2.1.5 Sterktes en zwaktes

8.2.1.5.1 Sterktes

Memorix is een krachtig, webgebaseerd systeem dat ontwikkeld werd bovenop een open source basis. Hierdoor is er een grote beschikbare ontwikkelingsondersteuning aanwezig. Pictura hanteert voor zijn software een “één licentie - meerdere gebruikers” systeem (RHC-model). Wanneer een instelling software koopt van Pictura mag het dus een ongelimiteerd aantal gebruikers deze software laten gebruiken. De software bevat een flexibele beheersomgeving en biedt autorisatie tot op formulierniveau aan. De software is uitbreidbaar met modules voor beelden, audiovisueel materiaal, museum- en archiefoverzicht. Er kan ook een eigen webwinkel opgezet worden. Tot slot verleent Pictura een extra aantal diensten, zoals inhoudelijke begeleiding door een documentalist, het beheren van de server en het nemen van back-ups. Hierdoor liggen de vereisten voor technische kennis bij de klant zeer laag.

8.2.1.5.2 Zwaktes

Hoewel het gebruik van commerciële applicaties heel wat voordelen kan bieden, zoals goede ondersteuning en eventuele bijkomende services zoals hosten van website en archieven, kan de vraag gesteld worden of de prijs die hiervoor dient betaald te worden effectief een meerwaarde oplevert. In de eerste plaats kan er nagegaan worden in welke mate de software aan de noden van de klant tegemoet komt en of er geen betere gratis open source alternatieven beschikbaar zijn. De bedenkingen die gemaakt kunnen worden bij deze alternatieven zijn de aanwezigheid van ondersteuning, de beschikbare IT-kennis van de klant zelf en de nodige resources. Enkel door deze voor- en nadelen van beide applicaties tegen elkaar af te wegen, kan er bepaald worden of de prijs voor het commercieel pakket te verantwoorden is.

8.2.1.6 Conclusie

Alle door Pictura ontwikkelde applicaties zijn webgebaseerd. Om de collectie te beheren is dus bijgevolg alleen een computer met een internetverbinding en een wachtwoord nodig. Dit betekent dat de gebruikers compleet onafhankelijk van plaats (kantoor of thuis) en tijd (tijdens de werkuren of in het weekend) kunnen werken. Deze software leent zich dan ook goed om bijvoorbeeld aan de hand van een grote groep vrijwilligers op een korte termijn

metadata in te voeren. De keuze voor open source applicaties als basis voor het systeem en Linux als besturingssysteem voor de server houdt de kostprijs binnen de perken. Bovendien hanteert Pictura een “één licentie, meerdere gebruikers” systeem. Toch dient er door de klant zelf nagegaan te worden of deze te betalen prijs te verantwoorden is.

8.2.2 Adlib

8.2.2.1 **Introductie**

Adlib Information Systems levert software voor archieven, bibliotheken en musea (ook in die volgorde). Zo biedt het een applicatie voor archieven (Adlib Archief), musea (Adlib Museum) en bibliotheken (Adlib Bibliotheek). Deze toepassingen kunnen gemakkelijk geïntegreerd worden met alle andere toepassingen en beschikken daarbovenop over een aantal optionele uitbreidingsmogelijkheden. Bij de ontwikkeling van de software wordt de nadruk vooral gelegd op open systemen en internationale standaarden.

8.2.2.2 **Projectsamenvatting**

Adlib Information Systems levert Adlib-software af sinds 1978. De eerste versies van Adlib, door het Britse LMR ontwikkeld, draaiden op Prime-computers onder het Primos besturingssysteem. Omdat het einde van het Prime-tijdperk werd voorzien, kwam in 1992 de eerste MS DOS-versie van Adlib uit (toen nog Adlib Lite geheten), al snel gevolgd door een versie voor UNIX. Het Windows-platform werd in de jaren 90 steeds populairder waardoor ook een Windows-versie van Adlib niet kon uitblijven. De eerste Adlib voor Windows werd in 1995 uitgebracht wat een grote groei in het aantal Adlib-gebruikers tot gevolg had. Momenteel worden alle nieuwe ontwikkelingen op het .NET-platform gedaan.

8.2.2.2.1 *Adlib Archief*

Adlib Archief is speciaal bedoeld voor het beheren van collecties in historische archieven en rijksarchieven. Adlib Archief biedt uitgebreide functionaliteit en interfaces aan voor professioneel archiefbeheer, gebaseerd op een hiërarchische structuur in overeenstemming met de internationale standaard ISAD(G), en validatiebestanden die voldoen aan de ISAAR(CPF)-norm. Tot de standaard mogelijkheden behoren onder andere de catalogisering van aanwinsten en publicaties, met een extra koppeling naar afbeeldingen. Verder zijn er ruimschoots zoekmogelijkheden, en kan er gebruik gemaakt worden van interfaces met externe bestanden zoals Word en Excel-documenten, of webpagina's. Het hart van Adlib Archief is de catalogus, die archiefbeschrijving op meerdere niveaus volgens de ISAD(G)-norm (2e editie) ondersteunt. De informatie die moet worden ingevoerd in catalogusrecords van een hoger niveau, kan ook automatisch worden ingevuld door overerving van de lagere

niveaus. Desgewenst kan de gebruiker deze overerving echter op ieder niveau onderdrukken. De velden van elk catalogusrecord worden gegroepeerd in overeenstemming met de ISAD(G)-elementen van een beschrijving, en gemakshalve op tabbladen gepresenteerd. De gebruiker kan eenvoudig doorheen de tabs bladeren, zodat het invoeren en wijzigen van gegevens snel en efficiënt verloopt. Koppelingen naar afbeeldingen en andere externe bestandstypen worden ondersteund.

8.2.2.2 Adlib Musea

Adlib Museum is een applicatie die speciaal gemaakt is voor collectie- en informatiebeheer in musea. Het hart van Adlib Museum wordt gevormd door het collectiebestand. Daarin kunnen objecten volledig volgens de Spectrum-standaard worden beschreven. De objectgegevens zijn onderverdeeld in handige groepen die worden gepresenteerd op tabbladen. De gebruiker kan eenvoudig tussen de tabs wisselen, zodat het invoeren en wijzigen van gegevens snel en eenvoudig kan gebeuren. Tabbladen die niet gebruikt worden, kunnen uitgeschakeld worden. Adlib ondersteunt koppelingen naar afbeeldingen en andere externe bestanden zoals Word-documenten, Excel-spreadsheets en PDF-bestanden. Het objectenbestand in Adlib Museum bevat standaard onder andere de volgende rubrieken en mogelijkheden:

- Identificatie;
- Fysieke kenmerken;
- Iconografie;
- Opschriften en merken;
- Associaties;
- Tentoonstellingen;
- Documentatie;
- Verwerving en waarde;
- Standplaats en bruikleen;
- Conditie en conservering;
- Reproducties;
- Nummers en relaties;
- Opmerkingen;

Verder bestaat onder meer de mogelijkheid om:

- Objecten met elkaar te verbinden;
- Meerdere titels per object te vermelden;
- Toelichtingen en bronvermeldingen op te nemen;

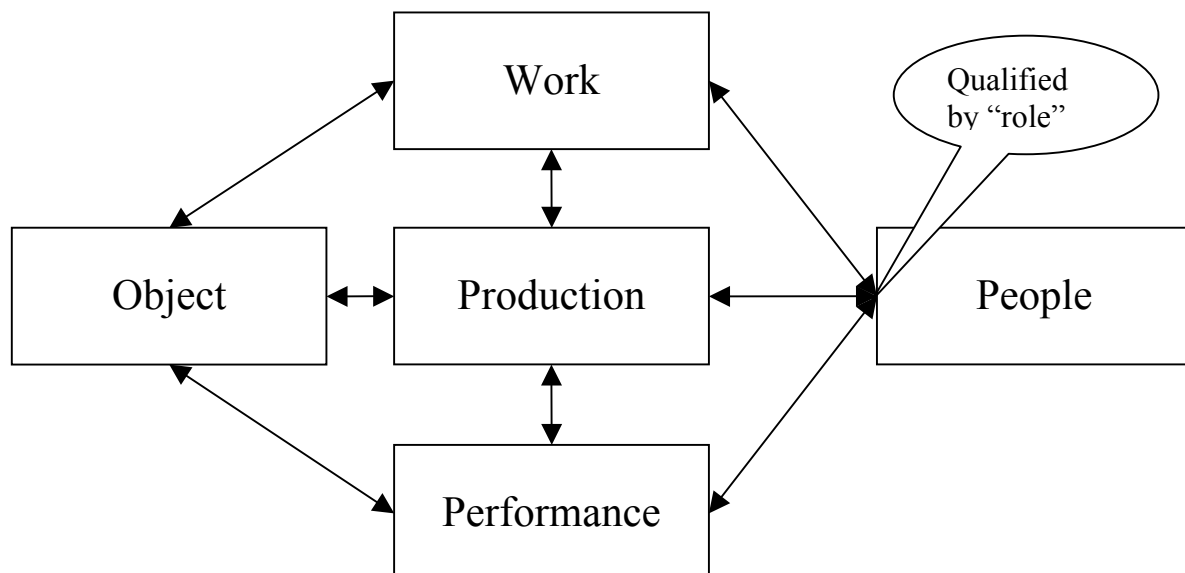
- Voorstellingen met trefwoorden te ontsluiten;
- Associatieve gegevens op te nemen;

8.2.2.3 Technologische overwegingen

8.2.2.3.1 Vereisten

Adlib designer

Een bijzondere eigenschap van ADLIB is dat er een toolkit met het systeem mee wordt geleverd, waarmee de klant zelf applicaties kan bouwen of aanpassen. Deze toolkit, ADLIB Designer, stelt een gebruiker in staat om een ADLIB systeem geheel aan de eigen vereisten aan te passen. Deze flexibiliteit heeft er voor gezorgd dat ADLIB ook in theateromgevingen is toegepast. De gegevensstructuren die in een theater toegepast worden hebben veel gemeen met bijvoorbeeld die uit de museale wereld. Bovendien beschikken theater(archieven) vaak ook over een keur aan realia en documenten (props, kostuums, scripts, bladmuziek). Zo gebruikten het “Theater Instituut Nederland” (TIN, <http://www.tin.nl>) en het Londense “Royal Opera House” (<http://www.royalopera.org>) de ADLIB toolkit om een combinatie te maken van ADLIB Xplus (bibliotheek, museum en archief) en een eigen structuur voor het vastleggen van gegevens over werken, producties en uitvoeringen. Hierbij wordt gebruik gemaakt van het volgende model:



De verschillende record types en hun relaties worden door middel van de professionele interface, een Windows “forms” applicatie, onderhouden.

Adlib Internet Server-Module

De Adlib Internet Server-module draait op een Windows-server. Indien men de nieuwste technieken wil toepassen, kan men het best gebruik maken van Microsoft Internet Information Server (standaard meegeleverd bij Windows), geïnstalleerd op een Windows 2000 of 2003-server. De Adlib-applicaties zijn getest voor MS Internet Explorer en Netscape Navigator. Internet Explorer wordt aangeraden in verband met de betere ondersteuning van de opmaakmodellen. De browsers moeten versie 4.0 of hoger zijn.

Support

Adlib Information Systems bevat een support afdeling die de technische ondersteuning voor Adlib gebruikers verzorgt. De medewerkers op deze afdeling voeren dataconversies uit van andere databases naar Adlib, maken applicatieaanpassingen en programmeren desgewenst aanvullende rapportagemogelijkheden voor uw gegevens in de Adlib-databases. Een belangrijke taak van de afdeling Support is het verzorgen van de helpdesk. Klanten met een onderhoudscontract kunnen hier kosteloos terecht met vragen over het gebruik van Adlib. Naast deze support afdeling biedt Adlib Information Systems ook gebruikersgerichte cursussen aan. Verder bestaan er ook Adlib Gebruikersgroepen. Dit zijn onafhankelijke verenigingen die de belangen van Adlib klanten vertegenwoordigen. Elke Adlib gebruiker kan lid worden van zo'n vereniging. Via hun website willen de Adlib Gebruikersgroepen communicatie van en kennis over Adlib stimuleren en verspreiden. Men kan er onder andere een forum en een prikbord terugvinden [119].

Architectuur

Adlib Information Systems gebruikt ASP en XML-technologieën voor de ontwikkeling van webapplicaties. Omdat deze technieken niet Adlib-specifiek zijn, is de Adlib-kennis die een applicatieontwikkelaar nodig heeft om een Adlib webapplicatie te bouwen beperkt. Om een Adlib Webapplicatie te bouwen heeft men alleen enige kennis nodig van Adlib-bestanden, zoals de bestandsnamen, velden en indexen. Verder moet men dus enkel een ASP scripttaal (bijvoorbeeld JavaScript) beheersen en bekend zijn met XML-technieken.

8.2.2.4 Gebruik en institutioneel beleid

Adlib genereert duidelijke en consequent opgebouwde schermen en bezit contextgevoelige hulp. Daarnaast zorgen de validatielijsten en thesaurus ervoor dat het invoeren van gegevens gemakkelijk en accuraat verloopt. Dit resulteert in betrouwbare en eenduidige bestanden die een solide basis vormen voor succesvol databeheer. De pakketten bevatten een exportassistent waarmee de data onder andere in kommagescheiden formaat en XML kunnen worden geëxporteerd. Andere formaten kunnen worden toegevoegd. De Adlib Internet Server is de schakel tussen de catalogus, de collecties en het internet of intranet. Door deze module aan een Adlib-systeem toe te voegen, kan de database online toegankelijk gemaakt worden. Online bezoekers gebruiken gewoon hun eigen webbrowser en hebben dus geen speciale software nodig.

8.2.2.5 Metadata en access

Metadata

Divers multimediaal materiaal is aan de verschillende recordtypes gekoppeld via een Dublin Core metadata record. De opslag wordt binnen Adlib voorzien door middel van een XML formaat. Als database engines kunnen de Adlib proprietary database, Microsoft SQL server of Oracle gebruikt worden.

Access

In Adlib kunnen toegangsrechten worden toegekend aan individuele gebruikers en groepen van gebruikers. Hierdoor kan de toegang tot gegevens en tot bepaalde functies zoals wijzigen, verwijderen of afdrukken, onder controle worden gehouden. Voor instellingen die hun catalogus via internet of intranet ter beschikking willen stellen, heeft Adlib Information Systems de Adlib Internet Server ontwikkeld. Door middel van een webbrowser kunnen online museumbezoekers gemakkelijk de catalogus doorzoeken. Deze module bevat een volledig zelf te definiëren zoekoptie en weergave van zoekresultaten, en geeft volledige controle over welke data toegankelijk is voor het publiek en welke niet.

8.2.2.6 Sterktes en zwaktes

Zwaktes

Voor opleidingen en support dient betaald te worden.

Sterktes

Adlib Producten leggen zich toe op een bepaalde sector/domein en profiteren zo van de uit deze sector aanwezige ervaring en kennis. Verder is er een zeer goede support onder de vorm van een helpdesk en gebruikersgroepen.

8.2.2.7 Conclusie

Adlib Information Systems, de maker van Adlib de verschillende software producten, heeft in haar lange bestaan al veel bibliotheken, musea en archieven geautomatiseerd. Een groot aantal medewerkers is afkomstig uit de bibliotheek- en de museumwereld en kent daardoor de problemen en wensen van deze sector als geen ander. Door nauw samen te werken met gebruikers en door het toepassen van de laatste technologische ontwikkelingen wordt de Adlib-software voortdurend verbeterd. Het resultaat is gebruiksvriendelijke software en een snel groeiend aantal gebruikers. Adlib Information Systems ondersteunt haar gebruikers met tal van faciliteiten. Voor advies en ondersteuning kan men terecht bij de helpdesk, zowel telefonisch als via het internet. Een breed scala aan opleidingen draagt zorg voor het optimaal gebruik van deze applicaties.

9 Streaming technologieën

9.1 Streaming Servers

9.1.1 Apple Darwin Streaming Server

Darwin Streaming Server is de open-source tegenhanger van de Quicktime Streaming Server [120] [121]. De server werkt zowel onder Windows, Mac, Linux, Solaris als BSD. Naast het Quicktime formaat kan Darwin ook MP4 en 3GPP streamen. Hierbij worden zowel HTTP, RTSP, RTP als IceCast ondersteund. Voor het streamen van geluid ondersteunt Darwin o.a. AAC, Apple Lossless, PCM en MP3. Voor video wordt ondersteuning geboden voor H.263, MPEG-4 en H.264/AVC.

9.1.2 Adobe Macromedia Flash Media Server 2

Macromedia Flash Media Server is een propriëtaire streamingserver van Adobe die werkt onder Windows of Linux [122]. De Flash Media Server ondersteunt enkel het eigen FLV-formaat. Dit formaat ondersteunt VP6 als videocodering en MP3 als audiocodering. Het streamen gebeurt via het eigen RTMP-protocol. Ondersteuning voor H.264 en HE-AAC wordt snel verwacht.

9.1.3 Windows Media Server 9

Windows Media Server is een propriëtaire streamingserver van Microsoft die werkt onder Windows 2003 [123]. De Windows Media Server ondersteunt enkel het streamen van de eigen video- en audioformaten. Hierbij worden UDP, TCP en HTTP ondersteund als protocollen. Windows Media Server kan ASX, ASF, WMA en WMV-bestanden streamen. Deze kunnen Windows Media Video videostreamen bevatten en Windows Media Audio audiostreamen (zowel professional, voice als standaard). Ook PCM-bestanden worden ondersteund.

9.1.4 RealNetworks Helix Server/Helix DNA Server

RealNetworks Helix Server is een propriëtaire streamingserver van RealNetworks die werkt onder Windows, Linux en Solaris. Sinds 2003 kent deze Server ook een open-source tegenhanger met wie RealNetworks Helix Server een deel van de broncode en features deelt namelijk Helix DNA Server [124] [125]. Beiden ondersteunen RAM, 3GPP en MP3-streaming en dit met behulp van de transportprotocollen RTSP, RTP, HTTP, TCP en UDP. De versie van RealNetworks voegt hier verder nog ondersteuning toe voor onder andere Windows Media, Quicktime, MPEG-4, H.263 en MPEG-4 AVC.

9.1.5 Apache

Apache is een open-source webserver ontwikkeld voor zowel Windows, Mac, Linux als BSD en Solaris [126]. Apache ondersteunt enkel het HTTP-protocol en kan via dit protocol alle formaten streamen die dit ondersteunen [127].

9.1.6 Orb

Orb is een gratis, niet open-source, programma, ontwikkeld voor Windows, dat toelaat audio en video te streamen [128]. Dit programma is gericht op privépersonen en heeft tot doel dat deze overal en met elk toestel toegang hebben toch hun eigen verzameling video's en muziek. Orb ondersteunt hierbij alle audio- en videoformaten waarvoor een DirectShow-filter beschikbaar is. Orb zal deze dan converteren naar een ASX, RAM of 3GPP-bestand naargelang welk formaat door de cliënt wordt ondersteund. Streaming gebeurt via het HTTP en UDP.

9.1.7 VLC

VLC, of Video LAN Client, is een open-source mediaspeler die beschikbaar is voor o.a. Windows, Mac, Linux en Pocket PC [129]. Naast het afspelen van video en audio biedt deze mediaspeler ook de mogelijkheid de content te streamen. Dit kan gebeuren via HTTP, MMSH, RTSP en UDP. VLC biedt ondersteuning voor de containerformaten PS, TS, OGG, ASF, MP4, MOV, RAW en WAV. Indien gewenst kan VLC ook voor het streamen de content hercoderen.

9.1.8 Samenvatting

	Server (Client) platformen	Container-formaten	Video-formaten	Audio-formaten	Transport-protocols
Apple Darwin Streaming Server	Windows, Mac, Linux, Solaris, FreeBSD (Windows, Mac)	MOV, MP4, 3GPP	H.264/AVC, MPEG-4, H.263	AAC, Apple Lossless, A-Law, AMR Narrowband, IMA, MACE, Qdesign Music 2, Qualcomm PureVoice, u-Law, PCM, MP3	HTTP, RTSP, RTP, IceCast
Adobe Macromedia Flash Media Server 2	Windows, Linux (Windows, Mac, Linux, Solaris, PocketPC)	FLV	VP6	MP3	RTMP
Microsoft Windows Media Server 9	Windows 2003 (Windows, Mac)	ASX, ASF, WMA, WMV	Windows Media Video	Windows Media Audio 9.1 (Standard, Voice, Professional), Professional, PCM	UDP, TCP, HTTP, MMS
RealNetworks Helix Server 11	Windows, Linux, Solaris (Windows, Mac, Linux, Solaris, Irix)	Quicktime, RAM, WMV, WMA, 3GPP	RealVideo, Windows Media Video, MPEG-4, H.264/AVC, H.263	RealAudio 10 (Stereo, StereoSurround, Multichannel), RealAudio 10 5.1, Windows Media Audio, MP3	RTSP, RTP, HTTP
Apache	Windows, Mac, Linux, BSD, Solaris	ASX, RAM, SDP		Via HTTP: alles	HTTP, HTTPS,
Orb	Windows	ASX, RAM, SDP (3GP)	MPEG-1/2/4, Windows Media Video, H.264/AVC Alles waarvoor DirectShow-filter beschikbaar is	AAC, RA, 3GP, MOV, QT, OGG, FLAC, AMR, M4P, MP3, Windows Media Audio, PCM Alles waarvoor DirectShow-filter beschikbaar is	HTTP, UDP
VLC	Windows, Mac, Linux, Solaris, FreeBSD, BeOS, PocketPC	PS, TS, OGG, ASF, MP4, MOV, WAV, RAW	MPEG-1/2/4, Windows Media Video, H.263, MJPEG, Theora, FLV, H.264/AVC	MP1/2/3, Windows Media Audio, AC3, AAC, HE-AAC, Vorbis, Flac	HTTP, MMSH, RTP, UDP

9.2 Audio searching

Wanneer men beschikt over een grote verzameling geluidsbestanden zal al snel de noodzaak optreden om deze verzameling eenvoudig te kunnen doorzoeken. Het zoeken door een geluidsdatabase verloopt immers heel anders dan bij het “gewone” tekstueel doorzoeken van teksten zoals bij zoekmachines als Google en zoekfuncties in webbrowsers of Office-pakketten. Zo gaat men er bij Google vanuit dat de gebruiker snel visueel door de mogelijke resultaten kan “scannen”, wat bij audiobronnen met hun sequentiële karakter niet kan. De eisen voor de relevantie van de zoekresultaten liggen bij geluidsbronnen dus een stuk hoger door een gebrek aan mogelijkheden om snel door de resultaten te scannen.

Voor audio searching zijn verschillende technieken en methodes mogelijk. De meest eenvoudige werkwijze (met de grootste overeenkomst tot het gewoon tekstueel doorzoeken) is het zoeken op metadata in of over het bestand. Een andere mogelijkheid is het zoeken via een speech-to-text module, waarbij de geluidsbestanden in tekst worden omgezet en vervolgens worden doorzocht. Tot slot zijn er ook technieken die gebruik maken van de karakteristieken van het geluidsbestand zelf, zoals stiltes en nadrukken, om het doorzoeken van geluidsbestanden te vergemakkelijken.

9.2.1 Zoeken op metadata

Het zoeken op metadata is de meest eenvoudige manier van zoeken, zoals ook zoeksites Yahoo [130] en Altavista [131] doen. Ze doen dit o.a. met informatie rondom het fragment die zich op de website van dat geluidsfragment bevindt. De metadata kan bestaan uit informatie binnenin de bestanden, zoals bv. de id3-tag bij mp3-bestanden, maar kan zich ook in een apart bestand bevinden zoals bij Dublin Core, MPEG-7 en dergelijke meer [132].

Het gebruik van metadata kent echter verschillende zwakke punten. Allereerst is de metadata niet steeds even correct, consistent en compleet. Het invullen van metadata is immers een tijdrovende en foutgevoelige bezigheid. Bovendien is er naast objectieve metadata, zoals titel en auteur, ook subjectieve metadata. Deze informatie hangt dus volledig af van de persoon die de gegevens heeft ingevuld. Voor deze manier van zoeken is men dus afhankelijk van derden en de uitgebreidheid en correctheid waarmee de gegevens zijn ingevuld.

Om zo goed mogelijke resultaten te verkrijgen bij het zoeken, wordt het gebruik van bv. een thesaurus bij het doorzoeken van een verzameling aangeraden. Waar de ene persoon

“veteraan” invult, kan de andere “oudstrijder” invullen waardoor slechts een deel van de beschikbare informatie zou weerhouden worden bij het zoeken. Ook het gebruik van “slimme” zoektechnieken, die schrijffouten en dergelijke “negeren”, kunnen een grote hulp bieden.

Tot slot moet opgemerkt worden dat de meeste metadata slaan op het gehele geluidsfragment. Dit zorgt ervoor dat er vaak nog binnen het geluidsfragment moet gezocht worden naar de relevante informatie en dat de metadata aldus niet alle informatie van binnen het geluidsfragment bevat. Dit ook kan dan lijden tot vals negatieve resultaten.

9.2.2 Zoeken via Speech-to-Text

Een andere zoektechniek bestaat erin transcripties van de geluidsfragmenten te maken. Dit kan zowel handmatig gebeuren als automatisch via spraaktechnologie.

Deze techniek kent verschillende voordelen tegenover het zoeken via de metadata. Zo kan met deze techniek alles automatisch gebeuren, kan de gehele tekst doorzocht worden en is het mogelijk de relevante stukken uit een geluidsfragment op een eenvoudige manier terug te vinden.

Transcripties kunnen op verschillende manieren gebruikt worden. Allereerst bieden ze de gebruiker, indien accuraat genoeg, de mogelijkheid informatie te vinden zonder de geluidsbron te moeten beluisteren. De transcripties zijn ook op een eenvoudige manier automatisch te doorzoeken (net als elke andere tekst). Verder bieden ze een overzicht en de mogelijkheid een snelle blik te werpen op de tekst om zo de relevante stukken terug te vinden. Ook biedt de transcriptie de kans fouten terug te vinden, die bij het omzetten naar tekst kunnen zijn ontstaan. Dit kan de accuraatheid van de zoekresultaten verhogen. De gebruiker kan immers de fouten die tot een vals positieve of vals negatieve hit leiden negeren of alsnog gebruiken.

Het mag duidelijk zijn dat het cruciale onderdeel bij transcriptie de spraakherkenning is. Zelfs in de meest ideale omstandigheden biedt spraakherkenningssoftware geen 100% accuraatheid. Het corrigeren van deze teksten vergt nog steeds veel manueel werk. Een nog groter probleem kan zich stellen wanneer de originele geluidsfragmenten zich in slechte staat bevinden, denken we maar aan opnames op oude langspeelplaten of cassettebanden. Dan is er een voorbewerking van het originele fragment nodig en zal de accuraatheid van de transcriptie nog steeds laag zijn. Dit geldt natuurlijk ook voor geluidsfragmenten waarin een dialect wordt gesproken waarmee de spraakherkenningssoftware niet omkan.

Tot slot geven de gebruikers van de speech-to-text aanpak aan dat deze nog een belangrijk nadeel kent. Bij het overzetten van geluid naar tekst gaat namelijk elke vorm van intonatie verloren. Dit kan tot verkeerde interpretaties leiden en maakt duidelijk dat geluid niet zomaar door tekst mag worden vervangen.

9.2.3 Doorzoeken door karakteristieken (pauzes,...)

Een derde techniek maakt gebruik van de karakteristieken van het geluidsfragment. Deze techniek is voornamelijk een ondersteunende techniek die kan helpen bij het doorbladeren of beluisteren van interviews.

Bij deze techniek probeert men zoveel mogelijk informatie te halen uit het akoestische karakter van het geluidsfragment. Men zal met deze techniek ook proberen het geluidsfragment verder onder te verdelen in semantisch betekenisvolle onderdelen. Daarbij zal men pauzes, nadrukken, wijzigingen in volume of kracht van het geluid proberen detecteren. Al deze gegevens kunnen aangeven waar een onderdeel van het geluidsfragment stopt (pauzes) of waar zich belangrijke informatie bevindt (nadrukken, stijgend stemvolume, etc.).

In [133] wordt een voorbeeld beschreven van een systeem dat deze laatste twee technieken samenvoegt. De geluidsbron wordt eerst opgesplitst in “paragrafen” op basis van de akoestische kenmerken. Vervolgens wordt van elk van deze paragrafen een transcriptie gemaakt. In het zoekprogramma wordt dan voor elk segment weergegeven hoe relevant het is voor de gegeven zoektermen. Bovendien wordt dit ook apart aangegeven voor elke ingegeven zoekterm.

Daarnaast is er ook SpeechSkimmer [134]. Dit systeem wordt gebruikt voor het snel “doorbladeren” van audiobronnen. Daarvoor gebruikt SpeechSkimmer de informatie over bv. pauzes en nadrukken om van de verschillende onderdelen een mate van informatie te bepalen. Dit laat toe om een geluidsfragment intelligent in tijd te comprimeren. Zo kunnen pauzes en haperingen worden overgeslagen of bij hogere snelheden hele onderdelen die minder belangrijk zijn. Naast deze intelligentere technieken laat SpeechSkimmer ook toe om bv. het geluidsfragment sneller af te spelen.

10 Beheermodellen

Intellectuele eigendomsrechten zorgen ervoor dat individuen en organisaties kunnen profiteren van hun creatieve en intellectuele inspanningen. Meer specifiek zorgen ze ervoor dat anderen er voordeel uithalen op kosten van de rechtmatige eigenaar. Wat vroeger een vrij ongecompliceerd rechtsgebied was, is nu ineens heel complex geworden. Technologische omwentelingen hebben er immers voor gezorgd dat het nu heel gemakkelijk en goedkoop geworden is om digitale kopieën te maken. Om aan deze veranderende condities te voldoen, is de wet op intellectuele eigendom aangepast waardoor het beheer van deze eigendomsrechten ineens ook complex is geworden.

Het is belangrijk te weten en te identificeren welke wettelijke beschermende rechten er aanwezig zijn vooraleer men die bronnen hoe dan ook begint te digitaliseren. Het meest relevante recht om in ogenschouw te nemen wanneer men culturele bronnen wenst te digitaliseren, is auteursrecht (id est: de bescherming van expressie van ideeën). Alle soorten bronnen (van muzikale, over artistieke, tot literaire) worden beschermd door het auteursrecht als ze origineel zijn. Verder zijn er natuurlijk nog een resem andere rechten op het reproduceren van deze “echte” werken zoals daar zijn: morele rechten, privacy recht, vrijheid van expressie, dataprotectie, onfatsoenlijkheid en persoonlijkheidsrecht. Los van al deze rechten moet ook de legaliteit van de instelling in kaart gebracht worden. Deze moet immers voldoende credentials hebben om in deze materie te kunnen optreden als broneigenaar of als brongebruiker.

Een eerste stap in het beheren van de rechten van (digitale) culturele bronnen is identificeren welke rechten er al bestaan in alle collecties binnen het museum, het archief of de bibliotheek. Alle culturele instellingen zouden een audit moeten uitvoeren op al hun activa om een inzicht te krijgen wie welke intellectuele eigendomsrechten in handen heeft. Een museum zou bijvoorbeeld een afbeelding van een schilderij uit zijn collectie kunnen tonen op zijn website. De rechten van het schilderij worden gehouden door het museum, maar waarschijnlijk heeft de fotograaf bijkomende rechten op de foto “an sich” van dat schilderij. Nog andere rechten kunnen zich bevinden bij de ontwerper en beheerder van de website.

Het museum moet, waar nodig, onderhandelen met de rechtenhouders om toestemming te verkrijgen de mondelinge historische bronnen te mogen gebruiken. Deze toestemmingen moeten vertaald worden in licenties. Deze licenties zouden de aard en de scope van de

inhoud van de bronnen moeten weerspiegelen, alsook de manieren waarop deze gebruikt kunnen worden, de duur van die licentie en waar nodig ook de prijs van de licentie. Dit zou voor elke bron gedaan moeten worden die het museum, het archief of de bibliotheek wenst te exploiteren.

Het volstaat niet om eenvoudigweg te onderhandelen over de licenties om het bronmateriaal te mogen gebruiken. De rechten moeten ook worden beheerd. De rechthouders zouden zowel de rechten als de licenties verbonden aan een bron moeten beschrijven. Deze informatie zou steeds samen met de andere records over deze bron bewaard moeten blijven. Slechts weinig intellectuele eigendomsrechten licenties blijven immers altijd geldig. Daarom is het uiterst belangrijk om bij te houden wanneer de rechten verstrijken en eventueel nieuwe onderhandelingen op te starten voor het herverkrijgen van de nodige licenties.

Musea en andere culturele erfgoedinstellingen moeten hun intellectuele eigendomsrechten beschermen. Ze moeten op een ondubbelzinnige, veilige manier licenties onderhandelen met individuen en organisaties die op één of andere manier gebruik wensen te maken van hun bronnen. Stappen moeten immers genomen worden om ervoor te zorgen dat onrechtmatig gebruik niet mogelijk is. Vaak volstaat het gewoon om waarschuwingen of gebruikersverklaringen te tonen. Voor de waardevolle digitale bronnen kan men beter technische middelen inzetten om deze te beschermen tegen ongeoorloofd gebruik. Deze kunnen gaan van het ter beschikking stellen van lage resolutiebeelden, het digitaal versleutelen van de digitale bron, digitaal watermerken, het beveiligen met paswoorden of enkelvoudige webcasting.

Wanneer men beslist om -hetzij zelf, door derden door middel van een gegeven licentie of gewoonweg door het verwerven van gedigitaliseerde content van andere instellingen-bronnen te digitaliseren, zijn er dus, zoals hierboven vermeld, heel wat acties uit te voeren zowel bij het verwerven van de juiste permissies, als bij het beheren van de bronnen en natuurlijk ook bij het uiteindelijk beschermen van die bronnen.

11 Literatuurlijst

- [1] : van Hooland, S. (2007). Publieke ontsluiting van erfgoed: (on)zin van folksonomies voor de culturele sector. Studienamiddag: Participatie in e-cultuur, 27 maart 2007, Brussel.
- [2] : Aroyo, L. (2007). About the Semantic Web, Folksonomies & Users in Digital Cultural Heritage. Bridging to the end-user: cultural heritage information personalization. Studienamiddag: Participatie in e-cultuur, 27 maart 2007, Brussel.
- [3] : Bauwens, J. & Nulens, G. (2005). ICT in het culturele veld: de virtualiteit van het virtuele, Steunpunt Re-creatief Vlaanderen.
- [4] : Mechant, P. & Michiels, K. (2007). Het virtuele kunstencentrum van de toekomst: zoektocht naar een innovatief webplatform voor virtuele cultuurbeleving. Lulu.com.
- [5] : Vande Winkel, R. (2004). Mondelinge, historische bronnen in Vlaanderen. Een kritische bevraging naar de huidige expertise (kennis, evaring en realisaties) inzake de productie, bewaring en ontsluiting van mondelinge, historische bronnen in Vlaanderen als vorm van oraal erfgoed. Eindverslag onderzoeksproject "Van Horen Zeggen" (I) Mondelinge, historische bronnen bewaren en ontsluiten.
- [6] : Vande Winkel, R. (2005). Mondelinge historische bronnen in Vlaanderen. Een stand van zaken en enkele beleidsaanbevelingen. Mores 6(1), 9-14.
- [7] : Prevenier, W.; Howell, M. & Boone, M. (2000). Uit goede bron: introductie tot de historische kritiek. Leuven: Garant.
- [8] : CATCH (Continuous Access To Cultural Heritage), <http://www.nwo.nl/catch>
- [9] : CHOICE (CHArting the informatiOn landscape employing ContExt information), <http://ems01.mpi.nl/CHOICE/menu.html>
- [10] : MuSeUM (Multiple-collection Searching Using Metadata), <http://staff.science.uva.nl/~kamps/museum/>
- [11] : STITCH (Semantic Interoperability to Access Cultural Heritage), <http://www.cs.vu.nl/STITCH/>
- [12] : CHoral (Access to oral history) : <http://hmi.ewi.utwente.nl/choral/>
- [13] : MITCH (Mining for Information in Texts from the Cultural Heritage), <http://ilk.uvt.nl/mitch/>
- [14] : MunCH (Multimedia aNalysis for Cultural Heritage), <http://ilps.science.uva.nl/munch/index.html>
- [15] : RICH (Reading Images in the Cultural Heritage), <http://www.rich.unimaas.nl>
- [16] : SCRATCH (SCRipt Analysis for the Cultural Heritage) : <http://www.ai.rug.nl/alice/nwo-catch-scratch/>
- [17] : CHIP (Cultural Heritage Information Personalization), <http://www.chip-project.org/>
- [18] : WITCHCRAFT (What is Topical in Cultural Heritage: Content-based Retrieval Among Folksong Tunes), <http://www.cs.uu.nl/research/projects/witchcraft/>
- [19] Simple Knowledge Organization Systems maintained by W3C's Semantic Web Deployment Working Group (SWDWG), <http://www.w3.org/2004/02/skos/>
- [20] : GATE-framework, Sheffield university, <http://gate.ac.uk/>
- [21] : Apolda plugin, <http://apolda.sourceforge.net>
- [22] : Lucene plugin, <http://lucene.apache.org/>
- [23] : Weka plugin, <http://www.cs.waikato.ac.nz/~ml/index.html>

- [24] : TreeTagger plugin, <http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/>
- [25] : Sesame triple store, <http://www.openrdf.org/>
- [26] : Findlay, P. (2006). End users, metadata and copyright – who mentioned digitisation? The Archival Sound Recordings Project. Presentation at Learning on Screen Conference, September 2006
- [27] : British Library's objectieven voor digitalisering, <http://www.bl.uk/about/policies/digital.html>
- [28] : British Library's DOM (Digital Object Management Programme),
<http://www.bl.uk/about/policies/dom/homepage.htm>
- [29] : British Library's ASR 1 & 2 (Archival Sound Recordings),
http://www.jisc.ac.uk/whatwedo/programmes/programme_digitisation/project_bl_sound_archive.aspx
- [30] : British Library's Archival Sound Recordings Service, <http://sounds.bl.uk>
- [31] : National Library of Australia's PANDORA project, <http://pandora.nla.gov.au>
- [32] : National Library of Australia's Digital Collections, <http://www.nla.gov.au/digicoll/index.html>
- [33] : National Library of Australia's PADI project, <http://www.nla.gov.au/padi/>
- [34] : Opbouw WAV PCM soundfile, <http://www.sonicspot.com/guide/wavefiles.html>
- [35] : Korte samenvatting Wave64, <http://reference.wolfram.com/mathematica/ref/format/Wave64.html>
- [36] : EBU RF64 specificatie, http://www.ebu.ch/CMSimages/en/tec_doc_t3306-2007_tcm6-42570.pdf?display=EN
- [37] : Opbouw AIFF soundfile, <http://www.borg.com/~jglatt/tech/aiff.htm>
- [38] : XMF Version 1.0 specification, <http://www.midi.org/about-midi/specinfo.shtml>
- [39] : XMF Version 1.01 specification update, [http://www.midi.org/about-midi/xmf/rp39spec\(xmf\).pdf](http://www.midi.org/about-midi/xmf/rp39spec(xmf).pdf)
- [40] : XMF Version 2.0 specification update, [http://www.midi.org/about-midi/xmf/rp43spec\(xmf2\).pdf](http://www.midi.org/about-midi/xmf/rp43spec(xmf2).pdf)
- [41] : OGG containerformaat, <http://www.xiph.org/ogg/>
- [42] : Speex codec, <http://www.speex.org/>
- [43] : Theora codec, <http://www.theora.org/>
- [44] : Vorbis codec, <http://www.vorbis.com/>
- [45] : Matroska containerformaat, <http://www.matroska.org/>
- [46] : EBML, <http://ebml.sourceforge.net/>
- [47] : Voorbeeld van audio tags, <http://www.matroska.org/technical/specs/tagging/example-audio.html>
- [48] : MXF (Material eXchange Format), <http://www.smpte-mxf.org/>
- [49] : MP4 Specificatie, <http://www.digitalpreservation.gov/formats/fdd/fdd000155.shtml>
- [50] : 3GPP Specificatie, <http://www.3gpp.org/ftp/Specs/html-info/26244.htm>
- [51] : ASF Specificatie, <http://www.microsoft.com/windows/windowsmedia/forpros/format/asfspec.aspx>
- [52] : MOV Specificatie,
<http://developer.apple.com/documentation/QuickTime/RM/Fundamentals/QTOverview/index.html>
- [53] : AVI informatie, <http://www.jmcgowan.com/avi.html>
- [54] : FLV (Flash Video), <http://www.digitalpreservation.gov/formats/fdd/fdd000131.shtml>
- [55] : FLV (Flash Video) officiële specificatie, <http://www.adobe.com/licensing/developer/>
- [56] : MARC 21 Field List, <http://www.loc.gov/marc/bibliographic/ecbdlist.html>
- [57] : Understanding MARC, <http://www.loc.gov/marc/umb/>
- [58] : MARCXML, <http://www.loc.gov/standards/marcxml/>

- [59] : Dublin Core, <http://dublincore.org/>
- [60] : Dublin Core Qualified Metadata voor het nieuwe millenium,
<http://igitur-archive.library.uu.nl/DARLIN/2005-0526-200728/PeereboomIP032000.pdf>
- [61] : Metadata en Dublin Core, <http://www.ascleiden.nl/Pdf/Darc/handoutmetadata.pdf>
- [62] : Dublin Core bij het Deventer IJselfront-project, <http://igitur-archive.library.uu.nl/DARLIN/2006-0830-200254/ZandhuisIP032002.pdf>
- [63] : Dublin Core in samenwerkingsprojecten en publieksgerichte ontsluiting,
<http://www.den.nl/docs/20050816173630>
- [64] : Encoded Archival Description, <http://www.loc.gov/ead/index.html>
- [65] : EAD Help Pages, <http://www.archivists.org/saagroups/ead/>
- [66] : METS specificatie, <http://www.loc.gov/standards/mets/>
- [67] : METS Documentatie,
<http://www.loc.gov/standards/mets/METS%20Documentation%20draft%20070310p.pdf>
- [68] : LOM Specificatie, http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
- [69] : LOM-RDF binding,
<http://kmr.nada.kth.se/papers/SemanticWeb/LOMRDFBinding-ARIADNE.pdf>
- [70] : LOM-XML binding,
<http://www.cs.kuleuven.ac.be/~erikd/LOM/20030115/MILCA-LOM-LTSC.zip>
- [71] : LOM gerelateerd software aanbod, <http://www.cancore.ca/editors.html>
- [72] : BibTeX specificatie, <http://www.ecst.csuchico.edu/~jacobsd/bib/formats/bibtex.html>
- [73] : MPEG-7 Overview, <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
- [74] : MPEG-7 Voorbeeld, http://www.lim.dico.unimi.it/didatt/materiali/MPEG7_Tutorial.pdf
- [75] : SPECTRUM, <http://www.mda.org.uk/spectrum.htm>
- [76] : ISAD(G), <http://www.vvbad.be/node/1159>
- [77] : SEPIADES, <http://www.knaw.nl/ECPA/sepia/home.html>
- [78] : ABC Model, <http://metadata.net/harmony/ABCV2.htm>
- [79] : Harmony Project, <http://metadata.net/harmony/Results.htm>
- [80] : MarcOnt, <http://www.marcont.org>
- [81] : MarcOnt Initiative – Bibliographic description and related tools,
<http://library.deri.ie/servlet/showPDF?docId=http%3a%2f%2flibrary.deri.ie%2fresource%2f19fbf1ff&chapter=1&view=pdf>
- [82] : MPEG-21 Overview,
<http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>
- [83] : Cornell University & University of Virginia (2003-2007). Fedora project, <http://www.fedora.info>
- [84] : Cornell University & University of Virginia (2003-2007). Fedora download,
<http://www.fedora.info/download/>
- [85] : Cornell University & University of Virginia (2003-2007). Fedora community,
<http://www.fedora.info/community/>
- [86] : Cornell University & University of Virginia (2003-2007). Fedora future,
<http://www.fedora.info/community/fedorafuture.shtml>

- [87] : URL (Uniform Resource Locator), http://en.wikipedia.org/wiki/Uniform_Resource_Locator
- [88] : MIME (Multipurpose Internet Mail Extensions), <http://www.iana.org/assignments/media-types/>
- [89] : REST (Representational State Transfer), <http://en.wikipedia.org/wiki/REST>
- [90] : SOAP (Simple Object Access Protocol), <http://xml.coverpages.org/soap.html>
- [91] : W3C Resource Description Framework. <http://www.w3.org/RDF/>
- [92] : Cornell University & University of Virginia (2003-2007). Fedora service framework, http://www.fedora.info/wiki/index.php/Fedora_Core_Repository_Service
- [93] : Cornell University & University of Virginia (2003-2007). Fedora preservation working group, http://www.fedora.info/wiki/index.php/Working_Group:_Preservation
- [94] : Cornell University & University of Virginia (2003-2007). Fedora workflow working group, http://www.fedora.info/wiki/index.php/Working_Group:_Workflow
- [95] : Open Archives Initiative, <http://www.openarchives.org/>
- [96] : Protocol for Metadata Harvesting, <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [97] : ISO Open Archival Information System, <http://nost.gsfc.nasa.gov/isoas>
- [98] : Metadata Encoding & Transmission Standard (METS), <http://www.loc.gov/standards/mets/>
- [99] : MPEG-21 Part 2 Overview (DIDL), <http://xml.coverpages.org/mpeg21-didl.html>
- [100] : Cornell University & University of Virginia (2003-2007). Fedora tools en services, <http://www.fedora.info/tools/index.shtml>
- [101] : MIT Labs & HP (2002-2007). Introducing DSpace, <http://dspace.org/introduction/index.html>
- [102] : MacKenzie Smith (oktoker 2003). Interview: A journey into DSpace, <http://www.biomedcentral.com/openaccess/archive/?page=features&issue=7>
- [103] : MIT Labs & HP (2002-2007). DSpace distributie, <http://sourceforge.net/projects/dspace>
- [104] : MIT Labs & HP (2002-2007). DSpace gebruikers, <http://wiki.dspace.org/index.php/DspaceInstances>
- [105] : MIT Labs & HP (2002-2007). DSpace System Manager, <http://www.dspace.org/implement/sys-man.html>
- [106] : MIT Labs & HP (2002-2007), DSpace Wiki, <http://wiki.dspace.org>
- [107] : MIT Labs & HP (2002-2007). DSpace Mailinglijsten, <http://www.dspace.org/feedback/mailling.html>
- [108] : Horsman, P. & Pompe, K. (december 2005). Building a Digital Archive: A Dutch Experience, http://www.rlg.org/en/page.php?Page_ID=20865#article2
- [109] : Dublin Core Metadata Initiative (1995-2007), <http://www.dublincore.org>
- [110] : Dublin Core Library Application Profile (september 2004), <http://www.dublincore.org/documents/library-application-profile/>
- [111] : JeromeDL, <http://sourceforge.net/projects/jeromedl>
- [112] : JeromeDL Wiki, <http://wiki.jeromedl.org>
- [113] : JeromeDL Gids, <http://www.jeromedl.org/guide/index.html>
- [114] : JeromeDL Bug Tracker, <http://jira.jeromedl.org/jira/secure/Dashboard.jspa>
- [115] : BRICKS Website, <http://www.brickscmmunity.org/>
- [116] : LGPL licentie, <http://dev.brickscmmunity.org/LGPL>
- [117] : Bricks Bug Report, http://dev.brickscmmunity.org/Report_Bugs
- [118] : Bricks Mailinglist, http://dev.brickscmmunity.org/Subscribe_to_the_BRICKS_Mailing_Lists

- [119] : Adlib user groups, <http://www.adlibug.nl/wb/>
- [120] : Darwin Streaming Server, <http://dss.macosforge.org/>
- [121] : Darwin Streaming Server Developer Connection ,
<http://developer.apple.com/opensource/server/streaming/>
- [122] : Flash Media Server, <http://www.adobe.com/products/flashmediaserver/>
- [123] : Windows Media server, <http://www.microsoft.com/windows/windowsmedia/forpros/server/server.aspx>
- [124] : Helix Server, <https://helixcommunity.org/>
- [125] : RealNetworks Helix Server – Helix DNA Server comparison,
https://helix-server.helixcommunity.org/2006/devdocs/helix_server_comparison
- [126] : Apache Server, <http://httpd.apache.org/>
- [127] : Audio and Apache HTTPD,
http://sander.vanzoest.com/talks/2002/audio_and_apache/
- [128] : ORB Server, <http://www.orb.com/>
- [129] : Video LAN Client, <http://www.videolan.org/>
- [130] : Yahoo audio zoekrobot, <http://audio.search.yahoo.com/audio>
- [131] : Altavista audio zoekrobot, <http://www.altavista.com/audio/default>
- [132] : ID3, audio file data tagging format, <http://www.id3.org/>
- [133] : Steve Wittaker, John Choi et al. (1998), What you see is (almost) what you hear: design principles for user interfaces for accessing speech archives, Fifth International Conference on Spoken Language Processing, Sydney
- [134] : Barry Arons (1997), SpeechSkimmer: a system for interactively skimming recorded speech, ACM Transactions on Computer-Human Interaction (TOCHI), Volume 4 , Issue 1 (March 1997), Special issue on speech as data, Pages: 3 – 38

12 Bijlagen

12.1 Bijlage 1: schriftelijke vragenlijst

Deze vragenlijst beoogt het in kaart brengen van de huidige situatie en behoeften inzake beheer en ontsluiting van mondeling historische bronnen in Vlaanderen. Met mondeling historische bronnen bedoelen we (historische) opnames van interviews, getuigenissen en verhalen.

Het is belangrijk voor ons onderzoek dat u bij het invullen van deze vragenlijst niet uit eigen naam spreekt, maar uit die van uw organisatie. Teneinde een totaalbeeld te verkrijgen, mag u uiteraard andere medewerkers betrekken bij het invullen van deze vragenlijst.

Op het onderste deel van deze inleidende pagina mag u uw coördinaten nalaten zodat wij u gemakkelijk kunnen contacteren indien nodig. De vragenlijst start op de volgende pagina.

Alvast bedankt voor uw samenwerking en tot binnenkort!

Indien u nog vragen heeft kan u ons contacteren via:

Tel: 09 / 264 91 83

e-mail: Laurence.hauttekeete@UGent.be

Coördinaten van de geïnterviewde(n)

Naam van de organisatie:

Naam/namen van de respondent(en):.....

Functie:.....

Telefoonnummer:.....

Email:.....

A. Aard en omvang van de collectie

Dit onderdeel peilt naar de huidige omvang van de mondelinge bronnen van uw organisatie. Het gaat over alle materiaal dat uw organisatie bezit, zowel analogoos als digitaal. Bovendien wensen we te achterhalen op welke wijze deze bronnen vandaag worden vastgelegd en bewaard.

1. Wij zouden graag enig inzicht verwerven in de aard, omvang en staat van de audiovisuele collectie van uw organisatie.

- a. Kan u voor elk van onderstaand types opname aanduiden hoeveel mondelinge bronnen in de collectie bij benadering gestockeerd¹ en ontsloten² zijn?

Kan u voor elk van deze bronnen (in de kolom ernaast) ook aankruisen in welke staat deze materialen zich hoofdzakelijk bevinden?

Type	Aantal gestockeerde interviews	Aantal ontsloten interviews	Zeer slecht	Slecht	Noch goed, noch slecht	Goed	Zeer goed
Audiomondelinge bronnen mondelinge bronnen					
Video mondelinge bronnen mondelinge bronnen					

- b. Hoeveel mondelinge bronnen worden jaarlijks bij benadering aan de audiovisuele collectie toegevoegd?

..... mondelinge bronnen

1. In welke mate is uw organisatie tevreden over de manier waarop de audiovisuele collectie vandaag gestockeerd wordt?

Duid op onderstaande schaal de tevredenheid van uw organisatie aan tussen 0 (= zeer ontevreden) en 10 (= zeer tevreden)

Zeer ontevreden ⇒ zeer tevreden										
0	1	2	3	4	5	6	7	8	9	10

¹ Met stockeren bedoelen we het louter bewaren van opnames

² Met ontsluiten bedoelen we het gestructureerd beschrijven via een classificatiesysteem zodat bronnen snel en efficiënt teruggevonden kunnen worden door internen en/of externen

2. Zijn er plannen om in de nabije toekomst wijzigingen door te voeren inzake de manier van stockeren van de audiovisuele collectie?

Kruis uw keuze aan en omschrijf kort

- Neen
 Ja

Indien ja, welke?:.....

3. Op welk type niet-file-gebaseerde drager zijn de originele mondelinge historische bronnen gestockeerd? (Met niet-file-gebaseerde dragers bedoelen we dragers waarbij de opgeslagen mondelinge bronnen niet als bestand worden opgeslagen)

Kruis per type opname uw keuze aan. Meerdere antwoorden mogelijk

AUDIO

- Wasrol
 Fonogram
 Vinyl
 Magnetische draad
 Magnetische band
 Compactcassette
 DCC (Digital Compact Cassette)
 DAT (Digital Audio Tape)
 Andere (specificeer):

- Onze organisatie bewaart geen audio-opnames

VIDEO

- Pellicule
 Videoband (Betamax, Video2000, VHS)
 1"
 Betacam (SP, Digital, SX, IMX, HDCAM)
 DV
 MiniDV
 Digital8
 Andere (specificeer):

- Onze organisatie bewaart geen video-opnames

4. Heeft uw organisatie een kopie (als back-up) van de originele mondelinge bron?

Kruis uw keuze aan

- Ja
 Neen (indien neen, ga naar vraag 6)

5. **Van hoeveel materiaal, in % van de totale hoeveelheid aanwezig niet file-gebaseerd materiaal, heeft uw organisatie een kopie als back-up?**

Vul het percentage in

..... %

6. **Kan u aanduiden aan hoeveel mondelinge bronnen de volgende 3 types van randdocumentatie gelinkt zijn?**

Vul het percentage in

..... % van alle bronnen heeft een bandinhoudfiche erbij

..... % van alle bronnen heeft een identificatiefiche erbij

..... % van alle bronnen heeft een contract erbij

7. **Hoeveel procent van de mondelinge bronnen is bij benadering getranscribeerd?**

Vul het percentage in

..... % heeft een volledige transcriptie

..... % heeft een gedeeltelijke transcriptie

(indien u tweemaal 0% heeft geantwoord, ga rechtstreeks naar **vraag 11**)

8. **Op welke drager(s) worden deze transcripties bewaard?**

Omschrijf kort

.....

9. **Hoeveel van de transcripties bezit u ook in digitale vorm? (vb: MS word, pdf, gescand, ...)**

Vul het percentage in

..... % van alle transcripties

10. **Zijn er mondelinge bronnen in uw organisatie die enkel als transcriptie beschikbaar zijn en waar dus geen audio- of video-opname bijhoort?**

Kruis uw keuze aan

- Ja
 Neen

11. In welke mate is uw organisatie tevreden over de manier waarop randdocumentatie (transcriptie, bandinhoudfiche, identificatiefiche, contract) van de mondelinge bronnen vandaag beheerd wordt?

Duid op onderstaande schaal de tevredenheid van uw organisatie aan tussen 0 (= zeer ontevreden) en 10 (= zeer tevreden)

Zeer ontevreden ⇒ zeer tevreden										
0	1	2	3	4	5	6	7	8	9	10

12. Zijn er plannen om in de nabije toekomst wijzigingen door te voeren inzake de manier van beheer van randdocumentatie?

Kruis uw keuze aan en omschrijf kort

Neen

Ja

Indien ja, welke?:

B. Digitale collectie

Dit onderdeel wenst te achterhalen in welke mate de huidige collectie mondeling historische bronnen digitaal beschikbaar is. Met de term 'digitaal' bedoelen we het proces waarbij bronnen loskomen van de drager en dus **louter file-based** worden.

Een digitaal bestand kan men derhalve op elke digitale speler consulteren. Ook kan men het bestand zonder enig kwaliteitsverlies wijd en snel kopiëren en distribueren.

In deze vragenlijst zullen we een onderscheid maken tussen de zogenaamde 'born digital' bronnen, interviews die digitaal geregistreerd worden, en bronnen die naar een digitale file-based vorm worden geconverteerd (gedigitaliseerd).

13. Zijn er mondelinge bronnen die door uw organisatie digitaal worden geregistreerd of die u digitaal van derden ontvangt of verkrijgt (born digital)?

Kruis uw keuze aan

Neen (ga rechtstreeks naar **vraag 20**)

Ja (vervolg hieronder de vragenlijst)

14. Hoeveel van volgende types opname betreft naar uw schatting born digital?

Vul het aantal in

Audio mondelinge bronnen

Video mondelinge bronnen

15. Op welke dragers wordt dit digitaal materiaal opgeslagen?

Kruis voor elk type opname die van toepassing is uw keuze aan. Meerdere antwoorden per type mogelijk

	Audio	Video
Diskette (8, 5 1/4, 3 1/2)		
CD-r(w)		
ZipDisk		
MiniDisk / Hi-MD		
LaserDisk		
DVD-r(w)		
Digital tapes (LTO-x)		
Hard Disk		
Flash geheugen		
HD DVD		
BlueRay DVD		
Ik weet het niet		

16. In welke codeer- en compressieformaten³ is het digitale materiaal opgeslagen?

Kruis voor elk type opname uw keuze aan. Meerdere antwoorden per type mogelijk

AUDIO

- MPEG-1 Layer II
- MPEG-1 Layer III
- AAC
- HE-AAC
- Dolby Digital (AC3)
- Apple Lossless
- ATRAC (Sony)
- FLAC
- Real Audio
- Vorbis
- WMA
- Andere (specificeer):
- Onze organisatie bewaart geen audio-opnames
- Ik weet het niet

VIDEO

- MPEG-1
- MPEG-2
- MPEG-4 ASP (DivX, XVID)
- MPEG-4 AVC
- MPEG-4 SVC
- AVS

³ Een **codeer- of compressieformaat** is een signaal-transformatieformaat (analoog of digitaal) waarbij een zo groot mogelijke compressiewinst gekoppeld wordt aan een zo klein mogelijk kwaliteitsverlies.

- MJPEG
- Real Video
- VC-1
- WMV
- Andere (specificeer):
- Onze organisatie bewaart geen video-opnames
- Ik weet het niet

17. Waarom heeft uw organisatie voor deze codeerformaten gekozen?

Omschrijf kort uw motivatie

.....
.....

18. In welke containerformaten⁴ is het digitale materiaal opgeslagen?

Kruis uw keuze aan. Meerdere antwoorden mogelijk

- Audio Interchange File Format (Apple) (.aif)
- AU (Sun) (.au)
- Wave (IBM, Microsoft) (.wav)
- 3GP
- MP4 (.mp4)
- ASF (.asf)
- AVI (.avi)
- DMF (.dvix)
- DPX
- Flash Video (FVL)
- Matroska (.mtk)
- MXF
- AAF
- NUT
- Ogg (.ogg)
- Quicktime (.qt)
- Real Media (.rm)
- Andere (specificeer):
- Ik weet het niet

19. Waarom heeft uw organisatie voor deze containerformaten gekozen?

Omschrijf kort uw motivatie

.....
.....

U mag vraag 20 overslaan en de vragenlijst vervolgen vanaf **vraag 21**

⁴ Een **container- of opslagformaat** is een computer-bestandsformaat dat verschillende types van data kan bevatten (al dan niet gecodeerd/gecomprimeerd) (Cfr. een container).

20. U heeft aangegeven dat uw organisatie geen 'born digital' mondelinge bronnen registreert of ontvangt. Heeft uw organisatie in de nabije toekomst plannen voor het zelf digitaal registreren van mondelinge bronnen?

Kruis uw keuze aan en motiveer kort

- Onze organisatie doet zelf geen registraties
- Neen
 Indien neen, waarom niet?:.....
- Ja
 Indien ja, op welke termijn?.....

21. Worden er door uw organisatie niet-file-gebaseerde bronnen naar digitale vorm omgezet (digitalisering)?

Kruis uw keuze aan

- Neen (ga rechtstreeks naar **vraag 33**)
- Ja (vervolledig hieronder de vragenlijst)

22. Hoeveel bronnen bezit uw organisatie die naar digitale vorm (file gebaseerd) omgezet zijn?

Vul het aantal in

Audio mondelinge bronnen
 Video mondelinge bronnen

23. Op welke dragers wordt dit gedigitaliseerd materiaal opgeslagen?

Kruis voor elk type opname die van toepassing is uw keuze aan. Meerdere antwoorden per type mogelijk

	Audio	Video
Diskette (8, 5 ¼, 3 ½)		
CD-r(w)		
ZipDisk		
MiniDisk / Hi-MD		
LaserDisk		
DVD-r(w)		
Digital tapes (LTO-x)		
Hard Disk		
Flash geheugen		
HD DVD		
BlueRay DVD		
Ik weet het niet		

24. In welke codeer- en compressieformaten is het gedigitaliseerde materiaal opgeslagen?

Kruis voor elk type opname uw keuze aan. Meerdere antwoorden per type mogelijk

AUDIO

- MPEG-1 Layer II
- MPEG-1 Layer III
- AAC
- HE-AAC
- Dolby Digital (AC3)
- Apple Lossless
- ATRAC (Sony)
- FLAC
- Real Audio
- Vorbis
- WMA
- Andere (specificeer):
- Onze organisatie bewaart geen audio-opnames
- Ik weet het niet

VIDEO

- MPEG-1
- MPEG-2
- MPEG-4 ASP (DivX, XVID)
- MPEG-4 AVC
- MPEG-4 SVC
- AVS
- MJPEG
- Real Video
- VC-1
- WMV
- Andere (specificeer):
- Onze organisatie bewaart geen video-opnames
- Ik weet het niet

25. Waarom heeft uw organisatie voor deze codeerformaten gekozen?

Omschrijf kort uw motivatie

.....
.....

26. In welke containerformaten is het gedigitaliseerde materiaal opgeslagen?

Kruis uw keuze aan. Meerdere antwoorden mogelijk

- Audio Interchange File Format (Apple) (.aif)
- AU (Sun) (.au)
- Wave (IBM, Microsoft) (.wav)
- 3GP
- MP4 (.mp4)
- ASF (.asf)
- AVI (.avi)
- DMF (.dvix)
- DPX
- Flash Video (FVL)
- Matroska (.mtk)
- MXF
- AAF
- NUT
- Ogg (.ogg)
- Quicktime (.qt)
- Real Media (.rm)
- Andere (specificeer):
- Ik weet het niet

27. Waarom heeft uw organisatie voor deze containerformaten gekozen?

Omschrijf kort uw motivatie

.....

.....

28. Welk materiaal komt in uw organisatie het eerst in aanmerking voor digitalisering?

Duid de drie belangrijkste criteria aan die door uw organisatie gehanteerd worden om niet-file-gebaseerd materiaal te digitaliseren?

Plaats '1' bij het belangrijkste criterium, '2' bij het tweede en '3' bij het derde belangrijkste criterium. Indien u slechts één criterium hanteert, hoeft u enkel een '1' te plaatsen enz...

Materiaal in de slechtste staat	
Oudste materiaal	
Meest geconsulteerde materiaal	
Gemakkelijkst te digitaliseren	
Belangrijkste of meest waardevolle materiaal	
Willekeurige selectie	
Op thematische basis (bv. n.a.v. een tentoonstelling)	

Andere criteria (specificeer):	
--------------------------------------	--

29. Wat gebeurt er na de digitalisering met de originelen?

Kruis uw keuze aan. Meerdere antwoorden mogelijk

- Worden vernietigd
- Worden bijgehouden zonder specifieke inspanningen
- Worden zo goed mogelijk bewaard en in stand gehouden
- Andere (specificeer):

30. In welke mate bent u tevreden over de diverse aspecten van de door uw organisatie gevoerde digitaliseringsaanpak?

Duid op onderstaande schaal voor elk van de aspecten de tevredenheid van uw organisatie aan tussen 0 (= zeer ontevreden) en 10 (= zeer tevreden)

	Zeer ontevreden ⇒ zeer tevreden										
Digitaliseringsproces	0	1	2	3	4	5	6	7	8	9	10
Aantal digitale bronnen	0	1	2	3	4	5	6	7	8	9	10
Codeerformaten	0	1	2	3	4	5	6	7	8	9	10
Beveiliging	0	1	2	3	4	5	6	7	8	9	10
Conversie analoog-digitaal	0	1	2	3	4	5	6	7	8	9	10

31. Zijn er plannen om in de nabije toekomst het digitaliseringsproces van de collectie te wijzigen?

Kruis uw keuze aan en omschrijf kort

- Neen
- Ja
 Welke:.....

32. Bewaart u bij file gebaseerde mondelinge bronnen een kopie of back-up?

Kruis uw keuze aan

- Ja
- Neen

U mag vraag 33 overslaan en de vragenlijst vervolgen vanaf deel C vraag 34

33. U heeft aangegeven dat uw organisatie geen mondelinge bronnen digitaliseert. Heeft uw organisatie in de nabije toekomst plannen voor het digitaliseren van de collectie?

Kruis uw keuze aan en motiveer kort

- Neen
Indien neen, waarom niet?:.....
- Ja
Indien ja, op welke termijn?.....

C. Situatie van ontsluiting

In dit onderdeel gaan we dieper in op de manier waarop de collectie van uw organisatie ontsloten is aan de hand van metadata. Met ontsluiten bedoelen we het gestructureerd beschrijven via een classificatiesysteem zodat bronnen op gelijk welke wijze snel en efficiënt kunnen teruggevonden worden door internen en/of externen. Onder 'metadata' verstaan we data die informatie verstrekken over de mondelinge bronnen.

34. Hoeveel procent van de totale collectie is naar uw schatting ontsloten?

Vul het percentage in

.....% (indien 0%, ga rechtstreeks naar deel D vraag 42)

35. Op welke manier is de totale collectie ontsloten?

Kruis uw keuze aan. Meerdere antwoorden mogelijk

- Intranet
- Internet
- Lokaal databasesysteem op computer (vb. Access, Excel)
- Fichebak
- Catalogus (in boekvorm)
- Fysieke consultatie (vb. via persoon aan de balie)
- Andere (specificeer):

36. Voor wie is de collectie mondelinge bronnen van uw organisatie toegankelijk?

Kruis uw keuze aan. Slechts één antwoord mogelijk

- Enkel voor internen
- Enkel voor externen
- Zowel voor in- als externen
- Voor niemand

37. Welke problemen ondervindt uw organisatie bij het ontsluiten van mondelinge bronnen?

Kruis uw keuze aan. Meerdere antwoorden mogelijk

- Geen
- Verouderings- en ontbindingsverschijnselen bij het materiaal
- Gebrek aan financiële middelen
- Gebrek aan opslagcapaciteit
- Catalogiseren van het materiaal (tekort aan metadata of inconsistent gebruik)
- Tekort aan afspeelapparatuur
- Verouderde afspeelapparatuur
- Tekort aan opnameapparatuur
- Verouderde opnameapparatuur
- Incompatibel worden van softwaresystemen
- Tijdsgebrek
- Verdwijnen/onbeschikbaarheid van het materiaal (bvb. door ontlening)
- Andere (specificeer):

38. Gebruikt uw organisatie standaarden voor de beschrijving van metadata?

Kruis uw keuze aan

- Ja
 Indien ja, welke?
 - Dublin Core Metadata Initiative
 - MPEG-7
 - SMPTE Metadata Dictionary
 - EBU P/Meta
 - SMEF
 - TV-Anytime
 - ABC-Ontology
 - Andere (specificeer): ...
- Neen
 Waarom niet?.....
 Hoe zorgt uw organisatie dan voor consistentie in de beschrijvingen?

- Ik weet het niet

39. Waarom heeft uw organisatie gekozen voor het formaat van metadata dat u vandaag gebruikt?

Omschrijf kort uw motivatie indien van toepassing

.....

40. Welke van onderstaande metadata worden momenteel door uw organisatie gehanteerd voor het beschrijven en ontsluiten van mondelinge bronnen?

Kan u daarnaast ook aangeven welke metadata wenselijk zijn om in de nabije toekomst te implementeren?

Kruis uw keuze aan

	Gehanteerde metadata	Gewenste metadata
Naam geïnterviewde		
Naam interviewer		
Transcriptie aanwezig		
Bandinhoudfiche aanwezig		
Identificatiefiche		
Contract aanwezig		
Titel van de serie of fragment		
Locatie interview		
Tijdstip (datum en tijd)		
Genre		
Doelgroep		
Thema		
Lengte (bvb. 0:25:10)		
Korte beschrijving van de inhoud van het fragment/item		
Kernwoorden		
Producer		
Datum en uur ingave in archief		
Aantal keren dat het fragment is opgevraagd, beluisterd of bekeken		
Rechthebbende		
Verwijzing naar Contract/clausule		
Eigenlijke rechten		
Codecs		
Resolutie		
Bitsnelheid		
URI media		
Sample frequency		
Bits per sample		
Andere (vul aan):		

41. In welke mate is uw organisatie tevreden over de manier waarop mondelinge bronnen vandaag ontsloten worden?

Duid op onderstaande schaal de tevredenheid van uw organisatie aan tussen 0 (= zeer ontevreden) en 10 (= zeer tevreden)

Zeer ontevreden ⇒ zeer tevreden										
0	1	2	3	4	5	6	7	8	9	10

D. Technische infrastructuur

Dit onderdeel polst naar de huidige situatie inzake technische infrastructuur, als ook software van uw organisatie.

42. Welke diverse softwaresystemen worden door uw organisatie gebruikt voor opname⁵, bewaring⁶ en ontsluiting⁷?

Kan u invullen welke systemen gebruikt worden voor de beide types wat betreft opname, bewaring en ontsluiting? Indien iets voor uw organisatie niet van toepassing is, laat dit veld dan blanco.

Kan u eveneens aangeven of de gebruikte softwaresystemen flexibel zijn in gebruik? (U kan bijvoorbeeld gemakkelijk functionaliteiten toevoegen, veranderen of weglaten in het systeem of het is gemakkelijk om je data te migreren en aldus over te stappen naar een andere software leverancier). Omcirkel wat van toepassing is voor uw organisatie. Indien niet van toepassing, omcirkel 'NVT'.

	Audio	Flexibel	Video	Flexibel
Opname		Ja/nee/NVT		Ja/nee/NVT
Bewaring		Ja/nee/NVT		Ja/nee/NVT
Ontsluiting		Ja/nee/NVT		Ja/nee/NVT

43. Welke diverse apparatuur wordt door uw organisatie gebruikt voor opname, bewaring en ontsluiting?

⁵ Softwaresystemen voor opname zijn opname- en editeerprogramma's zoals Audacity, Ulead, Adobe, ...

⁶ Softwaresystemen voor bewaring zijn databanken zoals Oracle, MySQL en Media Asset Management Systemen zoals Synergy, Ardome, ...

⁷ Softwaresystemen voor ontsluiting zijn Content Management Systemen zoals Drupal, Joomla, ...

Kan u invullen welke apparatuur uw organisatie gebruikt voor de beide types wat betreft opname, bewaring en ontsluiting? Indien iets voor uw organisatie niet van toepassing is, laat dit veld dan blanco.

	Audio	Video
Opname		
Bewaring		
Ontsluiting		

44. Kan u aangeven wat de huidige beschikbare opslagcapaciteit (computer, server, harde schijven) is voor file gebaseerd materiaal?

Vul het werkelijke aantal in

..... gigabyte

45. Zijn er plannen om in de nabije toekomst wijzigingen door te voeren inzake de technische infrastructuur, zowel wat betreft software als apparatuur?

Kruis uw keuze aan en omschrijf kort

- Neen
 Ja

Indien ja, welke?:.....

F. Rechtenbeheer

Dit onderdeel van de vragenlijst behandelt het omgaan met de rechten die aan mondelinge bronnen gekoppeld zijn.

46. Weet u welke rechten gelden op het materiaal uit de audiovisuele collectie?

Kruis uw keuze aan

- Neen
- Ja, de rechten van alle materiaal zijn gekend
- Ja, voor een gedeelte van het materiaal, voor een ander deel niet

47. Hoe zijn deze rechten vastgelegd (op contractuele basis, mondelinge overeenkomst...)? Welke zijn de belanghebbenden? Tussen welke partijen wordt een contract aangegaan?

Omschrijf kort de actuele situatie

.....

48. Gebruikt uw organisatie specifieke software, een databestand of een eigen beheerssysteem voor het rechtenbeheer?

Kruis uw keuze aan en motiveer kort

- Ja
 Welke?:.....
- Neen

49. Is er iemand in uw organisatie die meewerkt aan het opstellen van contracten, onderhandelingen over rechten, ...?

Kruis uw keuze aan

- Ja
- Neen

50. Is uw organisatie in het verleden reeds door derden aansprakelijk gesteld voor het onrechtmatig hergebruik van door u bewaarde mondelinge bronnen?

Kruis uw keuze aan

- Ja
- Neen

51. In welke mate is uw organisatie tevreden over de manier waarop het rechtenbeheer van mondelinge bronnen vandaag verloopt?

Duid op onderstaande schaal de tevredenheid van uw organisatie aan tussen 0 (= zeer ontevreden) en 10 (= zeer tevreden)

Zeer ontevreden ⇒ zeer tevreden										
0	1	2	3	4	5	6	7	8	9	10

52. Zijn er plannen om in de nabije toekomst wijzigingen door te voeren inzake de organisatie van rechtenbeheer?

Kruis uw keuze aan en omschrijf kort

- Neen
 Ja

Indien ja, welke?:.....

53. Indien u online mondelinge bronnen ter beschikking stelt: welke maatregelen neemt uw organisatie om digitaal materiaal te beveiligen en ongeoorloofde verspreiding tegen te gaan?

Omschrijf kort indien van toepassing

.....

G. Gezamenlijke ontsluiting

Het Van Horen Zeggen III-project onderzoekt de diverse behoeften op vlak van opname, bewaring en ontsluiting van mondelinge bronnen. Doelstelling is het ontwikkelen van een innovatieve applicatie met het oog op de ontsluiting van mondelinge bronnen.

54. Bent u geïnteresseerd in een gemeenschappelijk systeem voor ontsluiting van mondelinge bronnen? Met een gemeenschappelijk systeem bedoelen dat verschillende organisaties een zelfde systeem gebruiken voor hun ontsluiting.

Kruis uw keuze aan

- Ja
 Neen

55. Kan u kort omschrijven wat u als voor- en nadelen ziet van een gemeenschappelijk ontsluitingssysteem?

Omschrijf kort

Indien ja, welke zijn hiervan dan de mogelijke voordelen?

.....

Indien neen, welke zijn hiervan dan de mogelijke nadelen?

.....

.....

.....

.....

.....

H. Behoeften van de gebruiker

Dit laatste deel behandelt de functionaliteiten van het systeem voor de eindgebruiker beschouwd vanuit het perspectief van de collectiebeheerder.

Met eindgebruiker bedoelen we niet zozeer de residentiële gebruiker, wel die personen in instellingen die – al dan niet tegen betaling – op een structurele wijze gebruik wensen te maken van archiefmaterialen, zoals de medewerkers van wetenschappelijke, educatieve, socio-culturele en/of overheidsorganisaties

56. Welk van onderstaande functionaliteiten worden volgens u door de eindgebruiker al dan niet belangrijk geacht?

Kruis voor elke functie uw keuze aan

Hoe belangrijk zou u volgende mogelijkheden vinden om binnen het archief te zoeken:	Helemaal niet belangrijk	Niet belangrijk	Eerder niet belangrijk	Eerder belangrijk	Belangrijk	Zeer belangrijk
Eenvoudig zoeken (free text zoekveld, cfr. Google search)						
Doorzoeken van de collectie aan de hand van bepaalde metavelden (vb. titel, naam geïnterviewde,) (zoekwoord vrij in te geven)						
Voor een bepaald veld zoeken adhv een thesaurus (vb. bij thema kan je kiezen uit de vaste lijst: WOII, verkiezingen, de mijnbouw, sport, ...)						
Online bekijken of beluisteren van fragmenten (streaming)						
Online raadplegen van transcripties						
Mogelijkheid tot gratis downloaden van de mondelinge bronnen						

Mogelijkheid tot downloaden van de mondelinge bronnen tegen betaling						
Mogelijkheid tot het aanvragen van kopieën op CD of DVD						
Verfijnen van een zoekopdracht (zonder een volledig nieuwe zoekopdracht te lanceren)						
Suggesties op basis van gebruikers met gelijke interesse						
Suggesties op basis van categorieën (andere bronnen die in gelijkaardige categorieën vallen als het resultaat van uw oorspronkelijke zoekopdracht)						
Hoe belangrijk zou u volgende mogelijkheden vinden om binnen het archief te zoeken:	Helemaal niet belangrijk	Niet belangrijk	Eerder niet belangrijk	Eerder belangrijk	Belangrijk	Zeer belangrijk
Suggesties op basis van eigen ingevoerd profiel						
Mogelijkheid tot opzoeking per instelling						
Bewaren van de eigen selecties						
Mogelijkheid tot aanmaken van een eigen gebruikersprofiel						
Materiaal opzoeken aan de hand van gelijkaardige beelden (picture similarity)						
Een zoekopdracht opslaan voor later gebruik						
De mogelijkheid om te kiezen welke velden (metadata) en visuele informatie (keyframes, storyboard, ...) worden weergegeven in de resultatenlijst						
Online toevoegen van commentaar door gebruikers						
Online toevoegen van een score door de gebruikers						
Online toevoegen van trefwoorden door gebruikers						
Online omgeving voor discussie tussen eindgebruikers over bepaalde thema's						
Omgeving waar de eindgebruiker zelf content kan bewerken						
Online doorzoeken van de audiobestanden (op klanken)						
Bestanden raadplegen of downloaden in lage resolutie						

Bestanden raadplegen of downloaden in hoge resolutie						
Andere: (specificeer).....						

Van harte bedankt voor het invullen van de vragenlijst!

Wij willen er u nogmaals aan herinneren ons deze vragenlijst **tegen uiterlijk 7 mei** toe te sturen via de enveloppe die u bij deze vragenlijst heeft gevonden.

12.2 Bijlage 2: Respons schriftelijke vragenlijsten

Geantwoord

Erfgoedcel Antwerpen (Frank Herman)
Erfgoedcel Brugge (Ina Verrept & Karolien Steen)
Erfgoedcel Brussel (Ingeborg De Cooman)
Erfgoedcel Gent (Hendrik Defoort)
Erfgoedcel Hasselt (Sabine Hartmann)
Erfgoedcel Ieper (Eva Wuyts)
Erfgoedcel Kortrijk (Veerle Van den Abeele)
Erfgoedcel Leuven (Rebecca Gysen)
Erfgoedcel Meetjesland (Sylvia Matthys & Sandrine De Wilde)
Erfgoedcel Sint-Truiden (Wivina De Bus)
Erfgoedcel Tongeren (Ward Segers & Bruno Claes)
Erfgoedcel Waasland (Ode De Zutter)

AMSAB – Instituut voor Sociale Geschiedenis (Bart De Nil & Donald Weber)
AMVB (Patricia Quintens & Mariet Calsius)
KADOC (Peter Heyrman & Luc Schokkaert)
Liberaal Archief (Luc Pareyn)
SOMA (Anne Roekens & Jan Laplasse)
Stadsarchief Antwerpen (Matthias Van der Maesen)
VRT Beeld- en Geluidsarchief (Lieve Van Der Straeten & Chris Steyaert)

12.3 Bijlage 3: Rondetafelgesprekken

Erfgoedcellen:

Datum: 14 mei

Locatie: Vlaams Centrum voor Volkscultuur

Aanwezig:

Erfgoedcel Brugge: Karolien Steen
Erfgoedcel Hasselt: Sabine Hartmann
Erfgoedcel Land Van Waas: Ode De Zutter

Archiefinstellingen:

Datum: 14 mei

Locatie: Vlaams Centrum voor Volkscultuur

Aanwezig:

ADVN: Koen Van Keer
AMSAB: Donald Weber
AMVB: Mariet Calsius
KADOC: Luc Schokkaert
Liberaal Archief: Luc Pareyn

VRT Geluidsarchief:

Datum: 16 mei

Locatie: VRT

Aanwezig:

VRT: Lieve van Der Straeten en Chris Steyaert

SOMA:

Datum: 16 mei

Locatie: SOMA

Aanwezig:

SOMA: Jan Laplasse en Anne Roekens