

Overige vraagstukken

Versie 1.0 12-4-2016

Anje van der Lek

Het eerste projectresultaat, de informatieanalyse, heeft ons een dieper inzicht opgeleverd in de technische en archivistische structuur van de Videotulen. Dit leverde soms ook vragen en discussies op die in een eerdere fase nog niet aan de orde waren. In dit document geven we een overzicht van de (nieuwe) vraagstukken waar we tot nu toe in dit project zijn geconfronteerd.

1 Archiefstatus Documenten

Discussiepunt: moeten de bijlagen bij de vergadering (vergaderstukken/documenten) volledig worden beschreven?

In deze dataset zijn “documenten” de tekstbestanden (PDF) die bij als bijlagen/vergaderstukken bij de agenda’s op de website worden gepubliceerd.

In het Raadsinformatiesysteem zijn minimale metadata over de documenten vastgelegd. Alleen de naam van het stuk en bij welke vergadering of agendapunt het hoort, is in deze context van belang.

De originele documenten zullen vanuit Corsa op een later moment worden gearchiveerd en in het e-depot worden opgenomen. Dan zullen deze documenten van een complete set metadata worden voorzien. Hiermee worden ze dus dubbel gearchiveerd. Dit is in principe niet anders dan in de papieren situatie, waar raadstukken ook losstaan van stukken van de ambtelijke organisatie.

Het maken van een technische koppeling tussen de vergaderagenda’s en de stukken in hun volledige context zou technisch erg complex zijn. Het heeft geen meerwaarde voor de toegankelijkheid van de informatie.

Conclusie: Wij zien de documenten die onderdeel uitmaken van de dataset “Videotulen” als “bijlagen bij de vergadering” en het toevoegen van metadata over het document (zoals auteur, datum, versie etc) is in die context niet nodig.

1.1 Archiefstatus Back-up

Discussiepunt: moeten de audiovisuele bestanden die Den Helder als Back-up opslaat worden bewaard?

We hebben gezien dat de audiovisuele bestanden op servers van de gemeente Den Helder worden opgeslagen. Een belangrijke aanleiding voor dit project was het feit dat de backupvoorziening vol raakte en er moet worden geïnvesteerd in nieuwe hardware.

In de voorbereidingsfase van dit project gingen we er van uit dat de video ongecomprimeerd werd opgeslagen in de backup, en daarom misschien beter geschikt was om te bewaren als de gecomprimeerde video die door GO wordt opgeslagen.

In dit onderzoek is echter het volgende gebleken:

1. De videobestanden in de backup zijn opgeslagen in het Transport Stream bestandsformaat (TS) dit is een formaat dat gebruikt wordt voor DVD's en maakt gebruik van de MPEG-2 compressie. De videobestanden zijn dus ook gecompriemd.
2. Het bestandformaat dat door GO uitgeleverd (MP4) wordt is voldoende geschikt voor lange termijn archivering.
3. We hebben vastgesteld dat de audiovisuele bestandsformaten onlosmakelijk zijn verbonden met de agenda-metadata (o.a. agendapunten en pointers). De audiovisuele bestanden in de backup zijn niet gekoppeld aan deze metadata. De door GO geleverde audiovisuele bestanden zijn wel gekoppeld met de agenda-metadata, en hebben daardoor een betere archief-kwaliteit.

Conclusie: Na een succesvolle overdracht van de digitale raadsinformatie kan de backup dan ook verwijderd worden.

1.2 Webarchivering van Videotulen.

Discussiepunten: is het archiveren van de website van de gemeenteraad een goede methode voor het archiveren van Videotulen?

Tijdens het project is ook naar de mogelijkheid van webarchivering van videotulen gekeken. Hiermee wordt bedoeld dat de website waar de videotulen op worden gepubliceerd wordt gearchiveerd, op zo'n manier dat de links en audio / videobestanden nog net zo werken als op de "live" website die is gekoppeld aan het Raadsinformatiesysteem van GemeenteOplossingen (RIS/GO). Wij zijn van mening dat het archiveren van de Videotulen als website echter geen goede methode is om deze duurzaam toegankelijk te maken:

1. Webarchivering is geen oplossing voor preservatie van videobestanden

Een probleem bij webarchivering is het toegankelijk houden van de "embedded content", de plaatjes, video's en audiobestanden die via links in de website te benaderen zijn. Als een bepaald bestandsformaat in onbruik is geraakt, zoals nu bijvoorbeeld flash of eerder real-video, is deze embedded content niet meer af te spelen / raadplegen. In het webarchief zie je dan een foutmelding op de plaats waar de video af zou moeten spelen.

2. Metadata kan niet op het goede niveau worden toegekend

Een webarchief is een samengesteld digitaal object, dat verschillende bestanden bevat. In het geval van videotulen worden dan alle audio en videobestanden en PDF/A documenten samen met de HTML-code van de website in één file opgeslagen. Dus alle vergaderingen van 10 jaar zitten in één pakketje dat via een browser kan worden geopend.

Bij het overdragen aan een e-depot van dit "pakketje" kunnen dan niet de benodigde TMLO-metadata aan het archief worden meegegeven. Metadata die nodig is voor het vindbaar maken, beheren en preservatie van digitale informatie, moet op verschillende aggregatieniveaus worden toegekend. Daarmee wordt bedoeld, dat er informatie over de reeks vergaderingen, individuele

vergadering, maar ook informatie over elk individuele digitale bestand moet worden meegegeven. Omdat al deze verschillende bestanden in één file worden “ingepakt”, is het niet mogelijk om alle benodigde metadata op het juiste niveau mee te geven.

Discussiepunt: Maar is het dan niet vereist dat de originele website wordt gearchiveerd?

De audio/videoregistraties, documenten en de verbindende agenda-metadata vormen samen een dataset. Deze dataset is de neerslag van het vergaderproces. De website is een *interface* om deze dataset te raadplegen. Deze twee elementen, dataset en interface, moeten los van elkaar worden gezien.

De interface is niet bepalend voor de authenticiteit van de dataset. Als de dienstverlening van GO aan de gemeente Den Helder stopt, is website zoals deze nu is niet meer te gebruiken. Als men via een andere toegang -bijvoorbeeld een archiefportal- een video uit de overgebrachte dataset zou raadplegen, is dit nog steeds een authentieke registratie van de vergadering.

Conclusie: het archiveren van de website is niet vereist, en webarchivering biedt als methode niet genoeg mogelijkheden voor het toekennen van metadata aan het bestand Videotulen.

1.3 Toegang / interface

Discussiepunt: Moet het bouwen van een webinterface / toegang tot de Videotulen ook deel uitmaken van dit project?

Om het project minder complex te maken, is het maken van een toegang / interface op het archiefbestand “Videotulen” bewust buiten het project gehouden (*Projectplan p.8*)

In de projectdoelstelling was over toegankelijkheid opgenomen dat we *de voorwaarden scheppen om deze informatie op een later moment onafhankelijk van de leverancier toegankelijk te maken.* (Projectplan p.4)

Uitgangspunt was, dat in de informatieanalyse wordt vastgesteld hoe de dataset in elkaar zit en welke informatie essentieel is voor de zoek- en raadpleegmogelijkheden die de webinterface nu biedt. Als deze essentiële informatie niet verloren gaat, zou het mogelijk zijn om op een later moment een interface te maken met ten minste dezelfde (en misschien meer) functionaliteiten.

1. Dit uitgangspunt is moeilijk te testen zonder ook werkelijk een interface te bouwen.
2. Randvoorwaarde was, dat de toegankelijkheid via GO nog in stand zou blijven, en directe toegang tot het gearchiveerde bestand daarom nog niet nodig zou zijn. Het contract van Den Helder met GO loopt tot 2017 en het is niet zeker of dit wordt gecontinueerd. Het is daarom wellicht toch nodig om de toegankelijkheid van het archiefbestand op korte termijn te regelen.
3. Uit andere projecten is gebleken, dat het ontbreken van een zichtbare toegang wordt ervaren als een tekortkoming van het project. Het bieden van zo’n toegang is dus een succesfactor.

Conclusie: Het buiten het project houden van de toegang tot de videotulen heeft nadelen. De stuurgroep kan overwegen om het maken van een toegang toch uit te voeren.

1.4 Visie metadata

Discussiepunt: moeten de gestructureerde data uit het RIS in hun huidige vorm en structuur behouden blijven, of kan/moet alle data worden “gemapt” naar een metadata-element uit het TMLO?

De uitkomsten van dit discussiepunt zijn in behandeld in het document “Ontwerp aan te leveren informatiebestand”, de tekst hieronder is overgenomen uit dit document (Ontwerp p.7):

Onderscheid tussen metadata en gestructureerde informatieobjecten

Metadata wordt meestal gedefinieerd als “informatie over informatie”. In de praktijk is het onderscheid tussen “data” of “informatie” en “metadata” niet altijd makkelijk te maken. Zo is het in dit geval niet correct om de audiovisuele bestanden (MP4’s) te zien als de “te archiveren informatieobjecten”, en de gestructureerde informatie uit het RIS (de XML-export) als de daarbij horende metadata.

Het RIS bevat de agenda-informatie van vergaderingen en pointers naar een bepaald moment in de vergadering (agendapunt of spreker). De informatie in het RIS is dus niet puur informatie óver de audiovisuele bestanden, maar een informatieobject op zichzelf. Het is een dataset die de dynamische presentatie van agenda’s op de website mogelijk maakt.

In de context van het e-Depot maken we het volgende onderscheid:

- *metadata*: technische, administratieve of beschrijvende informatie óver een archiefbestand. Deze metadata worden gestandaardiseerd geleverd in ToPX/XML bestanden met de extensie “.metadata”. Deze metadata maakt het mogelijk om eenduidig te zoeken in het e-depot, en faciliteert tevens het duurzame beheer van de informatie. In paragraaf **Fout! Verwijzingsbron niet gevonden.** en **Fout! Verwijzingsbron niet gevonden.** [Ontwerp p. 7-9] wordt verder ingegaan op hoe deze metadatabestanden samengesteld dienen te worden.
- *gestructureerde informatieobjecten*: te archiveren informatieobjecten die bestaan uit gestructureerde, machineleesbare informatie. In dit project betreft dit de XML-export uit de database van het RIS. In hoofdstuk **Fout! Verwijzingsbron niet gevonden.** [Ontwerp p. 10-11] komen we hier op terug.

De XML-export wordt als een apart record opgenomen in het e-depot. Dit bestand bevat data die niet direct te mappen is met het TMLO/ToPX schema, zoals “pointers” (=tijdsmarkeringen die aangeven wanneer een agendapunt of bepaalde spreker begint) en gegevens (naam, partij) over sprekers. Deze data moeten wel bewaard blijven om een nieuwe toegang / viewer op te kunnen baseren.

Conclusie: De gestructureerde data uit het RIS in hun huidige vorm en structuur behouden blijven. Ze worden als XML bestand in het e-depot opgenomen. Daarnaast bevatten alle in het e-depot op te nemen bestanden ook een set metadata aangeleverd in het TMLO/ToPX formaat.