



Technology Scout naar de ondersteuning van andere soorten content in MediaMosa

Versie 1.0

Datum 3 december 2009

SURFnet/Kennisnet Innovatieprogramma

Inhoud

Samenvatting	3
Inleiding.....	4
Media in MediaMosa	5
Informatie structuur	5
Workflow binnen MediaMosa	6
Upload en analyse	7
Storage	7
Benodigde uitbreidingen en aanpassingen.....	8
Scheiding analyse en core workflow	8
Configureerbare workflow en tools	8
Upload functionaliteit	8
Analyse workflow	9
Core workflow	10
Metadata analyse en opslag	12
Inhoudelijke analyse van bestanden	12
Meerdere storage locaties.....	13
Transcodering en verwerking.....	13
Afspelen van files	13
Content type groepering	14
Conclusies en aanbevelingen	16

Samenvatting

MediaMosa¹ is het open source media content management systeem dat is ontwikkeld door SURFnet, Kennisnet en Madcap. Op basis van MediaMosa is het VP-Core videoplatform ingericht. In deze technology scout is onderzocht welke aanpassingen nodig zijn in MediaMosa om ook andersoortige media dan video op te slaan en te verwerken.

MediaMosa is een platform waar eindgebruikers applicaties (EGA) gebruik van kunnen maken voor het opslaan en afspelen van video bestanden. MediaMosa is daarbij voor de eindgebruiker van deze applicaties 'onzichtbaar'.

Wanneer MediaMosa ook geschikt wordt voor andere typen content dan video, kan door een EGA nieuwe functionaliteit voor eindgebruikers worden toegevoegd. Een voorbeeld hiervan is dat bij een filmfragment een biografie over de acteur en de uitgesproken tekst (in tekst formaat) beschikbaar kan worden gesteld.

In dit rapport worden een aantal aanpassingen en uitbreidingen voorgesteld om MediaMosa geschikt te maken voor andere content typen dan video. Voor de meeste van deze aanpassingen of uitbreidingen geldt dat ze al in het oorspronkelijke ontwerp van MediaMosa waren voorzien, maar niet, of slechts gedeeltelijk, zijn geïmplementeerd.

De benodigde aanpassingen en uitbreidingen zijn:

- Huidige verwerking vervangen door een configureerbare workflow
- Opdeling van de verwerking in twee aparte workflows: een analyse en een core workflow.
- Configuratie beheren door de EGA van de core workflow via de REST interface.
- Uitbreiding van de opslag van metadata met een record voor iedere mediafile voor de uit de analyse verkregen data.

Daarnaast wordt in deze technology scout een aanzet gegeven voor het inrichten van de analyse workflow door het uitwerken van een indeling van media typen in groepen en het aangeven van een toolset voor deze groepen.

Een eerste logische stap zou het uitbreiden en inrichten van MediaMosa zijn met het verwerken van afbeeldingen en audio bestanden (naast de huidige video's). Deze uitbreiding sluit functioneel logisch aan bij de huidige implementatie in VP-Core. Daarnaast zijn voor deze content typen goede tools gevonden voor het analyseren en verwerken. In een verdiepend onderzoek kan voor de overige typen een set van tools worden gevonden die de analyse en verwerking binnen MediaMosa kunnen afdekken.

¹ <http://mediamosa.org/>

Inleiding

MediaMosa is een robuust en flexibel en schaalbaar mediamanagement platform. Het vormt de basis van tal van succesvolle videodiensten die SURFnet en Stichting Kennisnet aan het onderwijs in Nederland aanbieden of gaan aanbieden (bijvoorbeeld SURFmedia). MediaMosa is door deze beide organisaties, en de ontwikkelaar Madcap², recent in een Open Source licentie beschikbaar gesteld.

VP-Core is de implementatie van MediaMosa binnen SURFnet. VP-Core ontsluit videocontent die opgeslagen staat op het gezamenlijke videoplatform van SURFnet en Kennisnet. Zo wordt VP-Core gebruikt als repository voor SURFmedia³. In het vervolg van dit document wordt gesproken over MediaMosa, waarbij dan ook de SURFnet specifieke implementatie VP-Core bedoeld wordt.

MediaMosa biedt ook de mogelijkheid om software en andere content als leerobjecten te distribueren. In deze technologieverkenning wordt onderzocht welke aanpassingen aan VP-Core gedaan moeten worden om het geschikt te maken voor de opslag en verspreiding van andere content dan videocontent.

In het eerste hoofdstuk wordt een beschrijving gegeven van de huidige verwerking van video files in MediaMosa. Het asset concept en de huidige workflow, van upload via analyse en transcoderen naar opvragen en afspelen worden toegelicht.

Het hoofdstuk 'Benodigde uitbreidingen en aanpassingen' geeft een overzicht van voorgestelde aanpassingen en uitbreidingen aan MediaMosa. Op basis van een analyse van het huidige ontwerp en implementatie is een aantal zaken gevonden waar, voor het ondersteunen van andere content, aanpassingen noodzakelijk zijn.

Tevens is een categorisering van soorten media files gemaakt. Filetypen zijn op basis van (technische) kenmerken ingedeeld in content groepen. Per content groep is een onderzoek gedaan naar geschikte tooling voor het verwerken van dergelijke bestanden. Op basis van de resultaten, beschreven in de hoofdstukken 'Benodigde uitbreidingen en aanpassingen' en 'Content type groepering', worden drie tools aanbevolen om toegepast te worden in MediaMosa.

In de conclusies en aanbevelingen wordt een overzicht gegeven van de bevindingen, de aanbevolen uitbreidingen en aanbevolen aanpassingen. Daarnaast wordt het uitvoeren van een aantal spikes (verdiepend onderzoek), voor een verdere verdieping, geadviseerd.

² <http://www.madcap.nl/>

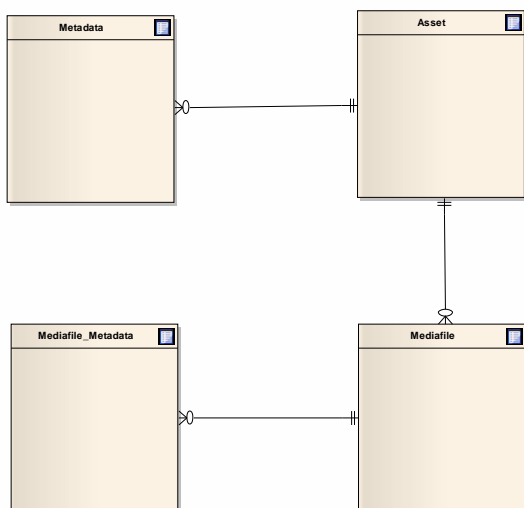
³ <http://www.surfmedia.nl/>

Media in MediaMosa

MediaMosa biedt een infrastructuur die gebruikt wordt door eindgebruikers applicaties (EGA); eindgebruikers werken niet rechtstreeks op MediaMosa. De EGA's roepen services aan die MediaMosa aanbiedt voor het opvragen, bewerken en opslaan van assets. Het afspelen van mediafiles wordt in MediaMosa afgehandeld door de play-proxy service. Deze proxy is verantwoordelijk voor de afscherming; media files mogen alleen aan geautoriseerde gebruikers beschikbaar gesteld worden.

Informatie structuur

MediaMosa kent een informatie structuur waarbij assets het basis element vormen (zie Figuur 1 Informatie structuur Asset). Eén asset bestaat uit een set van metadata, die zowel beschrijvend als technisch is, en daaraan gekoppelde mediafiles. Voorafgaand aan het uploaden van een mediafile naar MediaMosa wordt een asset aangemaakt waaraan het geuploade bestand wordt gekoppeld.



Figuur 1 Informatie structuur Asset

Een asset kan meerdere mediafiles bevatten; in het geval van video kunnen dit verschillende versies van het zelfde videofragment zijn die op verschillende wijzen gecodeerd zijn. Mediafiles bij een asset kunnen ook verschillend van aard en type zijn, zoals video stills bij een videofragment.

Een mediafile kent zijn eigen set van metadata. Deze bevat voornamelijk technische data, zoals de gebruikte codec, bitrate, fps enzovoort. Deze data wordt automatisch afgeleid uit de mediafiles zelf tijdens de analyse fase.

Afhankelijk van de geconfigureerde workflow kunnen, op basis van één (geüploade) mediafile, door MediaMosa meerdere mediafiles worden aangemaakt (transcodering naar ander formaat, creatie 'stills' bij videofragmenten) en aan de asset worden gekoppeld.

Workflow binnen MediaMosa

De workflow van de verwerking van een mediafile in MediaMosa kent de volgende stappen:

- **Uploaden** van het originele bestand. Deze stap valt uiteen in vier substappen: Het aanmaken van de asset, het creëren van een mediafile bij deze asset, het aanvragen van een upload ticket en de (uiteindelijke) upload.
- **Analyse:** Het bestand wordt geanalyseerd om het (technische) formaat te achterhalen. Op basis van het resultaat van deze analyse wordt een bestand getranscodeerd naar een ander formaat. Uitgangspunt bij de analyse is dat de file extensie van het bestand niet vertrouwd kan worden; MediaMosa bepaalt (en overschrijft) de toekenning van formaat en type die een EGA of gebruiker (impliciet) meegeeft.
- **Storage:** Het mediabestand wordt opgeslagen op het achterliggende opslagmedium.
- **Transcoderen:** Bij het uploaden van een mediafile (video) kan de opdracht meegegeven worden om deze te transcoderen in een aantal formaten.
- **Afspelen:** Deze laatste stap staat los van de workflow die wordt ingezet bij het aanmaken van de asset en het uploaden van mediafiles. Een eindgebruiker van een EGA verzoekt om het afspelen van een mediafile; de EGA geeft dit verzoek door aan MediaMosa.

Afspelen van mediafiles betekent in de MediaMosa context een keuze uit een van de drie soorten afspeel acties:

- Doorgeven van URL aan play proxy
- Doorgeven van blok(je) HTML code (met daarin link naar play proxy URL) voor het embedden van video binnen HTML
- Doorgeven van link voor download van bestand.

Binnen MediaMosa wordt de hierboven beschreven workflow deels uitgevoerd door een job mechanisme. Voor taken die een aanzienlijke verwerkingstijd en capaciteit nodig hebben, zoals transcoderen, worden jobs aangemaakt en ingepland.

Naast deze taken kent MediaMosa ook jobs voor het controleren van de integriteit tussen de informatie over assets in de database en de opgeslagen bestanden op het filesysteem en een job voor het opruimen van lege assets en mediabestanden.

Upload en analyse

In de huidige workflow is de analyse van de mediafiles beperkt tot het vaststellen van het (video)formaat en het bepalen van een aantal (technische) metadata-velden. Er vindt geen inhoudelijke controle of analyse plaats. Deze workflow wordt automatisch gestart na het uploaden van een bestand.

De analyse van geuploade mediabestanden wordt nu uitgevoerd door het file commando en de FFmpeg⁴ tool, speciaal geschikt voor het bepalen van videoformaten (en codecs). Het file commando vergelijkt de headers van bestanden met een set zogenaamde 'magic files' van bekende formaten en stelt daarmee het type vast. In de daarop volgende stap wordt FFmpeg gebruikt voor een verdiepende analyse van video bestanden.

De huidige ftp batch upload werkt (zoals de naam al aangeeft) alleen met ftp; er kunnen via ftp mediabestanden en metadata worden geupload naar MediaMosa. Samen met deze collectie bestanden wordt een Video Upload File (VUF) geupload, waarin instructies staan die toegepast worden op de geuploade files om deze in MediaMosa te importeren.

Storage

MediaMosa kent één werkwijze voor het opslaan van geuploade bestanden. Bij de huidige verwerking worden alle bestanden naar dezelfde locatie geschreven. Omdat er één mountpoint naar een achterliggend filesysteem wordt gebruikt, is het niet mogelijk onderscheid te maken tussen specifieke locaties. Het maken van een specifieke keuze op basis van kenmerken van het bestand, EGA of collectie is niet mogelijk.

⁴ <http://ffmpeg.org/>

Benodigde uitbreidingen en aanpassingen

In dit hoofdstuk worden de uitbreidingen en aanpassingen aan MediaMosa beschreven, die nodig zijn voor het kunnen verwerken van andere content typen. Het aantal wijzigingen oogt aanzienlijk, voor de meesten geldt echter dat deze een daadwerkelijke implementatie zijn van voorzieningen waarin bij het oorspronkelijke ontwerp al was voorzien.

Scheiding analyse en core workflow

Om een (betere) ondersteuning van meer content typen in MediaMosa mogelijk te maken, is een tweedeling van de workflow nodig. De huidige stappen in de workflow zijn (min of meer) hard coded vastgelegd en toegesneden op video bestanden. Door de verwerking in twee stappen te laten verlopen, een aparte analyse en metadata extractie stap gevolgd door een daadwerkelijke verwerking, ontstaat een meer flexibel en toekomst bestendig platform.

Configureerbare workflow en tools

Met een configureerbare workflow wordt een op stappen gebaseerd schema gerealiseerd, waarbij in iedere stap een voorgedefinieerde actie kan worden uitgevoerd op een bestand. Voorbeelden van dergelijke acties zijn het aanroepen van een (externe) tool voor het analyseren of het transcoderen van een bestand.

Zowel de analyse- als de core-workflow maken gebruik van dezelfde regel module, waarin de twee workflows worden aangemaakt en beheerd. De Drupal Rules module lijkt hier de meest geschikte kandidaat voor.

Voor het plaatsen van externe tools voor het uitvoeren van analyses (en/of bewerkingen) op bestanden kan de huidige MediaMosa opzet toegepast worden. Dit houdt in dat een tool via de command line van het operating systeem wordt gestart. De output vanuit de tool wordt naar MediaMosa gedirigeerd en daar verder verwerkt (tekst of XML gebaseerd). Dit mechanisme biedt een zeer flexibele methode om tools binnen MediaMosa te integreren.

Upload functionaliteit

De analyse workflow wordt geïnitieerd bij het uploaden van een bestand. Het huidige upload mechanisme voldoet ook voor andere content typen. Aanpassingen zijn hier niet noodzakelijk.

Wel bestaan er twee scenario's waarvoor een uitbreiding van de upload functie benodigd is:

- Het uploaden van batches van assets gebundeld in archieven.
- Het uploaden van rich media formaten waarbij de samenhang moet worden bewaakt.

Voor beide scenario's moet de upload functie van MediaMosa uitgebreid worden met de mogelijkheid om zip bestanden met daarin alle samenstellende delen van de asset samen met verwerkingsinstructies te verwerken. Dit zou geïmplementeerd kunnen worden als een uitbreiding het huidige ftp upload mechanisme (door het VUF bestand binnen de zip file te plaatsen).

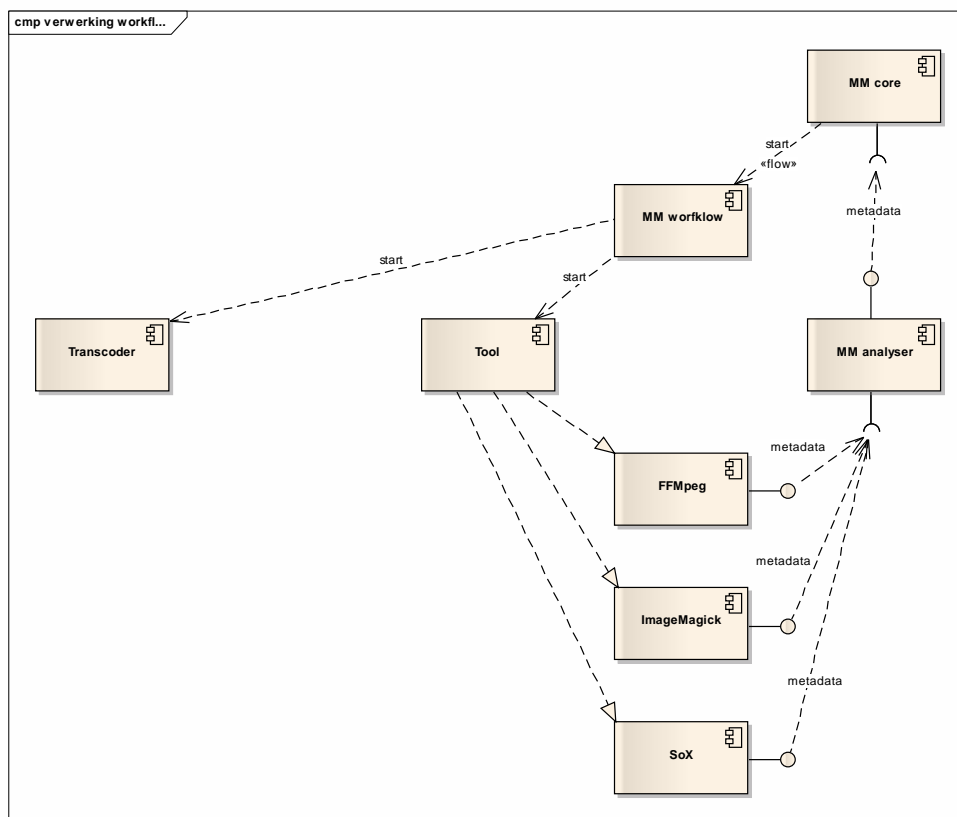
Analyse workflow

Het uitgangspunt voor de analyse workflow blijft dat MediaMosa zelf bepaalt wat het formaat en type van een mediabestand is, ook bij het werken met meerdere content typen. De analyse moet uitgebreid worden met extra stappen met per stap een geschikte tool om mediabestanden voldoende te kunnen analyseren.

Met de configureerbare workflow kan deze analyse worden ingericht binnen MediaMosa. Een geupload bestand kan door meerdere tools worden getest. Afhankelijk van het testresultaat kan een volgende analyse stap worden gestart of overgeslagen.

Het resultaat van de analyse workflow is de geaggregeerde metadata die bij het geuploade bestand wordt opgeslagen. De resultaten van iedere analyse stap worden samengevoegd tot één set metadata bij de file. De MediaMosa analyser component zal dit aan de Core doorgeven in RFC 822⁵ formaat.

⁵ <http://tools.ietf.org/html/rfc822>

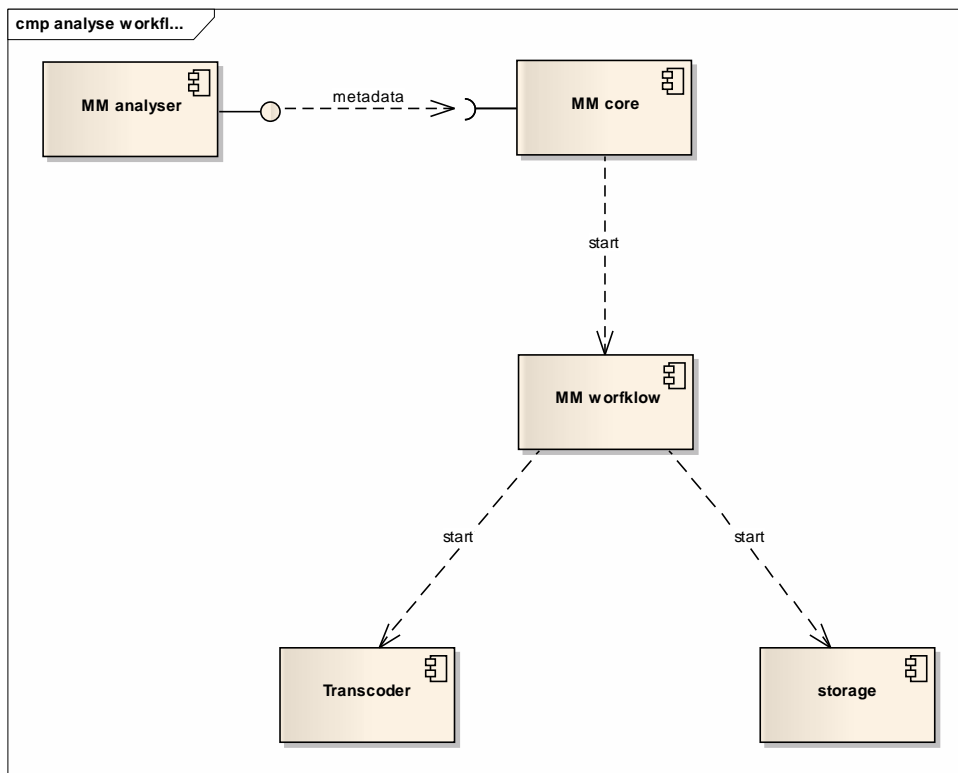


Figuur 2 Analyse workflow

In figuur 2 wordt de inrichting van de analyse workflow weergegeven. Vanuit MediaMosa core wordt deze geïnitieerd na de upload van een bestand. De workflow component start dan de analyse workflow, waarin geconfigureerd staat welke tools in welke volgorde gestart worden. Het resultaat van deze analyse(s) wordt geaggregeerd door de MediaMosa analyser component.

Core workflow

De Core krijgt een eigen workflow die op basis van de uitkomsten van de analyse, bepaalt welke stappen nodig zijn voor het daadwerkelijk verwerken. Dit kunnen acties zijn als het transcoderen naar een ander formaat of het plaatsen van een file in een specifieke collectie.



Figuur 3 Core workflow

In figuur 3 wordt deze workflow weergegeven. Metadata worden vanuit de analyse aangeleverd aan MediaMosa Core. Daar wordt de workflow gestart voor de Core verwerking. Binnen de Core workflow wordt op asset nivo gewerkt. De workflow kan beslissingen nemen gebaseerd niet alleen op de nieuwe metadata van de recent geuploadede file, maar kan ook de metadata van de asset en andere media files gebruiken.

Omdat de gewenste workflow op dit niveau bepaald wordt door de EGA, zal deze workflow vanuit de EGA beheerd en aangestuurd moeten worden. De REST interface tussen MediaMosa en de EGA zal uitgebreid moeten worden met operaties hiervoor. Bij het uploaden van een bestand kan de EGA een specifiek workflow profiel opgeven. Dit profiel is via (nieuwe) operaties vooraf gedefinieerd door de EGA in MediaMosa.

Metadata analyse en opslag

Naast het uitbreiden van het aantal analyse stappen om meerdere content typen te ondersteunen, zal binnen de analyse ook de ondersteuning van metadata worden uitgebreid. Afhankelijk van de toegepaste tools, het gehanteerde bestandsformaat en de kenmerken van het mediatype, zal de analyse van het mediabestand leiden tot een afwijkende en/of grotere set metadata dan nu toegepast wordt.

Binnen de metadata is onderscheid te maken tussen technische en beschrijvende data. De technische data geeft informatie over de technische eigenschappen van de mediafile, zoals gebruikte codec of hoeveelheid pixels. De beschrijvende data bevat informatie op eindgebruikers niveau, zoals titel, beschrijving en tags.

Binnen MediaMosa werd alle metadata die door FFmpeg uit een mediafile werd gelezen beschouwd als technische data. Bij het uitbreiden met andere content is deze aanname niet langer vol te houden. Formaten als jpeg bevatten zowel technische data, maar kunnen ook door de gebruiker gedefinieerde metadata bevatten (zoals tags).

Om hier in MediaMosa mee om te gaan zal de metadata verkregen uit de analyse van de file altijd en ongewijzigd beschikbaar blijven. De uit de analyse verkregen metadata wordt bij een asset opgeslagen in een eigen record zodat deze beschikbaar blijft ook na updates (bijvoorbeeld vanuit de EGA) op de metadata. Voor alle analyse tools die worden toegepast op een mediafile wordt de metadata samengevoegd en als één metadata set opgeslagen.

Inhoudelijke analyse van bestanden

Naast een analyse om bestandstype en -formaat te achterhalen en metadata te verwerven, is een inhoudelijke controle van media bestanden noodzakelijk, alleen al vanuit het beveiligingsaspect. Een controle op virusbesmetting en andere risico's kan worden ingericht door de beheerpartij (als storage service), een expliciete voorziening binnen MediaMosa hiervoor lijkt niet nodig.

Wel moet er in MediaMosa een notificatie of synchronisatie worden geïmplementeerd die dit bij het vinden en verwijderen van besmette bestanden in de MediaMosa administratie verwerkt.

In Kennisnet Groepen (2.0) is al een soortgelijke voorziening gebouwd; deze zou ook in MediaMosa toegepast kunnen worden.

Meerdere storage locaties

Op basis van het vastgestelde content type en fileformaat kan de workflow worden uitgebreid met een configureerbare keuze mogelijkheid op welke locatie bepaalde typen worden opgeslagen. Dit biedt mogelijkheden op sturing op het gebied van performance (speciale hoog performance opslag) en kosten (goedkopere hosting).

Transcodering en verwerking

De huidige opzet waarbij transcodering alleen plaats vindt na een upload (met expliciete transcodering opdracht), wordt uitgebreid. Afhankelijk van het bij de analyse vastgestelde type van een bestand, zal MediaMosa een andere verwerking starten. Voorbeelden van dergelijke verwerking is het aanmaken van stills bij video's of het omzetten van (powerpoint) presentaties naar flash.

Vergelijkbaar met de opzet bij de analyse zal per (groep van) bestandstype(n) een workflow configureerbaar worden, waarbij specifieke tool(s) aangestuurd kunnen worden voor het verwerken en/of transformeren van bestanden.

Naast de upload van 'losse' bestanden en VUF's via ftp te uploaden, zou de optie om een zip (of ander formaat) archief te kunnen verwerken op basis van een meegestuurd VUF een wenselijke uitbreiding zijn. Dit biedt de mogelijkheid om ook via andere protocollen dan ftp batch uploads te doen.

Deze uitbreiding zou ook toegepast kunnen worden voor rich content. In dat geval kan een asset opgebouwd uit meerdere samenhangende media typen in zijn geheel worden geüpload en aangemaakt. Afhankelijk van het formaat van de rich content zou de VUF instructie set uitgebreid moeten worden.

Afspelen van files

De afspeelmogelijkheden per bestandstype zijn sterk afhankelijk van de kenmerken van het desbetreffende type en de voor dit type in MediaMosa gedefinieerde workflow. De drie basis aktes zoals die nu gedefinieerd zijn: geef afspelen URL, geef embedded code en download bestand zullen per te onderscheiden (groep van) mediatype moeten worden gedefinieerd.

Content type groepering

Om een analyse te kunnen maken van de (on)mogelijkheden van ondersteuning van andere soorten content, is een indeling gemaakt van media typen. Deze indeling is gemaakt op basis van met name de beschikbare (open source) analyse tools én de beheer- en interactie-eigenschappen van de typen. De groepering en het plaatsen van bepaalde bestanden binnen deze indeling is deels arbitrair, maar biedt de benodigde handvatten om een analyse uit te kunnen voeren.

De groepering loopt van 'eenvoudige' tekst typen naar 'rich media', een categorie waarbij meerdere mediafiles van meerdere typen in samenhang uitgeleverd en afgespeeld kunnen worden. Hieronder worden de categorieën beschreven:

- Tekst: Hieronder vallen tekst documenten, maar ook bestanden met beperkte opmaak en (niet XML) wiki formaten.
- Schema: Hieronder vallen alle XML (en SGML) achtigen.
- Office/doc: De microsoft en open office document typen (.doc, .docx, .odt) en pdf.
- Office/spreadsheet: De microsoft en open office rekenblad typen.
- Office/presentatie: De microsoft en open office presentatie typen.
- Archief: Samengestelde bestanden als zip, gzip en jar.
- Afbeelding: Bekende afbeeldingformaten.
- Audio: Bekende audio formaten als .wav, .mp3.
- Video: De huidige set video formaten.
- Interactieve applicaties: Applicaties die binnen een browser uitvoerbaar zijn, als flex, flash, json, fx.
- Rich media: Specifieke formaten van samenhangende data, voorbeeld: SCORM⁶.

In de volgende tabel wordt een overzicht gegeven van de uitkomsten van deze analyse.

⁶ http://en.wikipedia.org/wiki/Sharable_Content_Object_Reference_Model

Contentgroep	Typen	Tools	Metadata
algemeen	alle	file	owner, filenaam, file datum
tekst	txt, rich text, (media)wiki		lastig te bepalen
schema	xml, xsd, html	parser, schema check	afhankelijk van schema
office	OOXML, ODF, PDF	office naar PDF converter, metadata extractor	uitgebreide schema's vooral gericht op 'gebruikersdata'
archieff	zip, rar, arj, jar	diverse	afhankelijk van type: variërend van technisch (compressie methode) tot inhoudelijk (manifest files binnen archief)
afbeelding	gif, jpeg, png, svg/xml, tiff	ImageMagick	afhankelijk van type en gebruik: nadruk meestal op technisch (kleurdiepte) tot inhoudelijk (tags, geodata)
audio	wav, wma, vorbis, mp3	SoX	doorgaans technisch, maar inhoudelijk wordt ook ondersteund door formaten (titel, artiest, album)
video	wmv, flash, mpeg4/h264, Quicktime	FFmpeg	doorgaans technisch
applicatie	flex, flash, silverlight, fx, json	meestal beperkt tot 'afspeel' ondersteuning	doorgaans technisch
rich	scorm	binnen ELO's, weinig tooling nog beschikbaar	afhankelijk van formaat (en kennis daarover)

Tabel 1 Indeling bestandstypen in contentgroepen

Tijdens deze analyse kwamen drie open source tools naar voren om toegepast te worden binnen MediaMosa. Naast het al gebruikte FFmpeg zijn dit ImageMagick⁷ voor het analyseren van afbeeldingen en SoX⁸ voor het verwerken van audio bestanden.

In de overige categorieën werden wel tools gevonden, maar kan op basis van deze eerste analyse nog geen keuze worden gemaakt.

⁷ <http://www.imagemagick.org/>

⁸ <http://sox.sourceforge.net/>

Conclusies en aanbevelingen

De huidige implementatie van MediaMosa is gericht op het verwerken van videobestanden. Bij het uitbreiden van de functionaliteit met ondersteuning voor andere content, kan aangesloten worden op het al aanwezige job mechanisme. Uitgangspunt blijft hierbij dat MediaMosa zelf de geuploadede mediafiles blijft analyseren.

Een tweede uitgangspunt blijft dat EGA's verantwoordelijk blijven voor de samenhang binnen assets. MediaMosa kent geen typering van assets, aan media files wordt wel (op basis van de analyse) een type toegekend. Voor rich media kan een hiërarchie van assets worden toegepast om de samenhang van media bestanden bij één hoofd asset op te slaan in de repository.

Om andere content effectief te ondersteunen zal de huidige hard coded workflow moeten worden vervangen door een configureerbare workflow, waarbij externe tools kunnen worden ingeschakeld voor het analyseren en verwerken van mediabestanden. De metadata set die op basis van de analyse wordt aangemaakt moet 'permanent' worden opgeslagen bij de mediafile.

Daarnaast zal de verwerking moeten worden gesplitst in twee zelfstandige workflows voor de analyse en voor de core. Voor het analyseren van meerdere contenttypen zijn meerdere stappen met meerdere tools nodig, waardoor een eigen workflow nodig is. In de MediaMosa Core is een workflow nodig die op basis van de uitkomsten van de analyse, de verdere verwerking en opslag aanstuurt.

Voor de Core workflow geldt dat deze bepaald wordt door het gewenste gedrag in de EGA. Daarom moet deze beheerd en aangestuurd worden vanuit de EGA via de REST interface. Bij het uploaden van bestanden kan een EGA een workflow profiel meegeven. Dit profiel is vooraf door de EGA (via de REST interface) gedefinieerd in MediaMosa.

MediaMosa biedt drie afspeelmogelijkheden voor mediafiles (afspelen via play proxy, HTML code met link naar play proxy voor embedding en download) die ook bij andere media typen ondersteund zullen worden. De configureerbare workflow moet de mogelijkheid bieden voor specifieke (groepen van) content typen hier per afspeelmogelijkheid een aanpassing voor te bereiden (bijvoorbeeld een powerpoint presentatie wordt door de play proxy aangeboden in flash formaat).

De verwerking met een configureerbare gesplitste workflow zal de flexibiliteit en toekomstvastheid van MediaMosa versterken. De beschreven groepering van media typen en de aanbevolen tools kunnen als basis dienen voor het uitwerken van de

workflow. In hoeverre de (standaard) mogelijkheden van MediaMosa hierbij moeten worden uitgewerkt zal uiteindelijk bepaald worden door het doel en de wensen van gebruikers van de EGA's.

Een eerste logische stap zou het uitbreiden en inrichten van MediaMosa zijn met het verwerken van afbeeldingen en audio bestanden (naast de huidige video's). Voor deze content typen zijn goede tools gevonden voor het analyseren en verwerken. In een verdiepend onderzoek (spike) kan voor de overige typen een set van tools worden gevonden die de analyse en verwerking binnen MediaMosa kunnen afdekken.