

NETWERK DIGITAAL ERFGOED

Annotatie Webarchief Groningen obv GTAA en OpenSKOS

Eindrapport

Project B3-1

Juni 2017

Auteur:
Lotte Baltussen

In opdracht van
Netwerk Digitaal Erfgoed

Inhoudsopgave

Managementsamenvatting	3
Netwerk Digitaal Erfgoed	5
Relaties met andere NDE-netwerkactiviteiten	6
<i>NDE houdbaar</i>	6
<i>NDE bruikbaar</i>	6
<i>Visuele Thesaurus voor Mode & Kostuums</i>	6
Betrokkenheid in het erfgoedveld	7
<i>Klankbordgroep</i>	7
<i>Bredere betrokkenheid</i>	7
<i>Presentaties</i>	7
Projectresultaat 1: metadatamodel voor het webarchief	8
Metadatamodellen voor webarchieven: algemeen	8
Metadatamodel - optie 1: MODS-lite	9
Metadatamodel - optie 2: OWMS 4.0	9
Metadatamodel - optie 3: MARC	10
Metadatamodel - optie 4: METS	10
Klankbordgroepdiscussie: erfgoedperspectief vs archiefperspectief	10
Conclusie	11
Projectresultaat 2: implementatie gtaa-onderwerpsas in mais-flexis 9	13
Inhoudelijke vraagstukken	14
<i>Inhoudelijke aansluiting</i>	14
<i>Van hiërarchie naar graph</i>	14
<i>Controle over het toevoegen van concepten aan de gtaa</i>	15
Technische implementatie: mogelijkheden en gemaakte keuzes	16
<i>Concepten aanroepen: live API implementatie vs een OAI-PMH harvest</i>	16
<i>Concepten annoteren: wat laat je zien, wat sla je op?</i>	19
<i>Concepten binnenhalen: relevante velden</i>	19
<i>Zoekresultaten duiden: context is everything</i>	20
<i>Het autocomplete-overzicht</i>	21
<i>Concepten kiezen en koppelen aan een record</i>	22
<i>Omgaan met dynamiek: veranderende en nieuwe termen</i>	22
<i>Omgaan met veranderende en verwijderde concepten</i>	23
<i>Nieuwe termen voorstellen</i>	25
Projectresultaat 3: onderzoek inzet persistent identifiers	27
Keuze maken uit de diverse PID-systemen	27
Bredere ontwikkelingen met betrekking tot PIDS	28
Conclusie	29
Inhoudelijk	29
Technisch	29
Resources	30
Vervolgstappen	31
Bibliografie	32
Bijlage 1 - mods-lite velden geïmplementeerd bij groningen archieven	34
Bijlage 2 - betrokkenen	36

Managementsamenvatting

Het Netwerk Digitaal Erfgoed (NDE) is een samenwerkingsverband dat zich richt op de ontwikkeling van een stelsel van landelijke voorzieningen en diensten voor het verbeteren van de zichtbaarheid, bruikbaarheid en houdbaarheid van digitaal erfgoed.

Eén van de speerpunten van het NDE is het faciliteren en ontwikkelen van een nationaal termennetwerk. Binnen dit netwerk kunnen instellingen bestaande thesauri van andere (erfgoed)instellingen gebruiken, om zo hun collectie te verrijken en te verbinden met andere collecties. Daarbij worden thesauri ook onderling met elkaar verbonden.

Binnen het NDE-project “Annotatie Webarchief Groningen obv GTAA en OpenSKOS”¹ (november 2015-november 2016) is een praktische use case uitgewerkt, om meer inzicht te krijgen in wat er nodig is om een nationaal gedeeld termennetwerk op te zetten. Binnen dit project is de use case uitgewerkt waarbij de Onderwerps-as van de thesaurus van het Nederlands Instituut voor Beeld en Geluid (de GTAA)² geïntegreerd is in het collectiebeheersysteem van het RHC Groninger Archieven, MAIS-Flexis.

Om de scope van het project behapbaar te houden is de Onderwerps-as gekoppeld aan een specifieke collectie: het webarchief. Groninger Archieven startte in juli 2015 met het project Webarchief Groningen, maar de tot nu toe rond de 100 gearchiveerde websites stonden voor aanvang van dit project nog niet in hun collectiebeheersysteem MAIS-Flexis en waren slechts zeer beknopt gemetadateerd op het platform van de archiverende partij (GW Crossmedia).

De doelstelling van dit NDE-project was tweeledig:

1. onderzoek naar de beste inhoudelijke en technische aanpak om de GTAA Onderwerps-as te integreren in MAIS-Flexis, om het webarchief handmatig te annoteren.
2. uitvoering van de implementatie van deze workflow in het webarchief van Groninger Archieven, wat leidt tot generieke aanbevelingen voor het opzetten van het termennetwerk en het gebruik maken van gedeelde thesauri door individuele erfgoedinstellingen.

Tijdens het project is tevens een metadatamodel voor het webarchief van Groninger Archieven gekozen, om de collectie te kunnen laten landen in MAIS-Flexis. Daarnaast is in samenwerking met het project Persistent Identifiers (PIDs) binnen het werkpakket NDE Houdbaar gekeken naar de mogelijke inzet van PIDS voor het webarchief.

De resultaten zijn als volgt:

Implementatie GTAA-Onderwerpsas

- In het termennetwerk moet een helder overzicht worden gemaakt van de beschikbare thesauri en op welke type collecties zij over het algemeen goed aansluiten, inclusief concrete voorbeelden en korte documentatie van de inhoud van de thesauri.
- Onderwerpen zijn lastiger te definiëren zijn dan Named Entities, zoals persoonsnamen en plaatsnamen. Daarbij heeft iedere collectie zijn eigen specifieke eigenschappen, die mogelijk even specifieke onderwerpen benodigen. Dit maakt gezamenlijk beheer enorm complex. Uiteindelijk zal voor een brede adoptie van een termennetwerk het noodzakelijk zijn voor instellingen om een verminderde mate van controle over gebruikte termen te accepteren.
- Onderwerpen toekennen wordt aanzienlijk vergemakkelijkt wanneer de context van een

¹ <http://www.den.nl/project/613/Annotatie-Webarchief-Groningen-op-basis-van-GTAA-en-OpenSKOS>

² <http://gtaa.beeldengeluid.nl/>

term wordt getoond in het overzicht met autocomplete-suggesties, zoals bredere of specifiekere termen en scope notes.

- Het is voor het faciliteren van een brede adoptie belangrijk dat er een deugdelijke infrastructuur wordt opgezet die enerzijds flexibel en laagdrempelig is, maar anderzijds zeer robuust en stabiel.
- Het annotatieproces kan vergemakkelijkt worden door dit vanuit één veld mogelijk te maken voor verschillende type onderwerpen (eg persoonsnamen, geografische namen). Aangezien aan een thesaurusterm een type gekoppeld is kan dit uiteindelijk weer verdeeld worden over meerdere velden en/of zoekfilters.
- De impact van het gebruik van thesauri uit het termennetwerk op de resources in organisaties moet verder worden uitgewerkt, zowel voor nieuwe thesaurusgebruikers, de thesaurusaanbieders en de leveranciers.

Metadatamodel voor webarchieven

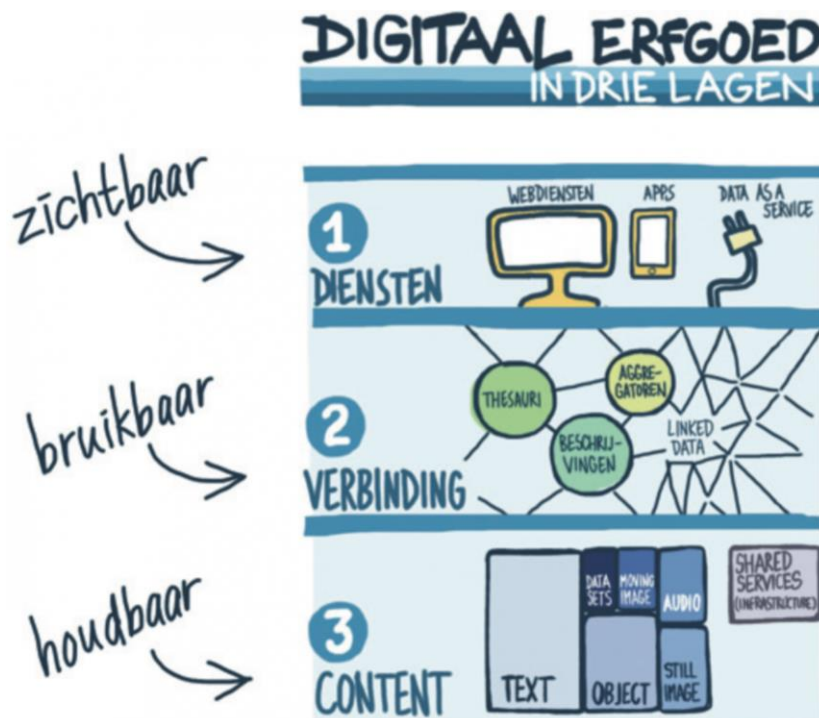
- Er zijn diverse standaarden (die zijn aangepast) voor het metadateren van webarchieven, maar er zijn nog weinig 'best practices'. Een extra complicatie is dat een webarchief kan worden aangelegd vanuit een erfgoedperspectief (collectievorming) of vanuit archiefperspectief (Archiefwet). Deze perspectieven resulteren ook in andere vereisten voor de metadatering.
- Wanneer er wensen binnen je organisatie zijn om in de toekomst over te stappen naar het gezamenlijke e-depot van het Nationaal Archief, houdt dan bij de implementatie van het metadatamodel van de webarchief rekening mee dat deze aansluit op TopX 2.3.
- De uitkomsten van dit project worden overgedragen aan het recentelijk door de NCDD opgerichte Kennisplatform Webarchivering, om te borgen dat de metadateringsdiscussie verder wordt gevoerd in een nationale context.

Inzet van PIDs voor webarchieven

- Er zijn verschillende systemen om PIDs te implementeren. De belangrijkste die binnen de erfgoedwereld gebruikt worden en aanbevolen worden voor de NCDD zijn Handle, URN/NBN en DOI. De PID-wijzer die binnen NDE Houdbaar door de NCDD is ontwikkeld biedt een goede houvast bij het maken van een keuze.
- PIDs kunnen op meerdere niveaus worden ingezet voor webarchieven: van een record van een crawl tot de individuele assets in een (W)ARC file. Binnen dit project is er besloten om te beginnen met PIDs voor records, aangezien het ook een algemene wens van Groninger Archieven is om PIDs in te gaan zetten voor alle records in MAIS-Flexis.

Netwerk Digitaal Erfgoed

Het project “Annotatie Webarchief Groningen obv GTAA en OpenSKOS” is een onderdeel van het werkprogramma Digitaal Erfgoed Bruikbaar van het NDE. Dit samenwerkingsverband richt zich op de ontwikkeling van een stelsel van landelijke voorzieningen en diensten voor het verbeteren van de zichtbaarheid, bruikbaarheid en houdbaarheid van digitaal erfgoed. Het netwerk is gestart op initiatief van het Ministerie van Onderwijs, Cultuur & Wetenschap (OCW) en past in de nationale strategie digitaal erfgoed.³ Deelnemers aan het samenwerkingsverband zijn onder meer de Koninklijke Bibliotheek, het Nederlands Instituut voor Beeld en Geluid, de Rijksdienst voor het Cultureel Erfgoed, de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW), het Nationaal Archief en kenniscentrum DEN. Binnen het werkprogramma van de NDE zijn drie werkpakketten overeenkomend met de drie lagen uit de strategie voor digitaal erfgoed, namelijk:



Afbeelding 1: het drie-lagenmodel binnen NDE. Door: Elco van Staveren | Denkschets.nl

Het werkpakket *Digitaal Erfgoed Bruikbaar* is gesplitst in drie deelprojecten:

1. het realiseren van een landelijk en erfgoedbreed netwerk van aan elkaar gekoppelde termenbronnen
2. het ontwikkelen van erfgoedbrede softwaretools voor rijkere ontsluiting en semantische zoektechnieken;
3. het ondersteunen van stimulerende voorbeeldimplementaties.

Dit project past daarmee als voorbeeldproject naadloos in de doelstellingen van *Digitaal Erfgoed Bruikbaar*.

³ <http://www.rijksoverheid.nl/documenten-en-publicaties/publicaties/2015/03/09/nationale-strategie-digitaal-erfgoed.html>

Relaties met andere NDE Netwerkactiviteiten

NDE Houdbaar

Binnen dit NDE werkpakket, geleid door de NCCD zijn twee relevante projecten uitgevoerd waarmee contact is gehouden en informatie mee is uitgewisseld:

- 'Persistent Identifiers'⁴: Vanuit dit project is input geleverd mbt de inzet van PIDs voor het webarchief van Groninger Archieven.
- 'Organiseren van coördinatie van webarchivering in Nederland'⁵: De projectleider van het project (Arnoud Goos, Beeld en Geluid) was lid van de klankbordgroep en heeft advies gegeven over het metadatamodel voor webarchieven. Vice versa heeft Lotte Baltussen zitting genomen in de klankbordgroep van 'Organiseren van coördinatie van webarchivering in Nederland'.

NDE Bruikbaar

Een belangrijk speerpunt van NDE is het realiseren van een 'knowledge graph'⁶ voor digitaal erfgoed. Dit is de verzameling van met elkaar verbonden terminologiebronnen, die als referentie dient voor het maken van verwijzingen naar plaatsen, personen, onderwerpen, tijdsperiode, gebeurtenissen in de beschrijvingen van de collecties en objecten.

Visuele Thesaurus voor Mode & Kostuums

Dit project (november 2015 - juni 2017) is getrokken door het Modemuseum in Antwerpen (MoMu) en geïnspireerd door het portaal Modemuze.nl. Het heeft tot doel het gebruik van gedeelde termenlijsten te motiveren binnen Vlaamse en Nederlandse musea met mode-erfgoed. In dat kader is er door een consortium van 11 musea gewerkt aan de versterking van de Art & Architecture Thesaurus (AAT), waarvoor 130 - al deels verwerkte - toevoegingen zijn aangeboden bij het Getty Research Institute. Hierbij is gefocust op de objecttypes en textuele materialen, die na vergelijking met de collecties duidelijk ontbraken in de Amerikaanse thesaurus.

Daarnaast organiseerde het project workshops over het opschonen en linken van collecties voor de projectpartners, om deze kennis vervolgens te verankeren in een handleiding voor het brede werkveld (in samenwerking met NDE). De eigen set termen is samen met de bestaande mode en kostuum-gerelateerde AAT-concepten ontsloten als de Visuele Thesaurus voor Mode & Kostuums. Inclusief de links naar op Wikimedia Commons geplaatste illustraties is deze thesaurus als SKOS beschikbaar. De thesaurus wordt geïmplementeerd in het platform Modemuze.nl, om de effecten te kunnen tonen aan het werkveld en dit verder bewust te maken van de kansen van Linked Data.

Voor dit project werkte het Antwerpse Modemuseum in de stuurgroep samen met Modemuze, NDE, VIAA, RCE en werd ondersteund door Spinqe, de Universiteit Gent en het RKD. Betrokken musea: Amsterdam Museum, Rijksmuseum Amsterdam, Centraal Museum, Paleis Het Loo, Fries Museum, Museum Rotterdam, Gemeentemuseum Den Haag, Koninklijke Musea voor Kunst & Geschiedenis, Koninklijk Instituut voor het Kunstpatrimonium, Modemuseum Hasselt, Modemuseum Provincie Antwerpen (MoMu).⁷

⁴ <http://www.nccd.nl/projecten/netwerk-digitaal-erfgoed/project-persistent-identifiers/>

⁵ <http://www.nccd.nl/projecten/netwerk-digitaal-erfgoed/project-coördinatie-webarchivering/>

⁶ <http://www.den.nl/project/614/Een-knowledgegraph-voor-de-integratie-van-erfgoedinformatie>

⁷ <http://modemuze.tumblr.com/post/145257284654/thesaurusproject-mode-erfgoed-beter-vindbaar>

Betrokkenheid in het erfgoedveld

Klankbordgroep

In het projectplan is een representatieve klankbordgroep benoemd die op diverse gebieden (thesauri, webarchief metadata, PIDs) gedurende het project advies heeft gegeven. Het ging om medewerkers van de volgende organisaties:

- Beeld en Geluid
- Groninger Archieven
- Koninklijke Bibliotheek
- Nationaal Archief
- Utrechts Archief

Er zijn meerdere bijeenkomsten in wisselende samenstellingen geweest om de belangrijkste ontwikkelingen en vraagstukken door te nemen.

Bredere betrokkenheid

Naast de klankbordgroep is er contact gezocht met andere experts in het erfgoedveld, van collectiebeheerders tot leveranciers van collectiebeheersystemen. Afhankelijk van de onderwerpen zijn deze experts betrokken bij 'klankbordgroep XL' bijeenkomsten, gebruikersonderzoeken, of middels één-op-één gesprekken.

Daarnaast is er een Basecamp-omgeving voor het project ingericht, waar zowel mensen uit bovenstaande organisaties als andere geïnteresseerden online op de hoogte konden blijven en feedback konden geven.

Voor het volledige overzicht van de betrokken organisaties en personen, zie Bijlage 2 - Betrokkenen.

Presentaties

Om de kennis die is opgedaan tijdens het project te delen zijn er op diverse evenementen presentaties gegeven:

- 16 december 2015. OCW, Den Haag. Leveranciersmiddag NDE.
- 24 mei 2016. Nationaal Archief, Den Haag. Interne projectpresentatie.
- 1 juli 2016. Beeld en Geluid, Hilversum. [AVA Net Symposium 2016](#).
- 22 september 2016. Reinwardt Academie, Amsterdam. Module Digitaal Erfgoed.
- 28 september 2016. Beeld en Geluid, Hilversum. [Conferentie Onderwijs en Linked Data](#).
- 7 april 2017. Brabanthallen, Den Bosch. [NDE-conferentie 2017](#).

Projectresultaat 1: Metadatamodel voor het webarchief

Tot de start van het project in november 2015 was het archief alleen vindbaar en doorzoekbaar binnen de beheersomgeving van de archiefpartij GW Crossmedia. Voordat de GTAA Onderwerps-as gebruikt kon worden om het webarchief van Groninger Archieven te annoteren, moest dit webarchief ten eerste beschikbaar worden in MAIS-Flexis. Ten eerste is er door DE REE een zogeheten 'Archiefeenheidsoort' (AET) geïmplementeerd, die een set aan eigenschappen bevat die van toepassing zijn op een specifieke collectie. Het webarchief van Groninger Archieven vooralsnog bestaat uit minder dan 100 URLs van websites met een relatie met het dagelijks leven van Groningen. Dit zal in de toekomst nog uitgebreid worden met meer van hetzelfde soort websites, maar wellicht ook met social media bronnen of apps. Hiertoe heeft de Archiefeenheidsoort de wat meer generieke en toekomstbestendige naam 'Internetbron' gekregen.

Na de implementatie van de AET kon deze gevuld worden met een metadatamodel dat geschikt is voor het beschrijven van webarchieven. De uitgangspunten waren dat er een metadatamodel werd gekozen dat:

1. aansluit bij de wensen en requirements van Groninger Archieven
2. flexibel is en in de toekomst uitgebreid kan worden als er nieuwe requirements bij komen
3. zo goed mogelijk aansluit bij de standaarden en denkwijzen van andere instellingen met webarchieven, zodat het bijvoorbeeld mogelijk wordt in de toekomst om webarchieven gezamenlijk aan te bieden en/of op te slaan in een gedeeld e-depot.

Groninger Archieven had al verkenningen gedaan naar metadatamodellen ism het Utrechts Archief en was hierbij uitgekomen bij het MODS-lite⁸ model, een afgeleide van het MODS-model⁹ dat door de Library of Congress is ontwikkeld voor onder meer het beschrijven van gearchiveerde website. Om te toetsen of dit model voldoet aan bovenstaande uitgangspunten, is de klankbordgroep van het project gevraagd zich over MODS-lite en mogelijke andere opties te buigen.

Metadatamodellen voor webarchieven: algemeen

Er zijn talloze standaarden en manieren om de metadata voor webarchieven te beschrijven, maar vaak zijn met name inhoudelijke beschrijvingen nog erg beperkt [Peterson 2015]. Dit komt doordat het nog een relatief nieuw type archief is én omdat er diverse lagen metadata te onderscheiden zijn. Zo definieert het OAIS model drie typen informatie, namelijk metadata over het opnemen van een collectie-item, het archiveren ervan én het toegankelijk maken.

Hieronder volgt een kort overzicht van gebruikte standaarden van belangrijke spelers in het veld. Belangrijk om hierbij op te merken is dat veel van de verder hieronder beschreven modellen een mengvorm zijn van bestaande standaarden die reeds in een organisatie werden gebruikt die later een webarchief zijn gaan aanleggen en andere standaarden (zoals Dublin Core en METS).

Enkele voorbeelden:

- Archive-It: een service van de Internet Archive waar instellingen zich op kunnen abonneren om websites te archiveren en beschrijven. Het metadatamodel van Archive-It bestaat uit de basisset van 15 Dublin Core elementen die naar wens aangevuld kan worden door

⁸ <https://www.loc.gov/standards/mods/v3/mods-userguide-lite.html>

⁹ <http://www.loc.gov/standards/mods/>

- gebruikers.¹⁰
- British Library: “Harvested Websites are stored in Archival Information Packages (AIP). Each AIP is described by a METS file. Operational metadata for resource discovery as well as archival metadata are normalized and embedded in the METS descriptor using common metadata profiles such as PREMIS and MODS.” [Enders 2010]

Metadatamodel - optie 1: MODS-Lite

Na de creatie van de AET kan het metadatamodel worden vormgegeven. Het voordeel van MAIS-Flexis is dat het zeer gemakkelijk is om zogeheten flexvelden toe te voegen en een metadatamodel uit te breiden, en het dus mogelijk is om te starten met een beperkte opzet en pas uit te breiden wanneer dit nodig blijkt te zijn.

Groninger Archieven heeft in overleg en samenwerking met het Utrechts Archief gekeken naar welk model geschikt is om hun beider webarchieven te metadateren, en zijn daarbij uitgekomen bij het MODS¹¹ model. MODS (Metadata Object Description Schema) is een beschrijvingsstandaard die is ontwikkeld door de Library of Congress, in eerste instantie om een meer simpele subset van MARC te creëren [Guenther 2003]. Het is een XML schema met een set bibliografische elementen die voor diverse doeleinden kan worden ingezet, zoals bibliotheekapplicaties en is gerelateerd aan het bibliografische formaat MARC 21. MODS wordt door de LoC en andere instellingen ook gebruikt als een metadatamodel voor webarchieven.¹²

Aangezien MODS nogal ‘zwaar’ en uitgebreid is, en Groninger Archieven en het Utrechts Archief simpel willen beginnen en velden in MAIS-Flexis gemakkelijk uit te breiden zijn, is ervoor gekozen om in eerste instantie alleen de velden in MODS-Lite¹³ in de AET te verwerken. Dit is een minder uitgebreide versie van MODS met elementen die specifiek zijn gebaseerd op de standaard 15 Dublin Core elementen.¹⁴ MODS-lite is echter nog steeds vrij uitgebreid, en bevat per element nog de mogelijkheid om attributen en sub-elementen toe te voegen, waarmee het ook compliant kan zijn met het meer verfijnde ‘Qualified Dublin Core’.¹⁵

Er zijn geen verplichte elementen binnen MODS. De enige vereiste is dat er per record minstens één element gevuld is.

De MODS-lite elementen die vooralsnog door Groninger Archieven en het Utrechts Archief zijn geselecteerd zijn [hier](#) te vinden.

Metadatamodel - optie 2: OWMS 4.0

[OWMS 4.0](#)¹⁶ is een metadatamodel dat in Nederland wordt gebruikt voor het metadateren van publieke overheidsinformatie op het internet. Het conceptueel model is gebaseerd op Dublin Core. In sommige gevallen zijn de definities specifiek gemaakt voor overheidsinformatie.

OWMS 4.0 houdt een minimale set van negen eigenschappen aan die toegepast moeten worden om

¹⁰ <http://dublincore.org/documents/dces/>

¹¹ <http://www.loc.gov/standards/mods>

¹² <http://www.netpreserve.org/sites/default/files/resources/Membersurvey.pdf>

¹³ <http://www.loc.gov/standards/mods/v3/mods-userguide-lite.html>

¹⁴ Echter, als je terug wilt van DC naar MODS, dan kun je wel informatie verliezen, aangezien DC platter is dan MODS.

¹⁵ <http://dublincore.org/resources/fag/#whatisthedifference> en <http://dublincore.org/documents/usaguide/qualifiers.shtml>

¹⁶ <http://standaarden.overheid.nl/owms/fag#n300>

OWMS conform te kunnen zijn:¹⁷

- Verplicht: Identificatie, Titel, Informatietype, Taal en Datum laatste wijziging.
- Verplicht mits van toepassing: Geldigheid en Locatie.
- Keuze uit minstens één van: Eindverantwoordelijke en Maker.

Een nadeel van OWMS is dat het een model is voor overheidsinformatie op het internet, maar niet voor webarchieven, zeker niet voor collecties zoals die van Groninger Archieven dat zich richt op het dagelijks leven van Groningers.

Metadatamodel - optie 3: MARC

Het MARC metadata model (MACHINE-Readable Cataloging) model is een metadatastandaard die voornamelijk wordt gebruikt voor het beschrijven van bibliografische records. Aangezien bibliotheken soms ook websites archiveren, gebruiken zij vaak het MARC model om deze webarchieven te beschrijven. De meest gebruikte variant van het model is MARC 21.¹⁸

Het is mogelijk om MARC naar MODS te mappen, maar omdat er hierbij informatie wordt 'platgeslagen', is het niet mogelijk om de informatie weer met dezelfde rijkheid terug te mappen van MODS naar MARC.¹⁹ De Library of Congress gebruikt MARC voor het beschrijven van deelcollecties van webarchieven. Voor de individuele crawls gebruiken ze MODS [Niu 2012]

Metadatamodel - optie 4: METS

METS (Metadata Encoding & Transmission Standard) is een framework voor het beschrijven van een digitale bron en alle componenten van deze bron. METS bevat zowel beschrijvende, technische en conserverings-metadata, en informatie over de interne structuur van de bron. METS kan met verschillende metadata schema's worden gevuld, zoals PREMIS (preserving), MODS en Dublin Core (beschrijving). Het is dus een relatief 'zwaar' en degelijk schema met diverse uitbreidings- en invulmogelijkheden. Het wordt onder meer gebruikt bij de British Library, dat een van de grootste webarchieven ter wereld beheert. [Enders 2010]

Klankbordgroepdiscussie: erfgoedperspectief vs archiefperspectief

MODS-lite kwam op basis van bovenstaande verkenning als winnaar uit de bus: het is een flexibel en schaalbaar formaat waar klein mee kan worden begonnen. Bovendien is MODS-lite goed gedocumenteerd en te mappen naar andere standaarden. Het is echter niet zo 'zwaar' als MODS en METS. Daarnaast is het een standaard die beter past in een archief als Groninger Archieven, in tegenstelling tot de bibliotheekstandaard MARC. Ten slotte valt OWMS af, omdat het meer betrekking heeft op 'live' overheidswebsites en voor zover bekend nog helemaal niet wordt gebruikt voor (web)archieven.

Binnen de klankbordgroep bleek echter dat het noodzakelijk was om de keuze van het metadatamodel goed door te lichten uit zowel een erfgoed- als archiefperspectief. De reden hiervoor was dat er cruciale conceptuele verschillen zijn tussen het beschrijven van webarchieven als cultureel

¹⁷ <http://standaarden.overheid.nl/owms/4.0/doc/NormatievespecificatieOWMS4.0v1.0.1.pdf>

¹⁸ <http://www.loc.gov/marc/>

¹⁹ <http://www.loc.gov/standards/mods/mods-mapping.html>

erfgoed (zoals Groninger Archieven dit goed) en webarchieven van overheden die vanwege de Archiefwet worden gearchiveerd.

Specifieke aanleiding voor de discussie over het archiefperspectief is het e-Depot van het Nationaal Archief. De basis voor het metadatamodel van het Nationaal Archief wordt gevormd door twee generieke 'toepassingsprofielen' die op dit moment worden gebruikt door de diverse overheden: het Toepassingsprofiel Metagegevens Rijksoverheid (TPR) en het Toepassingsprofiel Metagegevens Lokale Overheden (TMLO). Omdat het e-Depot van het Nationaal Archief gebruikt kan worden door alle overheden kan het de metadata van beide toepassingsprofielen verwerken in een standaard XML-formaat (TopX 2.3).²⁰ Als Groninger Archieven in de toekomst gebruik wil gaan maken van het e-Depot of wanneer andere overheden met een webarchief dit willen, moest er goed worden gekeken in hoeverre de keuze voor MODS-lite dit negatief zou kunnen beïnvloeden.

Hiertoe is ten eerste door het Nationaal Archief een mapping gemaakt van MODS-lite door Groninger Archieven naar TopX 2.3. Daarna zijn er gesprekken gevoerd om de resultaten door te spreken.

Conclusie

- Het maken van de mapping van MODS Lite naar TopX 2.3 bleek goed mogelijk. MODS Lite is wel rijker, wat betekent dat er bij een mapping enige metadata zal worden 'platgeslagen'.
- Het kan voorkomen dat een webarchief een losstaande collectie is, en/of dat individuele gearchiveerde websites onderdeel kunnen zijn van een inventaris (zoals bij het Nationaal Archief). In die gevallen moeten die ook direct in de inventarisbeschrijving worden opgenomen. Groninger Archieven heeft een minimaal aantal elementen gekozen van MODS Lite die we voor de webarchieven gebruiken die overeenkomen met wat zij gebruiken voor archiefbeschrijvingen, zodat een dergelijke koppeling mogelijk is.
- Er moet op landelijk niveau waar mogelijk afstemming komen voor metadatamodellen van webarchieven. Vele webarchiverende instellingen --zoals ook Groninger Archieven-- laten de archivering uitvoeren door een externe partij, waardoor deze over het algemeen geen onderdeel zijn van de eigen collectiebeheersystemen.
- Het is in TopX 2.3 niet mogelijk om een onderscheid te maken binnen het element 'Omschrijving', terwijl drie MODS Lite elementen hier naar te mappen zijn (Abstract, TableofContents, Subject). Een optie is om hier te werken met zogeheten qualifiers of meer specifieke velden, maar vooralsnog is dit een plat veld in TopX 2.3. Hoewel TopX bedoeld is voor opname van bestanden in het e-Depot van het Nationaal Archief en niet als eindgebruikermodel is dit zowel voor webarchieven als andere archieven een mogelijke beperking. Zeker omdat juist in 'Subject' in de uitwerking van MODS-Lite bij Groninger Archieven de GTAA Onderwerps-as trefwoorden te vinden zijn.
- Hierop aansluitend: een metadatamodel voor het beschrijven en beheren van een webarchief is nog niet een model dat gericht is op eindgebruikers. Wanneer webarchieven beschikbaar worden gesteld voor gebruikers kunnen er andere requirements ontstaan voor het presenteren en verrijken van metadata. [Dooley et al., 2017] [Baltussen et al., 2014]

Het beschrijven van webarchieven volgens vaste standaarden staat zowel in Nederland als wereldwijd nog in de kinderschoenen. [Dooley et al., 2017] [Webster, 2016] Binnen dit project is een verkenning gemaakt van de mogelijkheden en valkuilen. De MODS-Lite implementatie bij Groninger Archieven gaat voor nu uit van een relatief eenvoudige basisset aan velden die wanneer nodig kan worden uitgebreid, te weten:

²⁰ <http://www.nationaalarchief.nl/digitaal-archiveren/informatie-archiveren-het-e-depot/metadata-het-e-depot>

- Beschrijving
- Titel
- URL: website live
- URL: website archief
- Verantwoordelijke
- Onderwerp
- Geografisch trefwoord
- Type
- Datum: harvest
- Datum: publicatie
- GUID

Voor een specificatie van de implementatie en opmerkingen zie 1 - MODS-Lite velden geïmplementeerd bij Groninger Archieven. Belangrijk om hierbij aan te tekenen is dat deze velden zijn gekozen vanuit een erfgoedperspectief. Het webarchief van Groninger Archieven is vanuit dit perspectief aangelegd. Wanneer websites worden gearchiveerd door overheden vanuit de Archiefwet zijn met name extra / meer specifieke velden waarin wijzigingen van titels en URLs worden vastgelegd van groot belang.

Aanbeveling

- Afhankelijk van de aanleiding en frequentie van archiveren kan het raadzaam zijn om bepaalde velden toe te voegen of af te laten vallen ten opzichte van de implementatie in dit project.
- Wanneer er wensen binnen je organisatie zijn om in de toekomst over te stappen naar het gezamenlijke e-depot van het Nationaal Archief, houdt dan bij de implementatie van het metadatamodel van de webarchief rekening mee dat deze aansluit op TopX 2.3.
- De uitkomsten van dit project zullen dan ook worden overgedragen aan het recentelijk door de NCDD opgerichte Kennisplatform Webarchivering²¹ (waar de klankbordleden van dit project ook zitten in nemen), om te borgen dat de discussie niet eindigt met dit project, maar verder wordt gevoerd in een nationale context. Meld je aan op dit platform om op de hoogte te blijven van deze en andere ontwikkelingen binnen het Nederlandse webarchiveringsdomein.
- In het geval van Groninger Archieven is het webarchief in eerste instantie aangelegd als een aparte collectie, vanuit een erfgoedperspectief. Er zitten echter ook websites in de collectie van organisaties waarvan reeds inventarissen bestaan. De volgende stap zal zijn om deze websites ook onderdeel te maken van deze inventarissen.

NB: vlak voor het opleveren van dit rapport is er door de OCLC Web Archiving Metadata working group een draft rapport [Dooley, 2017] gepubliceerd met aanbevelingen voor metadatering van webarchieven. Deze zijn niet meegenomen in deze evaluatie. Echter, in het rapport staat onder meer een mapping van de aanbevelingen naar MODS velden. Wanneer OCLC het definitieve rapport publiceert, dient dit zeker geraadpleegd te worden door organisaties die een metadastandaard zoeken voor hun webarchief.

²¹ <https://informatie2020.pleio.nl/groups/profile/48637242/kennisplatform-webarchivering>

Projectresultaat 2: Implementatie GTAA-Onderwerpsas in MAIS-Flexis 9

Binnen dit project is EEN praktische use case uitgewerkt, om meer inzicht te krijgen in wat er nodig is om een nationaal termennetwerk op te zetten, waarbij erfgoedinstellingen gebruik kunnen maken van elkaars thesauri. Enerzijds maakt dit netwerk het mogelijk voor instellingen om bestaande thesauri te gebruiken, zodat de interne resources die nodig zijn voor het beheer beperkt worden. Door dit gedeelde gebruik én doordat thesauri onderling ook weer verbonden worden, ontstaan er relaties tussen verschillende en uiteenlopende collecties, doordat zij gebruik maken van dezelfde annotatiebronnen.

In dit project is de GTAA-Onderwerpsas van de thesaurus van Beeld en Geluid geïmplementeerd bij Groninger Archieven. Voor dit gerealiseerd werd zijn samen met leverancier DE REE en het Nederlandse erfgoedveld technische en inhoudelijke verkenningen gevoerd. Dit vond plaats in het begin van het project (december 2015-maart 2016). Vervolgens is overgegaan tot implementatie van de GTAA-Onderwerpsas en het metadatamodel voor het webarchief.

Deze implementatie en de implicaties voor het opzetten van een grootschalig termennetwerk in nationale context zijn vervolgens geëvalueerd middels gebruikersonderzoeken en discussiesessies. Deze evaluatie vond plaats in oktober-november 2016.

Hieronder volgt eerst een overzicht van de technische implementatie van de GTAA-Onderwerpsas in MAIS-Flexis 9 en worden vervolgens de resultaten van de evaluatie uitgelicht.

Over de GTAA²²

Het Nederlands Instituut voor Beeld en Geluid ontwikkelde met verschillende Nederlandse organisaties die audiovisueel erfgoed beheren de Gemeenschappelijke Thesaurus voor Audiovisuele Archieven (GTAA). De GTAA wordt gebruikt voor het inhoudelijk annoteren van historisch en actueel materiaal (bij Beeld en Geluid), met name audiovisueel materiaal). De thesaurus wordt ingezet in het handmatige beschrijvingsproces van de omroepen en Beeld en Geluid en voor automatische annotatietechnieken.

De GTAA is een 'meer-assige' thesaurus. Dit houdt in dat de termen (concepten) zijn ingedeeld in meerdere categorieën of assen, ook wel concept schema's genoemd. De categorieën van termen die voor het toegankelijk maken van het materiaal gebruikt worden, zijn:

- Onderwerpen, of wel begrippen (bijv. zwemmen, daklozen, binnenvaartschepen, audiovisuele archieven). Een variant op de as Onderwerpen is in gebruik voor het beschrijven van materiaal op het niveau van shots en geluiden, de as Onderwerpen beeld-geluid.
- Persoonsnamen, dit zijn eigennamen van personen die als inhoudelijk element voorkomen in een productie, of als Spreker, Cast-lid of Uitvoerende kunnen voorkomen. Voorheen bestond er nog een aparte as Makers, met persoonsnamen van makers en medewerkers aan de totstandkoming van audiovisuele producties. Deze is echter samengevoegd met Persoonsnamen.
- Namen, eigennamen van o.a. organisaties, muziekensembles, bevolkingsgroepen,

²² <http://gtaa.beeldengeluid.nl/>

oorlogen, verdragen, feest(dag)en, prijzen, producten, en titels van o.a. tijdschriften, boeken, films, radio- en televisieprogramma's (bijvoorbeeld: KLM, Hindoestanen, Sinterklaas, Golfoorlog, De aanslag).

- Geografische namen, namen van steden, dorpen, provincies, snelwegen, landen, etc.
- Classificatie, dit is een rubrieksindeling bij de schema's Onderwerpen en Onderwerpen beeld-geluid, die het systematisch zoeken op onderwerpen, ook van shots en quotes, mogelijk maakt.
- Genres, typering van een audiovisuele productie als geheel. Genre termen kenmerken de vorm, het filmgenre, de stroming of het type productie als geheel. Voorbeelden: actualiteiten, praatprogramma, magazine.

Binnen dit project is de Onderwerpsas gebruikt. Deze bevat iets meer dan 4.000 termen, waarvan er 846 top-level concepten zijn.²³ De onderwerpen zijn echter wel polyhiërarchisch georganiseerd. Dat wil zeggen dat één term aan meerdere top-level concepten gekoppeld kan zijn.²⁴

Alle termen in de Onderwerps-as zijn weer gekoppeld aan een bovenliggende Classificatie-as (de rubrieksindeling). Deze bestaat uit 100 concepten, waarvan 19 topConcepts^{25, 26}. De Classificatie-as is maximaal twee termen diep.

Inhoudelijke vraagstukken

Inhoudelijke aansluiting

Voor de start van het project is eerst op basis van een steekproef gekeken in hoeverre de Onderwerps-as van Beeld en Geluid inhoudelijk aansluit bij het webarchief van Groninger Archieven. Hiertoe zijn een zestal gearchiveerde websites met uiteenslopende onderwerpen (musea, lokale loopgroepen, kerken, buurtkroegen) bekeken en zijn er door een inhoudelijk expert van de GTAA minimaal drie relevante onderwerpstermen bij iedere website gezocht. Dit is vervolgens voorgelegd aan Groninger Archieven. De matches boden genoeg dekking, waarna het project werd gestart.

Van hiërarchie naar graph

Vervolgens zijn er verschillende gesprekken gevoerd met Groninger Archieven, DE REE en Beeld en Geluid over verdere inhoudelijke consequenties van het gebruiken van een thesaurus van een derde partij door Groninger Archieven.

Een belangrijk discussiepunt was de structuur. Een traditionele thesaurus heeft een hiërarchische boomstructuur, waarbij op het hoogste niveau termen liggen, die in subniveaus verder en daarmee specifiekere worden onderverdeeld.

Echter, zodra er gebruik wordt gemaakt van een thesaurus die via OpenSKOS of een ander Linked Data protocol beschikbaar wordt gesteld, verandert de boomstructuur naar een 'graph', een web van concepten (nodes) die door middel van diverse soorten relaties en gedefinieerde velden aan elkaar

²³ De Onderwerpsas is (vooralsnog) maximaal 5 niveaus diep (eg: wetenschap -> maatschappijwetenschappen -> recht -> auteursrecht -> portretrecht.).

²⁴ Dit komt in totaal 159 voor (vier termen hebben 3 broaderTerms, 155 termen hebben 2 broaderTerms).

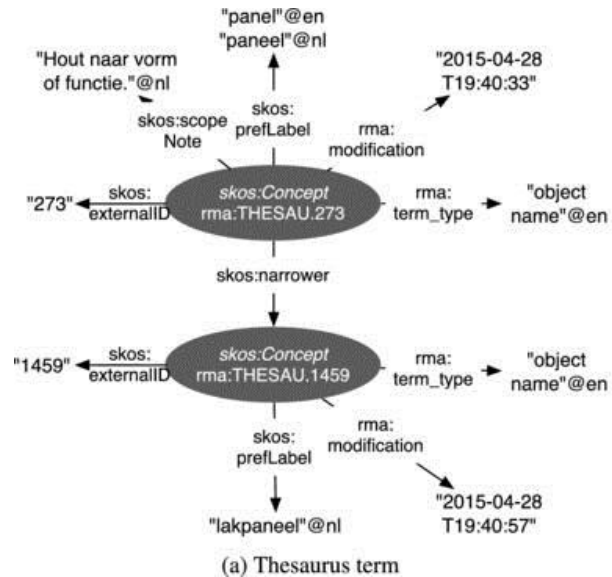
²⁵ 00X algemeen, 01L levensbeschouwingen, 02M maatschappij, 03B bestuur en geschiedenis, 04R ruimtelijke omgeving, 05E economie, 06G gezondheid, 07O opvoeding en onderwijs, 08W wetenschap, 09T techniek, 10V verkeer en vervoer, 11C communicatie en media, 12K kunst en cultuur, 13S sport en vrijetijdsbesteding, 14N natuur en milieu, 15A aardrijkskunde, 17B beeld/geluid, 18C concepts BBC Motion Gallery, 19R references BBC Motion Gallery.

²⁶ De Classificatie-as is niet polyhiërarchisch.

gerelateerd zijn.

Organisms

- NT1 Eukaryotes
- NT2 Animals
 - NT3 Aquatic animals
 - NT4 Aquatic mammals
 - NT5 Marine mammals
 - NT6 Whales
 - NT7 Ba een whales
 - NT8 Blue whale
 - NT8 Gray whale
 - NT8 Humpback whale
 - NT8 Minke whale
 - NT8 Right whales
 - NT9 Bowhead whale
 - NT9 Northern right whale
 - NT9 Southern right whale
 - NT7 Fossil whales
 - NT7 Toothed whales
 - NT8 Beaked whales
 - NT8 Beluga (Whale)
 - NT8 Dolphins (Mammals)
 - NT9 Bottlenosed dolphins
 - NT9 Killer whales
 - NT9 Pilot whales
 - NT9 River dolphins
 - NT8 Narwhal
 - NT8 Porpoises
 - NT8 Sperm whale



Afbeelding 2: Voorbeeld van een traditionele thesaurusindeling van een biologische taxonomie van organismes (links), en een 'graph' met nodes in een Linked Data thesaurus (rechts).

Dit zorgde bij de start van het project tot enige verwarring, aangezien het nog niet geheel duidelijk was hoe het graph model zich verhiel tot de standaard thesaurusmodule van MAIS-Flexis, en hoe de data van de GTAA ingelezen, opgeslagen en vervolgens beschikbaar moest worden gesteld. Om dit helder te krijgen waren meerdere gesprekken nodig, waarin de use case voor het gebruik van Linked Data aan de hand van praktische voorbeelden en workflows moest worden doorgenomen. Hiervoor was de bestaande documentatie van de GTAA en OpenSKOS nog niet voldoende. Daarbij hielp het om de online doorzoekbare en doorbrowsable versie van de GTAA te tonen.

Deze is in principe niet publiek beschikbaar, maar hielp om de samenhang tussen termen en de structuur inzichtelijk te maken. Hiermee werd ook de use case voor het inzetten van de samenhang tussen termen bij het zoeken door eindgebruikers duidelijk.

Controle over het toevoegen van concepten aan de GTAA

Naast de discussie over de structuur en de toepassing van Linked Data duurde het enige tijd voordat er consensus was over het mogelijke toevoegen van termen aan de GTAA door Groninger Archieven.

De meest ingrijpende oplossing zou inhouden dat zodra een documentalist van Groninger Archieven een concept toevoegt dat nog niet in de Onderwerps-as staat, dit concept bij Beeld en Geluid in de beheersomgeving van OpenSKOS binnen zou komen. Vervolgens moet de relevantie van de term van beoordeeld worden door een inhoudelijke specialist, die bepaalt of het concept moet worden toegevoegd, al bestaat, een mogelijk altLabel of hiddenLabel is, of moet worden verwijderd.

Aangezien de collectie van Groninger Archieven inhoudelijk zeer anders is dan die van Beeld en Geluid en omdat de Onderwerps-as generieke, high-level onderwerpen bevat is ervoor gekozen om deze kandidaatstermworkflow niet in te regelen. Wanneer Groninger Archieven een grote omissie tegenkomt, kan dit gemeld worden via een speciaal emailadres dat door de GTAA redactie wordt

beheerd.

Aanbeveling

- Maak binnen je organisatie de stap van hiërarchisch denken naar Linked Data denken. Veel erfgoedinstellingen die thesauri gebruiken denken logischerwijs vanuit de hiërarchie van de thesaurus. Linked Data thesauri hebben geen klassieke boomstructuur hiërarchie, maar hebben nodes die in een 'graph' verbonden zijn. Deze conceptuele omslag kan zeker in het begin van een dergelijk traject lastig zijn voor mensen die nog geen kennis hebben van Linked Data.
- Zorg voor goede documentatie over je thesaurus: de grootte van de diverse assen, de samenhang tussen termen, het beheer en de use case voor het gebruiken van de thesaurus door een derde partij.
- Voor gebruikers (zowel leveranciers als erfgoedinstellingen) is er goede, degelijke documentatie nodig van de beoogde of gebruikte thesaurus. Denk hierbij aan technische specificaties of het aanbieden van een doorzoekbare/browsebare versie van je hele thesaurus.
- Bepaal in overleg met de aanbieder van de thesaurus hoe eventuele nieuwe concepten kunnen worden voorgesteld. Identificeer of dit een volledig technisch ingeregelde workflow wordt, of dat dit via andere wegen (eg email) kan worden gedaan.

Technische implementatie: mogelijkheden en gemaakte keuzes

Concepten aanroepen: live API implementatie vs een OAI-PMH harvest

De eerste belangrijke keuze was hoe de concepten van de GTAA-Onderwerpsas technisch aangeroepen zouden worden. Hierbij waren de opties:

- de OPENSKOS API²⁷ gebruiken;
- een OAI-PMH harvest maken van alle termen, om deze vervolgens lokaal op te slaan en periodiek bij te werken.

Om de voor- en nadelen van deze opties helder te krijgen zijn de volgende vragen voorgelegd aan Beeld en Geluid, Groninger Archieven en DE REE:

- Hoe groot is de GTAA-Onderwerpsas en wat is het datamodel?
- Hoe veranderlijk is de GTAA-Onderwerpsas?
- Hoe moet data worden gesynchroniseerd als je gebruik maakt van de live API vs een OAI-PMH harvest?
- Hoe groot is de kans dat de API niet (goed) bereikbaar is?
- In hoeverre wil Groninger Archieven termen aanpassen en toevoegen?
- Voor welke implementatie hebben andere gebruikers van de GTAA gekozen en waarom?

Uiteindelijk bleken beide oplossingen hun voor- en nadelen te hebben.

Voordeel API

- Relatief weinig onderhoud, aangezien je geen eigen, lokale database maakt van de concepten.
- De API kent weinig technische problemen en de servers vrijwel altijd beschikbaar (weinig 'downtime').

²⁷ <http://openskos.beeldengeluid.nl/api>

- Afnemer werkt altijd met de actuele versie van de GTAA.

Nadeel API

- Je wordt als leverancier (in dit geval DE REE) afhankelijk van een dienst van een derde partij.
- Bij grootschalig gebruik van de API kunnen dermate veel calls worden gemaakt naar de server dat dit de performance negatief beïnvloedt. Dit brengt extra inzet en monitoring aan de kant van de aanbieder (in dit geval Beeld en Geluid) met zich mee.

Voordeel OAI-PMH harvest

- Als leverancier heb je zelf controle over de inrichting en de wijze van doorzoekbaar maken van de concepten, doordat je je eigen lokale database aanlegt.
- Deze oplossing reduceert het aantal calls naar de API.

Nadeel OAI-PMH harvest

- Extra werk aan de kant van de leverancier, omdat deze een lokale instantie moet inrichten en de concepten zelf moet indexeren.
- Wanneer meer klanten gebruik willen maken van de thesaurus, kan dit --afhankelijk van de technische implementatie-- betekenen dat er per klant een lokale database wordt geharvest, die per stuk moeten worden bijgehouden.²⁸

De Nederlandse Publieke Omroep maakt gebruik van de GTAA via de API. Zij gebruikt de Persoonsnamen-as van de GTAA om uitzendingen te annoteren, die vervolgens opgenomen worden in het archief van Beeld en Geluid.²⁹ Voor deze laatste use case is het gebruik van een API een betere keuze, aangezien het hier om een zeer grote as gaat binnen de GTAA waar veelvuldig nieuwe termen aan worden toegevoegd. Er kunnen dan ook eventuele timing issues ontstaan als er niet met de meest recente versie wordt gewerkt. Vanwege de snelle en constante werkprocessen bij de omroep is in dat geval een API een logischer keuze.

Leverancier DeventiT werkt ook met live APIs van diverse thesauri³⁰, en heeft een lokaal caching mechanisme voor termen. Hierdoor zijn ze minder afhankelijk van de uptime van de API. In het collectiebeheersysteem van Beeld en Geluid staat een lokale versie van de GTAA. Elke nacht wordt deze bijgewerkt met een harvest van de laatste wijzigingen in OpenSKOS. Dit is echter voor de mate van het gebruik bij Beeld en Geluid niet optimaal, omdat hierdoor de lokale versie van de GTAA niet gegarandeerd hetzelfde is als de live versie.

Er zijn dus diverse opties met ieder zijn voor- en nadelen. De GTAA-Onderwerpsas verandert relatief weinig én de mate van gebruik door Groninger Archieven is meer beperkt. Verder wilde DE REE controle hebben over opslag en presentatie van data bij de autocompletesuggesties, en betrof het project een pilot. Daarom is er binnen dit project gekozen voor een lokale OAI-PMH harvest.³¹

Aanbeveling

- Er bestaat niet één voorkeursscenario voor het inzetten van externe thesauri. Het zal bij het verder ontwikkelen van een grootschalig nationaal termennetwerk noodzakelijk zijn om goede praktische voorbeelden te geven van mogelijke scenario's en de overwegingen en lessons-learned die hieruit zijn voortgekomen.
- Bepaal eerst de specifieke use case(s) van je instelling. Werk vervolgens de verschillende voor- en nadelen van diverse functionele en technische scenario's uit.

²⁸ Dit is ook het geval bij de wijze waarop DE REE de GTAA nu in MAIS-Flexis 9 heeft geïmplementeerd.

²⁹ <https://www.beeldengeluid.nl/blogs/research-and-development/201403/beeld-en-geluid-thesaurus-ingeplugd/>

³⁰ <http://www.deventit.nl/atlantis-wordt-steeds-dynamischer/>

³¹ Deze functionaliteit is ingebouwd in MAIS-Flexis 9 en niet in MAIS-Flexis 8. Reden hiervoor is dat het eenvoudiger is om een web-based service als OpenSKOS in te bouwen in de flexibele webversie van MAIS-Flexis (versie 9).

Concepten annoteren: wat laat je zien, wat sla je op?

Een belangrijke, specifieke vraag voor prettig gebruik van een thesaurus is, welke gegevens over een concept/term getoond moeten worden tijdens het annoteren. Met andere woorden: wat heb je als documentalist nodig om de relevantie van een concept voor het record dat je beschrijft te bepalen?



Afbeelding 3: voorbeeld van het autocomplete-overzicht van de GTAA in MAIS-9 bij het invoeren van de eerste drie letters van de beoogde term 'verzet'.

Dit is afhankelijk van de thesaurus en de thesaurus-as die je gebruikt. Bij de as persoonsnamen moeten de velden 'Voornaam', 'Tussenvoegsel' en 'Achternaam' zichtbaar zijn, maar ook eventuele alternatieve namen, terwijl het bij onderwerpen prettig is om naar de voorkeursterm ook de alternatieve term te zien.

De beschikbare gegevens over een concept en zijn relaties staan in de OpenSKOS-documentatie van de GTAA³². Hieronder zal op hoofdlijnen ingegaan worden op de gemaakte keuzes voor de Onderwerps-as binnen dit project.

Concepten binnenhalen: relevante velden

Zoals hierboven is beschreven is er in dit project gekozen voor een lokale harvest van de Onderwerps-as via OAI-PMH. Wanneer je deze optie gebruikt, komen alle beschikbare informatie en relaties van een term hangen binnen. Vervolgens kun je desgewenst zelf een selectie maken binnen je eigen database-systeem om alleen hetgeen op te slaan dat voor jou relevant is. Als je gebruik maakt van de API kun je ook zelf bepalen welke velden je wilt tonen, maar loopt dit via een ander technisch proces.

Binnen dit project lag de focus op de volgende velden van de GTAA Onderwerps-as:

Veld	Beschrijving	Voorbeeld
URI	De unieke resource identifier van het concept. Deze is stabiel en uniek.	http://data.beeldengeluid.nl/gtaa/221033
pref	Voorkeurslabel van het concept	3D-techniek
uuid	In het zeer uitzonderlijke geval dat een concept wordt gedelete, dan verdwijnt de URI, maar niet de UUID.	0a39cbfc-a491-75f8-60b2-6b50418e299d

³² <http://openskos.beeldengeluid.nl/>

related	Gerelateerde term. Kan 0, 1 of meerdere keren voorkomen.	http://data.beeldengeluid.nl/gtaa/25521
scopeNote	Aanvullende informatie over de term.	NB: geen inhoudelijke onderwerpsterm
broader	Verwijzing met een URI naar de term die hiërarchisch direct boven een term ligt	http://data.beeldengeluid.nl/gtaa/26907
narrower	Verwijzing met een URI naar de term die hiërarchisch direct onder een term ligt	http://data.beeldengeluid.nl/gtaa/26713
altLabel	Alternatief label van de term	televisietorens (altLabel van zendmasten)
hiddenLabel	e.g. veelgemaakte spelfouten of alternatieve schrijfwijzen. Gebruikt om via zoeken bij het juiste prefLabel te komen.	kieviten (hiddenLabel van kievieten)
inScheme	Geeft aan in welk deel van de GTAA de term zit (in dit geval GTAA Onderwerps-as).	http://data.beeldengeluid.nl/gtaa/Onderwerpen

Tabel 1: overzicht van de gegevens die per concept gebruikt zijn binnen de implementatie van de GTAA in de MAIS-Flexis 9-omgeving van Groninger Archieven

Zoekresultaten duiden: context is everything

Bovenstaande velden geven de benodigde context van een concept weer. Dit is belangrijk, omdat onderwerpen multi-interpretabel kunnen zijn. Bijvoorbeeld: ‘bank’ is zowel een meubel waar je op kunt zitten als een financiële instelling.

Wanneer een documentalist zoekt naar een onderwerp, kunnen velden zoals scope notes, algemenere en specifiekere termen helpen bij het bepalen wat het juiste concept is. Daarnaast kan het tonen van dergelijke context inspireren tot het toevoegen van beter passende en/of meer termen.

Om te bepalen welke velden het beste getoond konden worden, werd een statistische analyse gemaakt van de ruim 4.000 concepten in de GTAA Onderwerps-as:

- Scope notes: aanwezig bij 12% van de concepten.
- Algemeneren termen: 78% van de concepten heeft 1 of meer algemenere term.
- Specifiekere termen: 18% van de concepten heeft 1 of meer specifiekere term. 5,5% heeft er meer dan 5. De term met de meeste specifiekere termen heeft er maar liefst 56.
- Gerelateerde termen: 75% van de concepten heeft 1 of meer gerelateerde term. 5% heeft er 10 of meer.
- Alternatieve labels: 30% van de concepten heeft 1 of meer alternatieve termen.

Het was lastig van tevoren in te schatten in hoeverre het tonen van de context van en relaties tussen concepten zouden helpen bij het interpreteren van de toe te voegen termen, of juist zouden zorgen voor een ‘information overload’ van documentalisten. Daarom werd er besloten om in de eerste versie alle gegevens en relaties te laten zien en vervolgens te testen wat de beste oplossing was. Uit deze tests bleek echter dat deze informatie erg gewaardeerd werd, zeker aangezien de algemenere, specifiekere en gerelateerde termen aangeklikt kunnen worden en vervolgens zeer eenvoudig kunnen

worden toegevoegd wanneer de documentalist deze (ook) relevant vindt.

Het veld hiddenLabel wordt niet getoond, maar wordt wel gebruikt om 'onder water' te zoeken. Voorbeeld: als een documentalist per ongeluk 'kieviten' intypt, wordt enkel het preLabel 'kievieten' en de andere bijbehorende velden getoond.

Beschrijven	
Titel *	Het Groninger Landschap
Onderwerpen (GTAA)	kieviten
Verantwoordelijke	kievieten (bron) algemener: weidevogels steltlopers

Afbeelding 4: voorbeeld van het 'onder water' doorzoekbaar maken van hiddenLabels.

Het autocomplete-overzicht

Hoe zien al deze velden er nu uit wanneer er door de Onderwerps-as van de GTAA wordt gezocht binnen MAIS-Flexis 9? In onderstaand voorbeeld heeft de documentalist de letters 'art' ingevuld om een record van de gecrawilde website van de David Bowie tentoonstelling in het Groninger Museum te annoteren. De documentalist ziet de relevante concepten verschijnen voor de zoekvraag 'art' in het autocomplete-overzicht.³³

Beschrijven	
Titel *	David Bowie Groningen
Onderwerpen (GTAA)	popmuziek tentoonstellingen art
Verantwoordelijke	artillerie (bron) algemener: landmacht
Type	artiesten , alternatieven: kunstenaars (bron)
Datering	specifieker: dompteurs imitators acteurs buikspekers cabaretiers majorettes straatartiesten jongleurs clowns komieken vuurvetdansers slangenbezoekers goochelaars algemener: artistieke beroepen
Datum harvest	algemener: artistieke beroepen
Samenvatting	gerelateerd: musici diskjockeys idolen zangers variété audities impresariaat fans fakirs
URL website live	artistieke beroepen , alternatieven: kunstenaars (bron) specifieker: musici diskjockeys artiesten architecten producenten beeldend kunstenaars choreografen muziekproducers decorontmodeontwerpers regisseurs grafisch kunstenaars fotografen schrijvers
URL website archief	algemener: beroepen gerelateerd: amateurs juweliers vrije beroepen kunst edelsmeden auteursrecht ambachten dienstverlenende beroepen

Afbeelding 5: voorbeeld van het autocomplete-overzicht van de GTAA in MAIS-9 bij het invoeren van de eerste drie letters van de beoogde term 'artiesten'.

De verschillende labels

De voorkeurstermen zijn te zien in **bold**. Het alternatieve label staat achter de voorkeursterm. De specifiekere, algemenere en gerelateerde termen staan onder de voorkeursterm. De scope note staat hier weer onder in *italics*. De documentalist kan ook klikken op specifiekere, algemenere en gerelateerde termen, om de informatie hierover op te roepen en ze daarna direct toe te kennen.

³³ De sortering van de resultaten kan worden ingesteld binnen in de module Applicatiebeheer van MAIS-9.

Onderwerpen (GTAA)	artiesten popm
Verantwoordelijke	popmuziek , alternatieven: <i>beatmuziek</i> (bron)
Type	specifieker: <i>hiphop rap dance punkmuziek blues r&b country & western rock-'n-roll hardrock reggae rhythm & blues folk soul</i> algemener: <i>muziekstromingen</i>
Datering	gerelateerd: <i>diskjockeys hitparades jukeboxes popfestivals bandleiders videoclip bands</i>
Datum harvest	notitie: <i>in ruimere zin te gebruiken voor populaire, lichte muziek</i>

Afbeelding 6: voorbeeld van een scope note bij de term 'popmuziek', die de documentalist meer context over een concept geeft bij het annoteren.

De URI: terug naar de bron

De documentalist kan op de link '(bron)' klikken om alle velden van een concept te zien. Deze link is de URI van het concept.

Concepten kiezen en koppelen aan een record

De documentalist kiest een concept door erop te klikken in de autocomplete lijst. Vervolgens wordt een koppeling gelegd tussen dit record en de onderliggende database van opgeslagen termen. Concreet wordt de volgende informatie van een GTAA-term aan het record gekoppeld (zie ook **Tabel 1**):

- **prefLabel**: dit is handig voor een documentalist om te kunnen zien wat er al is toegevoegd.
- **URI**: De unieke identifier op het web van een concept. Dit is het koppelvlak tussen de interne database en de meest recente versie van de GTAA Onderwerps-as. Wanneer er wijzigingen zijn van een concept, wordt de nieuwste informatie middels de URI opgehaald en opgeslagen in de database. Zo hoeft deze informatie niet per term, per record te worden gewijzigd, maar slechts op één plek.
- **UUID**: in het uitzonderlijke geval dat een concept wordt verwijderd (en daarmee ook de URI), kan de UUID altijd nog worden gebruikt om de oorspronkelijke term op te zoeken. Bij de records waar de verwijderde term reeds was toegevoegd blijft deze echter wel staan. De term kan dan echter niet nog aan andere records worden toegevoegd.

Aanbeveling

- Om de documentalist te helpen bij zijn of haar taak is het tonen van veel informatie over onderwerpen in het autocomplete-overzicht prettig. Voor onderwerpen zijn de volgende velden minimaal relevant: `prefLabel`, `broaderTerm`, `narrowerTerm`, `relatedTerm`, `altLabel`, `scopeNote`.
- Maak de `broaderTerm`, `narrowerTerm`, `relatedTerm` aanklikbaar in je autocomplete-overzicht, zodat de documentalist direct andere termen kan bekijken en toevoegen als dit gewenst is.
- Wat je toont en hoe je dit toont kan verschillen per type thesaurus. Persoonsnamen kunnen bijvoorbeeld andere velden hebben dan onderwerpen die bij autocomplete de benodigde context geven (eg: beroep, geboorte- en sterfdatum).
- Sla het `prefLabel`, de URI en de UUID op bij het record wanneer je een term uit een externe thesaurus toevoegt. Dit wordt tevens gezien als 'best practice' door het NIOD (verantwoordelijk voor de ontwikkeling van de WOII thesaurus³⁴) en door DEVENTit, die in Atlantic meerdere externe thesauri heeft geïmplementeerd.

Omgaan met dynamiek: veranderende en nieuwe termen

Naast het vraagstuk van het al dan niet tonen van alle mogelijke relevante velden bij het annoteren, is er veel discussie gevoerd over hoe om te gaan met veranderingen binnen concepten enerzijds, en de workflow voor het voorstellen van nieuwe termen anderzijds.

³⁴ <http://www.netwerkoorsbronnen.nl/wo2-thesaurus>

Omgaan met veranderende en verwijderde concepten

Wanneer je als (erfgoed)instelling gebruikt maakt van thesauri van derden is het van groot belang dat deze zowel inhoudelijk als technisch goed en deugdelijk onderhouden wordt zijn en dat concepten in principe niet verwijderd worden, aangezien je zelf niet verantwoordelijk bent voor redactie en onderhoud. Veelgebruikte, open thesauri die door erfgoedinstellingen of andere betrouwbare partijen worden aangeboden (eg: GTAA, GeoNames, Erfgoedthesaurus, AAT) zijn in principe zowel technisch als inhoudelijk degelijk. Daarbij is het van belang dat er een duidelijk contactpunt is voor vragen.

Hoewel concepten idealiter niet verwijderd worden, is een thesaurus wel een levend geheel. Zo kan het voorkomen dat een scope note wordt toegevoegd of aangepast, of dat er een hiddenLabel of relatedTerm aan een concept wordt toegevoegd. Een concept zelf kan ook aan verandering onderhevig zijn. Hoewel het uitgangspunt van een thesaurus is dat concepten inhoudelijk niet wijzigen, kan het wel voorkomen dat de benoeming van het prefLabel wordt aangepast, ze als altLabel opgaan in andere of zelfs verwijderd worden. Die veranderlijkheid is bij inhoudelijk deugdelijk beheer geen risico, maar juist een verrijking en verbetering.

Hoe ga je als eindgebruiker om met dergelijke veranderingen waar je zelf geen redactionele controle over hebt? Dit kan implicaties hebben voor de records waar deze aan zijn toegevoegd. Hoewel via de URI van een concept altijd de meest recente informatie kan worden opgevraagd, is dit bij een verwijderd concept niet mogelijk, aangezien de URI dan ook verdwijnt. Een dergelijke verwijdering van een concept komt bij de GTAA (en thesauri in het algemeen) echter niet tot nauwelijks voor.

Om dit te kunnen doen is het echter noodzakelijk om dergelijke substantiële wijzigingen te monitoren. Binnen OpenSKOS is er een 'changes feed' waarmee kan worden opgeroepen welke termen zijn veranderd. Op basis hiervan wordt de lokale database van DE REE bijgewerkt. Echter, wat er precies verandert binnen een concept en waarom is vooralsnog niet direct zichtbaar in de changes feed, alleen dat er een wijziging is in een concept .

Aanbeveling

- Er moeten best practices worden ontwikkeld voor het monitoren van wijzigingen in externe thesauri door eindgebruikers. De aanpassing aan OpenSKOS (versie 2.1, SKOSXL), waarbij ook URIs worden gekoppeld aan eg altLabels en hiddenLabels is een stap in de goede richting. Hiermee kan veel specifiekere een rapport worden gedraaid van wat er binnen een concept exact is gewijzigd.³⁵
- Maak --zeker binnen het door NDE beoogde termennetwerk-- afspraken over het wijzigen en met name verwijderen van concepten. Bij de GTAA komt dit zoals gezegd in principe nauwelijks voor, maar wellicht zijn andere thesauri (nog) niet zo stabiel.
- Voeg altijd, zoals bij de GTAA het geval is, de UUID van een concept toe aan een record. Deze kun je net als de URI gebruiken om een concept te 'resolven'. Het verschil met een URI is dat de UUID nog wel aanwezig is in de OAI-PMH changes feed als een concept wordt verwijderd. De UUID maakt het dus mogelijk om te zien dát een concept heeft bestaan, maar dat het verwijderd is. Hoe het concept er uitzag is echter niet meer te achterhalen bij de thesaurusaanbieder. Bij verwijdering van een concept kan de UUID in de changes feed wel worden gematcht met de UUID die lokaal is opgeslagen bij het record waar een concept eerder aan werd toegevoegd. Het zou beter zijn dat wanneer een concept wordt verwijderd, dat er een persistente plek is waar de inhoud wel zichtbaar blijft.

³⁵ Op moment van schrijven was deze versie in de acceptatiefase. De uitrol naar productie is gepland in Q3 2017.

Nieuwe termen voorstellen

Zoals eerder werd besproken maakt de Nederlandse Publieke Omroep (NPO) gebruik van de GTAA Persoonsnamen-as voor het annoteren van uitzendingen. Aangezien deze uitzendingen vervolgens direct instromen in het archief van Beeld en Geluid en de GTAA specifiek is ontwikkeld voor het annoteren van audiovisueel materiaal en uitzendingen van de Publieke Omroepen in het bijzonder, is het wenselijk dat de NPO zelf Persoonsnamen kan voorstellen als deze zich nog niet in de GTAA bevinden.

Kandidaatstermworkflow

Het toevoegen van termen verloopt via een zogenaamde kandidaatterm-workflow. Wanneer een documentalist van de NPO een persoonsnaam niet kan vinden in de GTAA die van toepassing is op een uitzending, wordt er een nieuwe term ingevoerd. Deze krijgt vervolgens een kandidaatsstatus en URI toegekend binnen de GTAA, waarna deze wordt beoordeeld door de GTAA-redactie. Een term kan worden geaccepteerd, verwezen naar een andere term, worden afgewezen of krijgt de status non-compliant. In het laatste geval wordt de term niet opgenomen in de GTAA, maar blijft deze wel bestaan en kan blijvend gebruikt worden door de NPO.

Waarom is hiervoor niet gekozen binnen dit project?

Voor dit project is ervoor gekozen om niet met bovenstaande kandidaattermworkflow te werken. De redenen hiervoor zijn:

- Onderwerpen \neq Named Entities: Onderwerpen zijn multi-interpretabel en qua toepassing afhankelijk van de achtergrond en voorkeuren van de invoerder. Named Entities verwijzen naar specifieke entiteiten, zoals personen, locaties en producten die uniek te onderscheiden en af te kaderen zijn.
- Een dergelijke workflow inregelen kost extra tijd en vervolgens onderhoud. Het is daarbij ook mogelijk om de redactie van de GTAA een email te sturen met voorstellen voor aanpassingen en wijzigingen.
- Het webarchief van Groninger Archieven is geen kerncollectie van Beeld en Geluid (waarbij uitzendingen van de NPO dit wel zijn).
- De Onderwerps-as is relatief klein, algemeen en stabiel en sloot al goed aan op het webarchief.
- Wanneer je gebruik maakt van externe thesauri accepteer je ook de autoriteit en expertise van anderen en betekent dit dat je zelf geen thesaurus hoeft te ontwikkelen en te onderhouden.
- Wanneer er echt belangrijke termen ontbreken die ook voor Beeld en Geluid zelf relevant zijn, is het altijd mogelijk om de GTAA-redactie te mailen om een voorstel te doen en via deze weg termen toe te voegen.

Zelf lokaal termen toevoegen?

Het is technisch wel mogelijk om zelf een term toe te voegen aan de lokale database harvest van de GTAA in Mais-Flexis. Een dergelijk lokaal toegevoegde term zou dan geen OpenSKOS URI krijgen, maar een eigen identifier en status binnen MAIS. Groninger Archieven en DE REE hebben hier echter van afgezien. Reden hiervoor is dat het niet hoeven aanleggen en beheren van een eigen thesaurus een grotere meerwaarde had dan het hebben van controle op de inhoud hiervan.

Namen-as

Eén probleem dat tijdens de gebruikerstests aan het licht kwam is dat ondanks de inhoudelijke compatibiliteitscheck enkele belangrijke termen niet in de Onderwerps-as staan, zoals 'Tweede Wereldoorlog' en 'Sinterklaas'. Deze bleken te staan in de Namen-as van de GTAA. Deze bevat eigennamen die een inhoudelijk element hebben, zoals onder meer namen van corporaties en andere termen die als eigennamen worden gezien, zoals feestdagen, bevolkingsgroepen, prijzen, radio- en televisieprogramma's, tijdschriften, boeken, films, oorlogen, verdragen, feestdagen en

fictieve figuren. Ook objecten die met een eigennaam worden weergegeven en geen geografische naam zijn, zoals bouwwerken, vervoermiddelen, bruggen en monumenten vallen onder de Namen-as. Kortom, er staan een boel termen in de Namen-as die ook een onderwerp kunnen zijn. Dit geldt overigens ook voor de termen in de assen Persoonsnamen en Geografische namen. Deze assen zijn uit elkaar getrokken, met de reden dat trefwoorden die een av-document inhoudelijk ontsluiten zijn onderverdeeld in de assen Onderwerpen; Persoonsnamen, Geografische namen en Namen.

De Namen-as is groter en anders ingericht dan de Onderwerps-as. Hij bevat ruim 30.000 concepten en er worden vaker concepten aan toegevoegd en wijzigingen in aangebracht (eg aanpassingen in scope notes, altLabels) dan aan de Onderwerps-as. Echter, de relaties zijn veel beperkter. Zo zijn er wel relatedTerms in de Namen-as, maar geen hiërarchische relaties (broader/narrowerTerms).

Aanbeveling

- Implementeer een kandidaattermworkflow alleen als hier een goede business case voor is, zowel aan de kant van de gebruikers als door de aanbieder.
- Named Entities zoals personen hebben een andere kwaliteit dan onderwerpen. Het zijn specifieke en uniek identificeerbare instanties. Onderwerpen zijn ‘fuzzier’ en ambigu, en de relevantie, naamgeving en hiërarchische relaties toevoegen van onderwerpen kan tot veel discussies leiden. De wijze waarop onderwerpen het beste een plek kunnen krijgen én onderhouden worden in een nationaal termennetwerk met gedeelde thesauri zal nog veel nader onderzoek behoeven.
- Er zitten concepten in de Namen-as die niet staan in de Onderwerps-as. Aangezien de Namen-as net als de Persoonsnamen en Geografische namen ‘Named Entities’ zijn, zou de meest logische optie zijn om deze alle toe te voegen voor onderwerpsnotatie, zodat ze kunnen worden toegekend als inhoudelijke term.
- Algemeener geldt dat vooraf zorgvuldig de inhoudelijke compatibiliteit en de reikwijdte en inhoud van de onderwerpsontsluiting van de afnemer in kaart te brengen zodat bepaald kan worden welke assen relevant zijn.
- Aanbieders van thesauri worden in toenemende mate een spil in een groter netwerk wanneer deze thesauri door derden worden gebruikt, en worden leverancier van een dienst. In de context van een grootschalig nationaal termennetwerk moet verder in kaart worden gebracht welke resources nodig zijn om vragen van gebruikers, termvoorstellen en verzoeken om technische aanpassingen te ondervangen.
- Er wordt van leveranciers van collectiebeheersystemen zoals De REE verwacht dat zij participeren in het termennetwerk. Hoewel deze welwillendheid er aan beide kanten zeker is (en er ook diverse gremia zijn waarin deze onderwerpen besproken worden) betekent dit ook voor leveranciers dat zij hiervoor resources moeten inzetten. Er moeten dan ook goede werkafspraken worden vastgesteld tussen de diverse actoren in het termennetwerk. Zo zal DE REE zelf geen kosten rekenen voor het gebruik van de GTAA Onderwerps-as, maar wel voor de lokale implementatie en koppeling aan de gewenste velden van de klant.

Projectresultaat 3: Onderzoek inzet Persistent Identifiers

Persistent Identifiers (PIDs) zijn permanente identificatienummers die gegarandeerd blijven werken, ook al verandert het webadres van een organisatie. Er zijn verschillende soorten PID's (Handle, URN:NBN en DOI) die ieder diensten aanbieden voor het aanmaken van een permanent werkende link.³⁶ Kort gezegd werkt het als volgt:

- Er wordt een PID aangemaakt voor een record bij een PID-dienst op basis van het webadres.
- Wanneer een webadres wijzigt (eg www.nationaalarchief.nl naar www.gahetna.nl) moet je de verwijzing van de PID aanpassen.
- Hierdoor maakt het niet uit welk webadres een record heeft, zolang de eindgebruiker bij het verwijzen ernaar maar gebruik maakt van de PID, in plaats van het webadres. Wanneer je de PID in je browser invoert, 'resolved' deze vanzelf naar het courante webadres.

Om de mogelijke inzet van PIDs voor het webarchief van Groninger Archieven te onderzoeken is er contact gezocht met het werkpakket Houdbaar³⁷ van NDE en de toenmalige projectleider van het Nationaal Archief. Naast de expertise die binnen Houdbaar aanwezig is, maakt de het Nationaal Archief de overstap naar MAIS-Flexis, waarbinnen het aanmaken van PIDs een key requirement is.

Keuze maken uit de diverse PID-systemen

Er zijn verschillende systemen om PIDs te implementeren. De belangrijkste die binnen de erfgoedwereld gebruikt worden en aanbevolen worden voor de NCDD zijn Handle, URN/NBN en DOI. Meer specificaties en achtergrondinformatie is te vinden op <http://www.ncdd.nl/pid/>, maar grofweg wordt URN:NBN met name gebruikt in de bibliotheekwereld, DOI voor het persistent verwijzen naar online artikelen en Handle bij archieven. DOI en Handle zijn goed uitwisselbaar. De hiërarchie uit de inventaris komt niet terug in de Persistent Identifier.

Binnen dit project is gebruik gemaakt van de PID-wijzer die binnen NDE Houdbaar is ontwikkeld (<http://www.ncdd.nl/pid-wijzer/>). Daarnaast is er advies uitgebracht vanuit het PID-project van NDE Houdbaar. Het resultaat was dat Handle het beste systeem zou zijn voor Groninger Archieven, zowel voor het webarchief, als voor de collectie in het algemeen.

Naast het feit dat Handle algemeen goed is toe te passen (en niet zoals URN:NBN en DOI voor specifieke soorten publicaties wordt gebruikt) is het Handle systeem flexibel. Er zijn wel de volgende aandachtspunten:

- Je moet Handle-verwijzingen zelf aanmaken en onderhouden. Dus: als de URLs van je website wijzigen of je stapt over naar een ander PID-systeem, moet je als organisatie de nieuwe URLs koppelen aan de bestaande Handles.
- Het is niet de bedoeling om PIDs aan te maken voor objecten die niet publiekelijk beschikbaar zijn.
- Idealiter worden PIDs voor records automatisch gegenereerd binnen MAIS-Flexis zelf.
- Je moet als organisatie je eigen zogeheten Handle-prefix aanvragen, zodat specifiek voor jouw instelling Handles kunnen worden aangemaakt. De kosten hiervoor zijn laag (ongeveer \$50 per jaar). Dit is een stap die gemaakt moet worden, ook al zit de PID-functionaliteit standaard in je collectiebeheersysteem.
- Een GUID of een URI is niet hetzelfde als een PID. Dergelijke identifiers kunnen net zo goed

³⁶ http://ncdd.nl/site/wp-content/uploads/2015/10/20150928_NDE_WP3_Persistent_Identifier_Projectplan.pdf

³⁷ <http://www.ncdd.nl/projecten/netwerk-digitaal-erfgoed/project-persistent-identifiers/>

uniek zijn als een PID, maar PIDs dien je in een extern systeem te registreren en onderhouden.

Bredere ontwikkelingen met betrekking tot PIDs

Groninger Archieven is actief betrokken bij belangrijke discussies en ontwikkelingen met betrekking tot de inzet van Persistent Identifiers binnen collectiebeheersystemen (inzetten op recordniveau, voor assets, etc). Zo wordt er afstemming gezocht binnen de MAIS-Flexis gebruikersgroep en in het bijzonder vanuit het Nationaal Archief, waarvoor PIDs binnen hun MAIS-Flexis architectuur een kerncomponent vormt.

Om zo goed mogelijk aan te sluiten bij de resultaten van bovenstaande ontwikkelingen, wacht Groninger Archieven nog met het definitief inzetten van PIDs.

De belangrijkste vragen die bij deze ontwikkelingen tzt beantwoord moeten worden met betrekking tot webarchieven is op welk niveau/welke niveaus de inzet van PIDs wenselijk is. Hiervoor zijn meerdere opties, zoals:

- Binnen MAIS-Flexis:
 - per crawl van een website op het niveau van het record van deze crawl.
 - een set van crawls van dezelfde URL.
 - een referentie naar het (W)ARC bestand van een crawl, en/of specifieke onderdelen binnen een (W)ARC
- Bij GW Crossmedia:
 - een referentie naar het (W)ARC bestand van een crawl, en/of specifieke onderdelen binnen een (W)ARC
 - de URL van een gearcheeerde website (of: per URL van iedere pagina van een gearcheeerde website)

Daarbij is de vraag hoe de workflow van het aanmaken van PIDs moet gaan lopen. Bij het Nationaal Archief wordt een verzoek voor het aanmaken van een PID naar de Handle-service worden gemaakt vanuit het collectiebeheersysteem, bij het aanmaken van een record. Het zou dan ook wenselijk zijn om deze PID tevens te koppelen aan de crawl-informatie bij GW Crossmedia. Dit kan handmatig, maar het is ook mogelijk om dit automatisch te laten doen. Wat hiervoor de beste aanpak is, zal nog nader worden bepaald.

Conclusie

De focus van dit project lag op het genereren van inzichten en aanbevelingen die relevant zijn voor het beoogde nationale termennetwerk voor digitaal erfgoed. Use cases en pilots zoals dit project zijn essentieel om inzichtelijk te maken wat de voordelen zijn van het gebruiken van thesauri van derden, en wat hiervan de impact is op erfgoedinstellingen en leveranciers. De belangrijkste conclusies zijn als volgt.

Inhoudelijk

In het termennetwerk moet een helder overzicht worden gemaakt van de beschikbare thesauri en op welke type collecties zij over het algemeen goed aansluiten, inclusief concrete voorbeelden en korte documentatie van de inhoud van de thesauri. Dit zal het begrip van de mogelijkheden van gedeelde thesauri vergroten.

Verder bleek dat Onderwerpen veel lastiger te definiëren zijn dan Named Entities, zoals persoonsnamen en plaatsnamen. Daarbij heeft iedere collectie zijn eigen specifieke eigenschappen, die mogelijk even specifieke onderwerpen benodigen. Dit maakt gezamenlijk beheer enorm complex. Een mogelijke oplossing kan zijn om thesauri waar mogelijk met elkaar te verbinden. Hierdoor ontstaan er wel dwarsverbanden tussen collecties, maar kan een instelling wel haar eigen thesaurus beheren en behouden. Een andere oplossing is om thesauri van derden te gebruiken, maar het ook mogelijk maken voor een instelling om op lokaal niveau ontbrekende termen toe te voegen. Dit schiet echter het doel voorbij van een gedeeld termennetwerk. Uiteindelijk zal voor een brede adoptie van een termennetwerk het noodzakelijk zijn voor instellingen om een verminderde mate van controle over gebruikte termen te accepteren.

De ontwikkeling van de WOII Thesaurus binnen het Netwerk Oorlogsbronnen is een perfecte use case om deze inhoudelijke vragen verder te beantwoorden. Vanuit dit project zullen belangrijke concrete vraagstukken en aandachtspunten in kaart worden gebracht die het gebruik van een onderwerpsthesaurus en hiermee ook de workflow voor het voorstellen van kandidaattermen door meerdere instellingen met zich mee brengt.

Technisch

Het is voor het faciliteren van een brede adoptie belangrijk dat er een deugdelijke infrastructuur wordt opgezet die enerzijds flexibel en laagdrempelig is, maar anderzijds zeer robuust en stabiel.

Een mogelijke oplossing voor het vereenvoudigen van het inhoudelijk beheer van een onderwerpsthesaurus, is om het technisch mogelijk te maken om meerdere thesauri binnen één onderwerpsveld in te zetten. Zo bleek in dit project dat de Namen-as een goede aanvulling zou zijn voor de Onderwerps-as. Daarbij zouden andere 'Named Entities' zoals Persoonsnamen en Geografische Namen ook als onderwerp kunnen worden toegevoegd. Wanneer dergelijke concepten ook beschikbaar zou zijn binnen hetzelfde onderwerpsveld, zou het niet nodig hoeven zijn om ofwel kandidaattermen voor te stellen, ofwel termen uit andere concept schemes over te hevelen naar/dupliceren in de Onderwerps-as.

Een voorbeeld van een organisatie die voor een dergelijke oplossing heeft gekomen is de Finse publieke omroep Yle. In hun systeem kunnen journalisten termen uit zowel de intern beheerde thesauri als andere bronnen zoals Wikidata toevoegen.³⁸ Tijdens het annoteren kan de gebruiker zien in de autocomplete-suggesties uit welke bron(nen) de termen komen en welk type term het betreft (eg: onderwerp, persoon, geografische locatie) en de beste match kiezen.

Het annotatieproces kan vergemakkelijkt worden door dit vanuit één veld mogelijk te maken voor verschillende type onderwerpen. Aangezien aan een thesaurusterm een type gekoppeld is kan dit uiteindelijk weer verdeeld worden over meerdere velden en/of zoekfilters. Een mogelijk nadeel is dat de gebruiker niet wordt geholpen om de inhoudelijke termen 'systematisch' toe te kennen, dus eerst de onderwerpen, dan persoonsnamen, dan geografische namen en dan de overige eigennamen. Voor beheer zou het ook wenselijk kunnen zijn om de concept schemes uit te splitsen.

Een dergelijke oplossing van het koppelen van meerdere thesauri aan één veld vereist wel een goed uitgewerkte business case en vervolgens een heldere presentatievorm. Het tonen van de bron van de thesaurus en het type term is hierbij een minimum, maar meer contextinformatie zoals scope notes en broader terms is ook wenselijk.

De vraag rijst echter welke thesaurus de bron is en/of leidend is, wanneer er meerdere thesauri beschikbaar zijn in één veld. Ten eerste omdat dezelfde term in meerdere thesauri kan voorkomen en ten tweede omdat dit kan betekenen dat er al koppelingen zijn gemaakt tussen deze thesauri.

Resources

Bij de diverse gesprekken en evaluaties die binnen het project plaatsvinden bleef de vraag naar voren komen: wat voor impact heeft het gebruik van thesauri uit het termennetwerk op de resources in mijn organisatie? Deze vragen kwamen vanuit drie soorten stakeholders: de nieuwe thesaurusgebruikers, de thesaurusaanbieders en de leveranciers.

De organisaties die thesauri van derden willen gebruiken, moeten in eerste instantie extra menskracht moeten inzetten en materiële kosten maken voor de technische implementatie. Daarnaast moeten inhoudelijke experts betrokken worden om de geschiktheid van de thesaurus te beoordelen en vervolgens om deze in gebruik te nemen. Wanneer er na implementatie technische aanpassingen nodig zijn, of er inhoudelijke vragen en verzoeken zijn zijn hier ook resources voor nodig. Daar staat tegenover dat het niet meer noodzakelijk is een eigen thesaurus te onderhouden. Daarbij wordt door het gebruik van een termennetwerk een verbinding gemaakt tussen diverse erfgoedcollecties, wat de kwaliteit verbetert en de vindbaarheid vergroot.

De thesaurusaanbieders krijgen op hun beurt weer te maken met technische en inhoudelijke vragen van de nieuwe gebruikers. Daarnaast komen er vragen vanuit leveranciers die de implementatie moeten doen van thesauri in de collectiebeheersystemen, of aanpassingen moeten maken vanwege een bredere set aan requirements in de systemen waarin thesauri worden aangeboden en beheerd. Dit betekent dus ook een extra benodigde inzet van resources aan de kant van de leveranciers.

Er zullen meer concrete business cases moeten worden uitgewerkt om inzicht te geven in de kosten, maar ook de baten van het delen van thesauri, voor alle drie stakeholder groepen.

³⁸ <http://wikimedia.fi/2016/04/15/yle-3-wikidata/>

Vervolgstappen

Tijdens het project zijn er nog enkele concrete mogelijke vervolgstappen gedefinieerd met betrekking tot de inzet van de GTAA bij Groninger Archieven.

Ten eerste is het de wens om het webarchief publiek toegankelijk te maken op de website van Groninger Archieven. Hierbij is het de wens om voor het zoeken en browsen de hiërarchische structuur van de GTAA in te zetten. Bijvoorbeeld: door bij records ook de broader en narrower terms te laten zien, op basis van zoekvragen filters en suggesties aan te bieden op basis van de thesaurus of door zoekvraagexpansie (waarbij eg ook alle narrower terms in de zoekvraag worden meegenomen).

Ten tweede staat het webarchief nu nog als aparte collectie in MAIS-Flexis, terwijl er ook websites in zitten van organisaties waarvan ook een inventaris bestaat. Het is de wens om deze gearchiveerde websites ook onderdeel te maken van de bijbehorende inventarissen.

De GTAA zou breder binnen Groninger Archieven kunnen worden ingezet. Voorbeelden hiervan is het mogelijk maken om Nadere Toegangen te annoteren met concepten uit de GTAA. De collectie van het Gronings Audiovisueel Archief (GAVA) zou ook goed gebruik kunnen maken van de GTAA. Het betreft een av-collectie (net als die van Beeld en Geluid), waarvan de amateurfilms voor een groot deels reeds beschikbaar is op het Amateurfilm Platform. Dit is het Nederlandse platform voor amateurfilmcollecties, dat wordt beheerd door Beeld en Geluid. Via de GTAA zouden er koppelingen ontstaan tussen de amateurfilmcollecties van Groninger Archieven en Beeld en Geluid. Hierbij zou het interessant zijn om te kijken in hoeverre het mogelijk is om automatisch termen te extraheren³⁹ uit de beschrijvingen van de films van het GAVA, aangezien deze erg rijk zijn en op fragmentniveau zijn toegevoegd.

³⁹ Hierbij zou geanalyseerd moeten worden of het niet beter zou zijn om niet de Onderwerps-as te gebruiken, maar de onderwerpen die specifiek zijn gedefinieerd voor shots en fragmenten. Dit is een uitgebreidere variant op de Onderwerps-as.

Bibliografie

“Atlantis Wordt Steeds Dynamischer!” DEVENit, April 1, 2016. <http://www.deventit.nl/atlantis-wordt-steeds-dynamischer/>.

Baltussen, Lotte Belice, Jaap Blom, Leïla Medjkoene, Radu Pop, Jasmijn van Gorp, Hugo C. Huurdeman, and Leidi Haaijer. “Hard Content, Fab Front-End: Archiving Websites of Dutch Public Broadcasters.” *Alexandria* 25, no. 1–2 / August 2014 (August 2014). doi:10.7227/ALX.0021.

“De Oorlog Op Zijn Plaats: Pilotproject Geocoding Afgerond.” *Netwerk Oorlogsbronnen*, July 18, 2016. <http://oorlogsbronnen.nl/nieuws/de-oorlog-op-zijn-plaats-pilotproject-geocoding-afgerond>.

Dooley, Jackie. “Feedback, Please! DRAFT Web Archiving Metadata Best Practices!” *Hanging Together*, June 1, 2017. <http://hangingtogether.org/?p=5998>.

Dooley, Jackie. “Slam Bam WAM: Wrangling Best Practices for Web Archiving Metadata.” *Hanging Together*, August 23, 2016. <http://hangingtogether.org/?p=5684>.

Dooley, Jackie M., Karen Stoll Farrell, Kammi Tim, and Jessica Venlet. “Developing Web Archiving Metadata Best Practices to Meet User Needs.” *Journal of Western Archives* Issue 2, no. Volume 8 (2017). <http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1079&context=westernarchives>.

Enders, Markus. “A METS Based Information Package for Long Term Accessibility for Web Archives,” 2010. <http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/enders-70.pdf>.

Ewing, Renata, and Lori Donovan. “Metadata in Archive-IT.” *Archive-It Help*, April 17, 2014. <https://webarchive.jira.com/wiki/display/ARIH/Metadata+in+Archive-It>.

Farnel, Sharon. “Metadata Workflows for Web Archiving: Thinking and Experimenting with Metadata in ‘bulk.’” presented at the Archive-It Partner Meeting, November 12, 2013. <https://webarchive.jira.com/wiki/download/attachments/61145091/Metadata%20Workflows%20for%20Web%20Archiving%20-%20AI%20Partner%20Meeting%20Nov%202013.pdf?version=1&modificationDate=1384778473058&api=v2>.

Free, Tessa. “Vinden Wat Erin Zit.” *Netwerk Oorlogsbronnen*, September 26, 2016. <http://oorlogsbronnen.wixsite.com/blog/single-post/2016/09/26/Vinden-wat-erin-zit>.

Guenther, Rebecca S. “MODS: The Metadata Object Description Schema,” *Libraries and the Academy*, 3, no. 1 (January 2003): 137–50.

NDSA. “NDSA Web Archiving Survey Report.” Washington, D.C.: NDSA Content Working Group, June 19, 2012. <http://www.digitalpreservation.gov/ndsaworkinggroups/documents/ndsawebarchivingsurveyreport2012.pdf>.

Niu, Jinfang. “An Overview of Web Archiving.” *D-Lib Magazine* 18, no. 3/4 (March 2012). doi:10.1045/march2012-niu1.

Peterson, Christie. “Archival Description for Web Archives.” *Medium*, June 22, 2015.

<https://medium.com/on-archivy/archival-description-for-web-archives-1d9dce8dcef0#.ly9xjwrlj>.

Priem, Matthias, Victor de Boer, Michiel Hildebrand, Johan Oomen, and Nico Verplancke. "Gemeenschappelijke Thesaurus Voor Uniforme Ontsluiting - Eindrapport," June 17, 2016. <http://viaa.be/assets/files/news/downloads/Eindrapport-GTUU-Taalunie.pdf>.

Roberto, Frankie. "Building the Wellcome Library Alpha – People." *Good, Form & Spectacle*, September 4, 2016. <https://goodformandspectacle.wordpress.com/2016/09/04/building-the-wellcome-library-alpha-people/>.

———. "Building the Wellcome Library Alpha – Subjects." *Good, Form & Spectacle*, September 4, 2016. <https://goodformandspectacle.wordpress.com/2016/09/04/building-the-wellcome-library-alpha-subjects/>.

Romaniuk, Laurentia. "Metadata for a Web Archive: PREMIS and XMP as Tools for the Task," February 26, 2014. <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=2755&context=libphilprac>.

Smith-Yoshimura, Karen. "Analysis of International Linked Data Survey for Implementers." *D-Lib Magazine* 22, no. 7/8 (July 2016). doi:10.1045/july2016-smith-yoshimura.

———. "Faceted Vocabularies." *Hanging Together*, October 31, 2016. <http://hangingtogether.org/?p=5739>.

Webster, Peter. "What Do We Need to Know about the Archived Web?" *Webstory: Peter Webster's Blog*, April 18, 2016. <https://peterwebster.me/2016/04/18/what-do-we-need-to-know-about-the-archived-web/>.

Wildenborg, Ykje, and Dieter Suls. "Gastblog: Schone, Gelinkte Modeverzamelingen Op Komst." *Digitaal Erfgoed Nederland*, June 2, 2016. <http://www.den.nl/blog/bericht/5379/>.

Bijlage 1 - MODS-Lite velden geïmplementeerd bij Groninger Archieven

MODS-element	MaisFlexis veld	Opmerkingen en aandachtspunten
title	Titel	Titel van de website. 1. Zorg voor herhaalvelden waar titelwijzigingen in vermeld staan (met datumstempels) en geef aan welke titel actief is. 2. Beslis wanneer je een nieuwe beschrijving van een website maakt. De KB heeft als uitgangspunt: maak een nieuwe beschrijving als zowel de naam als de url van de website verandert. Bijvoorbeeld: het College Bescherming Persoonsgegevens omgedoopt tot Autoriteit Persoonsgegevens met een nieuwe url. Maak dan een nieuw record, sluit de oude af en verwijst in beide records over en weer (zie ook metadatumveld relatedItem).
name	Verantwoordelijke	Meestal een corporatie die verantwoordelijk is voor het beschikbaar maken van de website. 1. Hiervoor kan net als voor onderwerpen een vaste namenlijst worden gebruikt. 2. Zorg voor herhaalvelden waar naamswijzigingen in vermeld staan (met datumstempels) en geef aan welke naam actief is.
typeOfResource	Type	Binnen dit MODS-element zijn geen subtypes, maar is sprake van een vaste waarde, te weten 'software, multimedia'. 1. De vaste waarde is dermate nietszeggend dat het goed zou zijn om hiervoor in nationaal verband een vaste lijst voor te ontwikkelen en af te wijken van de MODS-waarde (eg: social media kanaal, website, app).
genre	NVT	Niet toegepast in dit project. Groninger Archieven gebruikt geen genres voor types websites, enkel onderwerpen.
originInfo	Datering Datum harvest	Datering: Indicatie van hoe lang een website al bestaat, dit zal vaak niet bekend zijn. Datum harvest: Datum waarop de crawl is gestart. 1. Het is aanbevelingswaardig om een overzicht in te voeren van de uitgevoerde crawls en de crawl frequentie.
language	NVT	Vooralsnog is het gehele webarchief van Groninger Archieven Nederlands, vandaar dat dit veld nog niet is gebruikt.
physicalDescription	NVT	Gaat echt om fysieke, tastbare kenmerken.
internetMediaType	NVT	Dit element in MODS Lite betreft het classificeren van sub-elementen (eg JPEG). Het was te specifiek voor het gebruik binnen Groninger Archieven.

abstract	Beschrijving	Beschrijving van het record.
tableOfContents	NVT	Te gedetailleerd voor deze use case.
note	NVT	Wellicht wel een relevant veld voor toekomstig gebruik. Bijvoorbeeld: 1. Voor een webarchief is vastlegging van afspraken en handelingen van groot (juridisch) belang, aangezien voor het crawlen toestemming nodig is. In 'note' kun je bijvoorbeeld (met datumstempel) correspondentie met de webbeheerders kwijt 2. Dit veld kan gebruikt worden om aan te geven binnen welke speciale webcollectie deze website valt (indien van toepassing).
subject	Onderwerp Geografisch trefwoord	Onderwerp: de GTAA Onderwerps-as is hier de basis. Geografisch trefwoord wordt gevuld met een interne lijst van Groninger Archieven (gaat eg om locatie Verantwoordelijke of werkgebied)
classification	NVT	Het Classificatieschema dat boven de GTAA-Onderwerpsas ligt is niet gebruikt in dit project.
relatedItem	NVT	Wellicht relevant voor latere implementatie. Bijvoorbeeld voor het vastleggen van de geschiedenis van de website 1. voortzetting van [verwijzing naar record] 2. voortgezet als [verwijzing naar record] 3. opgegaan in [verwijzing naar record]
identifier	NVT	Eventueel gebruiken voor Persistent Identifiers.
location	URL website live URL website archief	URL website live: Live URL van de hoofdpagina van de gearchiveerde website. URL website archief: De URL van de hoofdpagina van de gearchiveerde site. 1. Wanneer een URL van de live site verandert, voeg dan herhaalvelden toe waar deze wijziging in vermeld staat (met datumstempels) en geef aan welke url actief is. 2. Voor sommige websites moet het mogelijk zijn om meerdere actieve url's op te geven. Denk aan een forum of een blog die onderdeel van de website zijn, maar een ander adres hebben (en beslis of je die misschien apart wilt opnemen in je archief), of documenten die op een ander domein staan (waar de webbeheerder dan wel verantwoordelijk is).
accessCondition	NVT	Wordt wellicht later toegevoegd om onderscheid tussen toegangsrechten te kunnen aangeven: 1. Publieke, online toegang. 2. Onsite toegankelijk 3. Niet toegankelijk (Dark Archive)
recordInfo	GUID	GUID (Globally Unique Identifier) van het record. Automatisch gegenereerd door Oracle binnen MAIS-Flexis.

Bijlage 2 - Betrokkenen

Wie	Instelling	Thesaurus: technische/inh oudelijke vraagstukken	Metadatamodel: webarchief	PIDs
Alma Wolthuis Media Manager Metadatabeheer	Beeld en Geluid	1		
Arnous Goos Projectleider webarchivering	Beeld en Geluid		1	
Josefien Schuurman Manager Informatiemanagement	Beeld en Geluid	1		
Tim Manders Senior Media Manager Optimalisatie	Beeld en Geluid	1		
Vincent Huis in 't Veld Senior Media Manager Optimalisatie	Beeld en Geluid	1		
Willem Melder Software engineer	Beeld en Geluid	1		
Henk van der Meer Collectiebeheerder	Groninger Archieven	1		
JaapJan Hoogstins Collectiebeheerder	RHC Groninger Archieven	1	1	1
Harry Romijn Adjunct directeur RHC Groninger Archieven	RHC Groninger Archieven	1	1	1
Margreet Visch Collectiebeheerder	Groninger Archieven	1		
Enno Meijers Informatiemanager	Koninklijke Bibliotheek	1		
Johan Stapel Informatiemanager	Koninklijke Bibliotheek		1	
Peter de Bode Collectievorming webarchief	Koninklijke Bibliotheek		1	
Arno van der Kolk Programmeur/developer	Leverancier: DE REE	1		
Bas van Schoonhoven Programmeur/developer	Leverancier: DE REE	1		
Erik de Ree Senior consultant	Leverancier: DE REE	1		
Gerhard Hijzeler Senior consultant/ projectmanager	Leverancier: DE REE	1		1

Peter van Diermen Directeur	Leverancier: DEVENTit	1		
Ykje Wildenborg Projectcoördinator Modethesaurus	MoMu	1		
Antal Posthumus Adviseur Recordkeeping	Nationaal Archief		1	
Ed de Heer Projectleider	Nationaal Archief	1		
Gijsbert Kruithof Informatiearchitect	Nationaal Archief			1
Hans van der Neut semi-vrijwilliger en thesaurusdeskundige	Nationaal Archief	1		
Jeroen van Luin Verwerving & beheerder van digitale archieven	Nationaal Archief		1	
Tine van Nierop Project Manager Open Access & Use	Nationaal Archief	1		
Edwin Klijn Manager Netwerk Oorlogsbronnen	NIOD	1		
Inge van Stokkom Thesaurusontwikkelaar	NIOD	1		
Lizzy Jongma Senior ICT projectleider Netwerk Oorlogsbronnen	NIOD	1		
Joop Vanderheiden Coördinator Ergoedthesaurus	RCE	1		
Reem Weda Informatiespecialist terminologieën	RKD	1		
Michiel Hildebrand Linked Data specialist	Spinque	1		
Engelien de Ruijter Informatiespecialist	Stadsarchief Rotterdam	1	1	
Marie Claire Dangerfield Expert Informatiebeheer	Stadsarchief Rotterdam	1		
Marjoleine Houben Medewerker digitale dienstverlening	Stadsarchief Rotterdam	1		
Mies Langelaar Preservation Officer	Stadsarchief Rotterdam	1		
Sarah Knuvelde Informatieprofessional Collecties	Stadsarchief Rotterdam	1		

Annelot Vijn Applicatiebeheerder	Utrechts Archief	1	1	
Mathijn Elhorst Manager Digitale Techniek	VPRO	1		
Shannon van Muijden Databeheerder collectie	Zuiderzeemuseum	1		

NB Lotte Baltussen: mijn speciale dank gaat uit naar Arnoud Goos, Vincent Huis in 't Veld, Willem Melder, Josefien Schuurman, (Beeld en Geluid), Arno van der Kolk, Bas van Schoonhoven (DE REE), JaapJan Hoogstins, Harry Romijn (Groninger Archieven), Gijsbert Kruithof, Remco van Veenendaal (Nationaal Archief) voor het kritisch meelesen en waar nodig corrigeren van dit eindverslag.