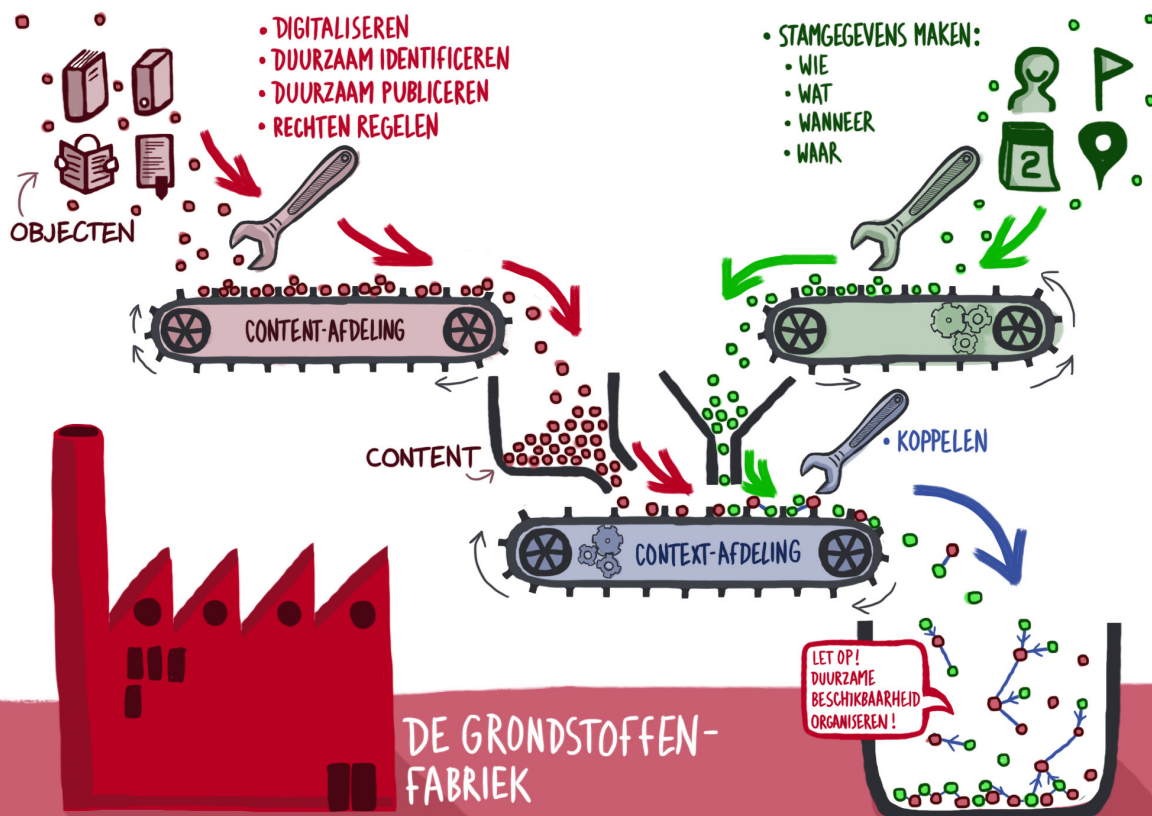


# Men neme een collectie ... en dan?

## Overzicht van informatietechnologie voor het toegankelijk maken van oorlogsbronnen

# Netwerk Oorlogsbronnen



DENKSCHETS.NL

Amsterdam, 15 oktober 2015



Ministerie van Volksgezondheid,  
Welzijn en Sport



Het rapport *Men neme een collectie ... en dan? Overzicht van informatietechnologie voor het toegankelijk maken van oorlogsbronnen* is het eindresultaat van werkpakket 4 'Zoeken, vinden en verrijken'. Dit onderzoek was onderdeel van het project Kenniscentrum Oorlogsbronnen dat is uitgevoerd van 5 januari tot 18 december 2015. Een werkgroep bestaande uit vertegenwoordigers uit het erfgoedveld was nauw betrokken. De namen van de leden van de werkgroep zijn opgenomen in bijlage A.

**Tekst:** Ivo Zandhuis / [www.zandhuis.nl](http://www.zandhuis.nl)  
**Illustratie voorkant:** Elco van Staveren / [www.denkschets.nl](http://www.denkschets.nl) (zie p. 77)  
**Opmaak:** Ellen Bouma / [www.ellenbouma.nl](http://www.ellenbouma.nl)  
**Opdrachtgever:** NIOD – project Kenniscentrum Oorlogsbronnen / [www.oorlogsbronnen.nl](http://www.oorlogsbronnen.nl)

# 1. Inhoud

<b>1.</b>	<b>Inleiding</b>	<b>5</b>
1.1	Grondstoffen	5
1.2	Beheer, behoud, beschikbaarstelling van grondstoffen	6
1.3	Gebruiksaanwijzing van dit overzicht van ontsluitingsstrategieën	7
<b>DEEL I: Grondstoffen</b>		
<b>2.</b>	<b>Digitale informatie</b>	<b>9</b>
2.1	Verschillende soorten digitale informatie	9
2.2	Metadata	11
2.3	Digitale manifestaties	13
2.4	Stamgegevens	14
2.5	User Generated Content	15
<b>3.</b>	<b>Metadata maken</b>	<b>16</b>
3.1	Catalogiseren/inventariseren: formeel beschrijven	16
3.2	Catalogiseren: inhoudelijk beschrijven	18
3.3	Inhoudelijk metadata geautomatiseerd bepalen	20
3.4	Rechten vastleggen	21
3.5	Fragmenten markeren	24
3.6	Preservation metadata vastleggen	25
3.7	Verbetervoorstellen van het publiek verwerken	26
<b>4.</b>	<b>Digitale manifestaties maken</b>	<b>28</b>
4.1	Bronnen selecteren	28
4.2	Digitale reproductie maken	30
4.3	OCR-en	33
4.4	Spraakherkenning	34
4.5	Data entry	35
4.6	(Semi-)automatisch gestructureerde data maken	36
4.7	Verbetervoorstellen van het publiek verwerken	38
<b>5.</b>	<b>Stamgegevens maken en ernaar verwijzen</b>	<b>39</b>
5.1	Authority-files gebruiken	39
5.2	Authority-files maken en publiceren	40
5.3	Thesauri gebruiken	42
5.4	Thesauri maken en publiceren	43
5.5	Plaatsnamenlijsten gebruiken	44
5.6	Plaatsnamenlijsten maken en publiceren	44
5.7	Verbetervoorstellen van het publiek verwerken	45
<b>6.</b>	<b>User Generated Content</b>	<b>46</b>
6.1	Eigen bijdragen vastleggen en beschikbaar stellen	46
6.2	Discussie faciliteren	47

## DEEL II: Beheer, behoud en beschikbaarstelling van grondstoffen

<b>7.</b>	<b>Het applicatielandschap</b>	<b>49</b>
7.1	Een drielagenmodel voor erfgoedontsluiting	49
7.2	De data laag beheert en behoudt	50
7.3	De verbindingslaag brengt samen en verbindt	50
7.4	De presentatielaag zorgt voor interactie met publiek	51
7.5	Samenwerking tussen de lagen	51
7.6	Toepassing bij een instelling	51
7.7	Toepassing in een domein	52
<b>8.</b>	<b>Beheren en behouden</b>	<b>53</b>
8.1	Architectuur rondom beheer en behoud	53
8.2	Inrichten van metadatabeheer	54
8.3	Inrichten van opslag	55
8.4	Inrichten van duurzame opslag	56
8.5	Persistent identifiers maken	59
8.6	Reacties opslaan	61
8.7	Aanbieden aan website of aggregator	62
8.8	Enkele juridische aspecten	63
<b>9.</b>	<b>Samenbrengen en verbinden</b>	<b>65</b>
9.1	Redactie voeren	65
9.2	Ophalen van grondstoffen	66
9.3	Verbeteren van grondstoffen	67
9.4	Diensten ontwikkelen en aanbieden	68
9.5	Enkele juridische aspecten	69
<b>10.</b>	<b>Interactie met publiek</b>	<b>70</b>
10.1	Vinden	71
10.2	Bekijken	71
10.3	Gebruiken	72
10.4	Verrijken	73
10.5	Gebruiksgegevens vangen	74
10.6	Enkele juridische aspecten	75
<b>11.</b>	<b>Conclusie</b>	<b>77</b>
	<b>Bijlages</b>	<b>79</b>
	A: met dank aan ...	80
	B: relatie met DE BASIS	81
	C: literatuur	82

# 1. Inleiding

Collectiebeheerders en samenwerkende instellingen kunnen de toegang tot oorlogsbronnen verbeteren door middel van verschillende activiteiten. Dit rapport biedt een overzicht van deze acties. Moderne collectiebeheerders en medewerkers die hiervan kennis nemen zijn in staat om te bepalen welke activiteit hij of zij kan uitvoeren om toegankelijkheid zo effectief mogelijk te verbeteren. De meest effectieve acties kosten het minste moeite, tijd en geld, en dragen maximaal bij aan de toegankelijkheid van het materiaal op internet.

De basis voor toegankelijkheid is de informatie over en afgeleid van de bronnen waarvoor zoeken, vinden en verrijken gewenst is. Daarom geeft hoofdstuk 2 een overzicht van de verschillende soorten informatie waarover we het hebben. Hoofdstukken 3, 4, 5 en 6 omschrijven acties waardoor deze informatie ontstaat.

Hoofdstuk 7 geeft een overzicht van de soorten applicaties die een nodig zijn om de informatie te beheren, behouden en beschikbaar te stellen. Hoofdstukken 8, 9 en 10 gaan in op respectievelijk de informatiesystemen gespecialiseerd in (1) beheer en behoud, (2) samenbrengen en verbinden, en (3) interactie met publiek.

## 1.1 Grondstoffen

Allereerst is het noodzakelijk om na te gaan of er voldoende digitale informatie beschikbaar is om het vinden mogelijk te maken. Deze digitale informatie vormt als het ware de grondstof voor het zoeken, vinden en verrijken.

### **Metadata**

Belangrijke grondstof voor het zoeken en vinden van bronnen is de metadata: data *over* de bronnen. Het rapport onderzoekt welke soorten metadata bijdragen aan de toegankelijkheid van materiaal. Hoe metadata tot stand komt is het onderwerp van hoofdstuk 3.

### **Digitale manifestaties**

Naast de digitale informatie *over* de bronnen kan de collectiebeheerder er voor kiezen de bronnen zelf digitaal aan te bieden. We spreken van digitale manifestaties. Hiertoe rekenen we zowel de scans, de digitale audio- en video-bestanden, de transcripties als de data in de bron zelf.

Als bijvoorbeeld een tekst helemaal digitaal beschikbaar is, kan een zoekmachine ingezet worden om op alle woorden in de tekst te zoeken. Als de tekst niet digitaal beschikbaar is, is de eerste actie daarom het realiseren van deze tekst. Nu is het mogelijk deze tekst over te (laten) typen, maar er zijn ook technieken om dit geautomatiseerd te laten verlopen. Voor een gedrukte tekst is dat Optical Character Recognition (OCR).

Welke acties leiden tot digitale manifestaties is omschreven in hoofdstuk 4.

## Stamgegevens

De metadata verwijst bij voorkeur naar gemeenschappelijke lijsten van auteurs, onderwerpen, plaatsnamen en dergelijke. Sommige collectiebeheerders spreken van terminologiebronnen of waardenlijsten. In dit rapport kiezen we er voor deze lijsten stamgegevens te noemen. Een authority-file (voor auteursnamen), een thesaurus (voor onderwerpsaanduidingen) of een gazetteer (voor geografische namen) zijn allemaal specifieke vormen van dit soort lijsten. In hoofdstuk 5 omschrijven we hoe een collectiebeheerder deze stamgegevens kan maken.

## User Generated Content

Behalve de informatie die een collectiebeheerder maakt, is er veel informatie beschikbaar die het publiek heeft gemaakt. Deze informatie noemen we *User Generated Content*. Ook hiervan zijn vele varianten denkbaar: verhalen, opmerkingen, discussies; het publiek laat van zich horen op verschillende manieren. Dit is het onderwerp van hoofdstuk 6.

## 1.2 Beheer, behoud, beschikbaarstelling van grondstoffen

### Beheer en behoud

Digitale informatie creëren is één ding, deze op een goede manier beheren een tweede. Een erfgoedinstelling gebruikt daarvoor verschillende soorten digitale informatiesystemen. Er is bijvoorbeeld een Collectie Beheer Systeem noodzakelijk, die alle metadata over de collectie bevat. Digitaal depot, Trusted Digital Repository, e-Depot zijn namen voor systemen die het duurzame behoud van informatie organiseren.

Een architectuur beschrijft de samenhang tussen de systemen. En deze architectuur moet aansluiten bij afspraken over bijvoorbeeld de omgang met en de vervanging van de informatiesystemen. In hoofdstuk 8 staan verschillende soorten standaardtoepassingen beschreven, die in een dergelijke architectuur een rol kunnen spelen om het beheer en behoud van de digitale informatie te organiseren.

### Samenbrengen en verbinden

De gecreëerde en beheerde informatie is echter nog steeds niet bij het publiek beschikbaar om het te kunnen zoeken, vinden of verrijken. De digitale informatie moet zodanig uit de beheersystemen beschikbaar zijn, dat er een website mee kan worden gebouwd. Daarnaast is het mogelijk grondstoffen aan te bieden aan een “aggregator”. Een aggregator brengt digitale informatie uit verschillende informatiesystemen samen en maakt gebruik van de verbindingen die collectiebeheerders en publiek hebben gelegd. Deze verbindingen bieden context aan de bronnen en zijn zichtbaar in de aggregator. Daarmee helpen ze bij het zoeken en vinden van materiaal.

Een aggregator heeft nog een ander belangrijk voordeel: het publiek kan op zoek naar materiaal zonder vooraf kennis te hoeven hebben van het instituut waar hij moet zoeken. Hoofdstuk 9 omschrijft welke technieken kunnen worden gebruikt om een aggregator in te zetten bij het verbeteren van zoeken, vinden en verrijken.

## Interactie met het publiek

Naast de systemen waarin de collectiebeheerder zijn grondstoffen beheert en de aggregators die de informatie samenbrengen, zijn er verschillende mogelijkheden om de informatie beschikbaar te stellen. Een website biedt bijvoorbeeld de mogelijkheid om het publiek te laten zoeken in de beschikbare bronnen. In een ander voorbeeld kan een app het publiek voorzien van originele bronnen tijdens een stadswandeling. Een onafhankelijke partij kan een online omgeving inrichten waarmee publiek informatie “cureert” - waarmee het publiek zelf materiaal kan verzamelen, groeperen en van aantekeningen voorzien. Hoofdstuk 10 geeft een overzicht van de technieken die een collectiebeheerder hier kan inzetten.

### 1.3 Gebruiksaanwijzing van dit overzicht van ontsluitingsstrategieën

De gebruiker van het rapport doorloopt de verschillende acties en onderzoekt of deze een effectieve bijdrage leveren aan het toegankelijker maken van een bepaalde collectie.

#### Voorbeeld

De vrijwilliger van een historische vereniging gaat aan de slag met de vraag hoe een lokaal ontdekt verzetsblaadje beter toegankelijk kan worden gemaakt. De vrijwilliger kent het overzicht van mogelijke acties en vraagt zich af: “heeft iemand het materiaal gescand?” Het antwoord blijkt ja, vervolgens controleert zij of de kwaliteit van de scans voldoende is. Vervolgvraag is: “waar zijn de scans opgeslagen?” Ze blijken op een harde schijf op het clubhuis te liggen. Hier valt iets te verbeteren aan het duurzame beheer. De vrijwilliger gaat na: “is er een collectieregistratie?” Dat wel. Iemand houdt een spreadsheet bij met al het materiaal dat de vereniging beheert. De spreadsheet is beschikbaar op de website van de vereniging en wordt op de webserver vervangen zodra er nieuwe informatie beschikbaar komt. Ook hier is iets te verbeteren aan het duurzame beheer van de informatie. Eerste activiteit die de vrijwilliger gaat uitvoeren is het inrichten van een omgeving waarin de scans en de metadata op een goede manier kunnen worden opgeslagen en beheerd. Van daaruit kan de vrijwilliger het materiaal beschikbaar stellen. Dit is effectiever dan bijvoorbeeld het materiaal te OCR-en. Het OCR resultaat zal namelijk in dit geval waarschijnlijk op harde schijven in de kast terecht komen, naast de harde schijven van de scans.

# DEEL I: Grondstoffen



## 2. Digitale informatie

Betere toegankelijkheid van bronnen is een samenspel van allerlei vormen van digitale informatie die van deze bronnen is afgeleid. Om de activiteiten voor betere toegankelijkheid goed te kunnen beschrijven, is het van belang deze verschillende soorten grondstoffen te onderscheiden.

### 2.1 Verschillende soorten digitale informatie

Voor de verschillende grondstoffen maken we onderscheid tussen *metadata* – informatie over een bron – en *digitale manifestaties* – een digitale weergave van de inhoud van de bron. Metadata wordt beter indien de collectiebeheerder gebruik maakt van *stamgegevens*: verwijzingen naar lijsten met bijvoorbeeld persoonsnamen, plaatsnamen en onderwerpen. En het gebruik van bronnen leidt tot informatie die is gemaakt door het publiek: *user generated content*. In deze paragraaf leggen we deze begrippen uit en in de volgende paragrafen zoomen we er verder op in.

#### Metadata

Van een boek is belangrijk om een beschrijving van het boek beschikbaar te hebben, die het voor een persoon mogelijk maakt dat in een gesprek meteen duidelijk is waarover het gaat. Met behulp van de tekst “*Oeroeg door Hella Haasse*” wordt in één keer duidelijk waarover we het hebben. Ook de signatuur van het boek helpt om duidelijk te maken over welk boek we het hebben. Deze informatie noemen informatieprofessionals de formele ontsluiting. Daarnaast wil een collectiebeheerder informatie vastleggen over het onderwerp, in dit voorbeeld “vriendschappen”. Informatieprofessionals noemen dit de onderwerpsontsluiting (Schneiders 1984). Op de beschikbaarstelling van een publicatie kunnen verschillende soorten rechten betrekking hebben. In dit geval is met name het auteursrecht van toepassing. Hierover moet informatie worden vastgelegd, zodat bij het beschikbaar stellen van een digitale manifestatie rekening met deze rechten kan worden gehouden.

Al deze soorten metadata hebben betrekking op het informatie-object zelf en zijn onafhankelijk van de manifestatie. Over de manifestatie is allerlei informatie van belang die het gebruik en de bewaring van de manifestatie mogelijk maakt: de preservation metadata.

Er is dus metadata die verschillende informatie *over* een bron vastlegt:

1. Formele metadata (formele ontsluiting)
2. Inhoudelijke metadata (onderwerpsontsluiting)
3. Informatie relevant voor rechten
4. Preservation metadata

#### Digitale manifestatie

De term *digitale manifestatie* wordt gebruikt voor elke digitale vorm waarin de inhoud van materiaal zelf is opgeslagen: een scan, een volledige weergave van een tekst, een

overzicht van de data in de vorm van een spreadsheet (Helwig and Bisscheroux 2015) (IFLA 1998). Een digitale manifestatie van *Oeroeg* is op internet niet beschikbaar.

Digitale manifestaties kunnen dus verschillende vormen aannemen:

5. Digitale reproducties (scans, audio-files)
6. Transcripties (soms “machineleesbaar” genoemd)
7. Gestructureerde data (de inhoud van een bron in de vorm van een tabel)

### **Stamgegevens**

De metadata bevatten informatie die uit een standaardlijst kan worden geselecteerd, zoals een naam van de auteur (bv. “Hella S. Haasse”). Een dergelijke lijst van gestandaardiseerde namen wordt een *authority-file* genoemd. Daarnaast kan een gestandaardiseerde term voor een onderwerp (bv. “vriendschap” in plaats van “vriendschappen”) worden toegekend. Voor dit soort lijsten wordt de term *thesaurus* of *onderwerpsclassificatie* gebruikt (Foskett 1969).

Als een verzamelterm voor thesauri, authority-files of bijvoorbeeld plaatsnamenlijsten, gebruiken we in dit rapport de term “stamgegevens”. Dit begrip lenen we uit de boekhoudsoftware. Software-ontwikkelaars gebruiken deze term in dat domein voor die gegevens die steeds hetzelfde zijn en waarnaar in boekingen kan worden verwezen, zoals adresgegevens van vaste klanten en leveranciers.

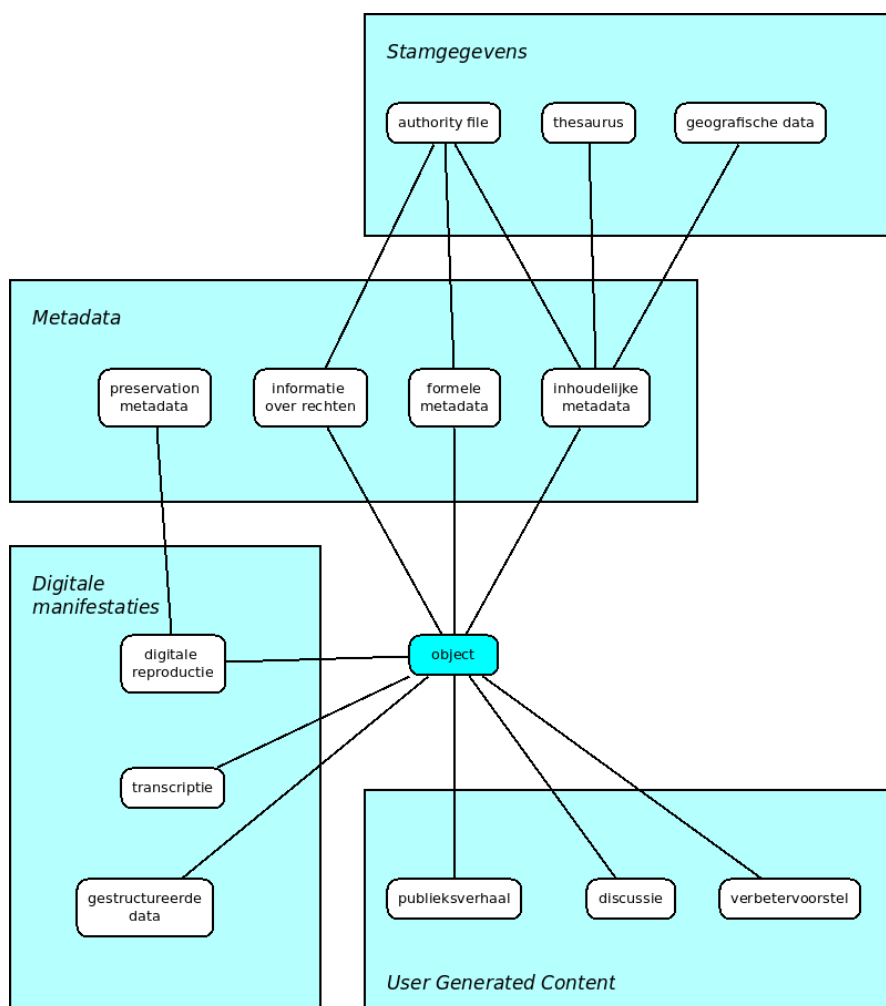
8. Thesauri
9. Authority files
10. Geografische data oftewel gazetteer

### **User Generated Content**

De gebruiker van de bron kan zelf op allerlei manieren informatie over de bron toevoegen. De lezer van *Oeroeg* kan bijvoorbeeld een boekverslag schrijven en deze publiceren als blog. Hij kan op Twitter of Facebook vertellen dat hij het boek leest. Een andere gebruiker maakt wellicht een samenvatting op wikipedia ([https://nl.wikipedia.org/wiki/Oeroeg\\_%28boek%29](https://nl.wikipedia.org/wiki/Oeroeg_%28boek%29)).

Er zijn verschillende soorten reacties denkbaar, die we in dit rapport scharen onder *User Generated Content*:

11. Publieksverhalen
12. Discussies
13. Verbetervoorstellen



Het bovenstaande schema geeft weer wat de relaties zijn tussen de verschillende grondstoffen. De samenhangende informatie helpt bij het vinden van het materiaal, mits collectiebeheerders de juiste applicaties toepassen om de informatie te beheren, te behouden en beschikbaar te stellen. Welke applicaties dat zijn, is het onderwerp van deel II van dit rapport.

## 2.2 Metadata

De belangrijkste en oudste grondstof die collectiebeheerders hebben gemaakt om toegankelijkheid van bronnen te verbeteren, is metadata: data *over* de bron.

### Formele metadata

Formele metadata zorgt voor ‘een korte beschrijving die volgens vaste regels zodanige kenmerken van documenten opsomt, dat deze als eenheid kunnen worden onderkend en van andere onderscheiden’ (Schneiders 1984:196).

Het onderscheiden van een object van ander materiaal is een belangrijke functie, omdat daarmee een onderzoeker aan een document of voorwerp kan refereren. Daarnaast geven de gekozen kenmerken een indruk van de context van het object. Wanneer is het gemaakt of verschenen? Wie heeft het archiefstuk gemaakt en waarom? Waar heeft het maken van de tekening plaatsgevonden? Maakt het object onderdeel uit van een groter geheel, zoals een archief, een collectie of een tijdschriftenserie?

De formele metadata bestaat uit verschillende metadatavelden die in een metadata-model zijn gedefinieerd. Elke soort bron heeft zijn eigen velden, gestandaardiseerde namen en zijn eigen normen voor het invullen van de velden. Archieven hebben immers andere eigenschappen dan boeken of afbeeldingen.

Bij het beschrijven van archieven maakt een collectiebeheerder onderscheid tussen beschrijvingen van verschillende groepen van archiefstukken. Ook een collectie van bijvoorbeeld museumobjecten als geheel heeft een eigen manier van beschrijven.

### **Inhoudelijke metadata**

Zoeken op de metadata die een object formeel beschrijven, heeft vooral zin als een gebruiker het object al kent. Hij maakt dan gebruik van de woorden die in een beschrijving zijn opgenomen.

Gebruikers willen echter vaak een bron vinden die gaat over een bepaald persoon (“wie”), een bepaald onderwerp (“wat”), een bepaalde plaats (“waar”), en/of een bepaalde periode (“wanneer”). Soms bevat de formele metadata van de bron deze informatie, maar dat is niet gegarandeerd. En waar het boek is gepubliceerd, kan hetzelfde zijn als de plaats waar het boek over gaat, maar ook dat is niet zeker.

De onderwerpsaanduiding die de collectiebeheerder vastlegt en koppelt aan het object is subjectief van aard: elke collectiebeheerder zal andere onderwerpskenmerken aan een object willen toekennen, die volgens hem of haar het onderwerp het beste weergeeft. Bovendien verstaat het publiek misschien wel hele andere dingen onder termen die de collectiebeheerder toekent. Ook kunnen termen in de toekomst van betekenis veranderen, wat invloed heeft op de vindbaarheid. De “staatsregeling” van 1798 zouden we tegenwoordig een “grondwet” noemen. In de omschrijving van deze staatsregeling komt het woord grondwet niet voor.

### **Informatie relevant voor rechten**

Het (al dan niet digitaal) beschikbaar stellen van materiaal is gebonden aan wetten. Is het materiaal openbaar of rust er geheimhoudingsplicht op? Is het materiaal privacygevoelig, zodat er speciale afspraken moeten worden gemaakt voordat de collectiebeheerder het materiaal aan publiek ter inzage geeft? Belangrijke vraag is: bij wie berust het auteursrecht? Welk auteursrecht is dat? En geeft dat auteursrecht de mogelijkheid om het materiaal zonder meer ter beschikking te stellen.

Bij het object is het daarom noodzakelijk informatie vast te leggen waaruit de collectiebeheerder kan afleiden of een object op een bepaalde manier beschikbaar mag worden gesteld. Te denken valt aan de rechthebbende, wanneer een recht vervalt en of er speciale afspraken zijn gemaakt.

### **Preservation metadata**

Als sprake is van een digitale bron, is het nodig extra metadata vast te leggen die betrekking heeft op zijn digitale aspecten. Je moet dan denken aan de namen en verwijzingen naar computerbestanden die gezamenlijk de bron vormen. En de volgorde

waarin de gebruiker de bestanden moet interpreteren. Als een publicatie is gescand, is de digitale bron een verzameling scans, die alleen in een bepaalde volgorde de publicatie weergeven. Het is nodig deze volgorde vast te leggen zodat een viewer straks weet welk plaatje er moet worden getoond als er op de “volgende bladzijde” wordt geklikt. Dit wordt structurele metadata genoemd.

Technische metadata legt vast hoe de digitale reproductie tot stand is gekomen. Dit helpt om in de toekomst informatie te kunnen verstrekken over de betrouwbaarheid van de digitale reproductie. Ook stelt de informatie de collectiebeheerder in staat om het duurzame beheer van de informatie te organiseren. Of hij een bestand bijvoorbeeld moet omzetten naar een ander softwareformaat omdat computer- en softwareleveranciers het oude formaat niet langer ondersteunen, blijkt uit de technische metadata. In de technische metadata ligt daarom het softwareformaat vast. Het softwareformaat is bovendien noodzakelijk om te bepalen welke tool een collectiebeheerder moet gebruiken om het document te kunnen bekijken.

Voor dit soort metadata gebruiken we de term preservation metadata.

## 2.3 Digitale manifestaties

Behalve de metadata, is een digitale manifestatie van de bron een belangrijke grondstof om de toegankelijkheid te verbeteren. Digitale manifestaties hebben als eigenschap dat ze een letterlijke weergave zijn van de inhoud van de bron. Die letterlijke weergave kan verschillende vormen aannemen.

### Digitale reproducties

Er bestaan verschillende technieken om van een analoog object een digitale reproductie te maken. Papieren bronnen resulteren in scans om op een website te tonen, of om als afbeelding af te drukken in een boek. Bij het scannen van papier bepaalt de scanner wat de kleur is op een klein deel van de bladzijde. Hoe meer punten het scanapparaat meet (in technische termen: hoe hoger de resolutie) hoe preciezer de reproductie is die de computer vastlegt.

Ook bij audiovisuele bronnen meet een scanapparaat het (analoge) geluidssignaal van een tape in zo klein mogelijke fragmenten (in technische termen: zo klein mogelijke samples). De collectiebeheerder kan de kopie van de digitale audiovisuele kopie gebruiken om beschikbaar te stellen op de website, maar ook programmamakers kunnen een fragment hergebruiken in een nieuwe televisie- of radio-uitzending.

Reproducties van museale objecten worden gemaakt door een fotograaf. Hierbij kan deze kiezen tussen verschillende technieken, waarbij het mogelijk is om een driedimensionale ervaring te benaderen.

### Transcripties

De mogelijkheid van een scan of digitale reproductie van audiovisueel materiaal blijft beperkt tot het kunnen bekijken van een bron via internet. Ook als de bron bestaat uit gedrukte woorden, is zoeken naar woorden in de tekst niet zonder meer

mogelijk. Hiervoor is een extra grondstof noodzakelijk: de volledige tekst in machineleesbaar formaat. De volledige tekst maakt het mogelijk om een woord in de tekst terug te vinden. Als iemand de tekst heeft gelezen of afgeluisterd en met de hand heeft ingevoerd, noemen we deze volledige tekst “transcriptie”. Er zijn ook technieken die dit handmatige resultaat benaderen: Optical Character Recognition (OCR) en spraakherkenning.

### **Gestructureerde data**

Sommige bronnen bevatten gestructureerde informatie. Het gaat daarbij bijvoorbeeld over handboeken met overzichten van begrippen, publicaties met lijsten van personen of registers met onroerend goed in archieven. De informatie in dit materiaal komt als digitale manifestatie het beste tot zijn recht als deze “gestructureerd” beschikbaar is. Dat wil zeggen dat de collectiebeheerder de informatie heeft opgeknipt in logische eenheden, die gezamenlijk een datamodel vormen. Deze eenheden zouden er dan uit kunnen zien als een spreadsheet of database tabel.

Gestructureerde data stelt een gebruiker in staat om de informatie in het document te combineren met andere tabellen. Ook kan nog preciezer het juiste document worden gevonden, zoals bijvoorbeeld de ontsluiting die is gemaakt om de juiste akte te kunnen vinden in de Burgerlijke Stand van 19<sup>e</sup> en 20<sup>e</sup> eeuw.

## **2.4 Stamgegevens**

Sommige grondstoffen hebben als kenmerk dat collectiebeheerders ze kunnen hergebruiken bij het vastleggen van formele of inhoudelijke metadata. De namen van auteurs bijvoorbeeld of gestandaardiseerde termen. We noemen ze stamgegevens.

### **Authority Files**

Het is onnodig om elke keer bij een formele of inhoudelijke beschrijving opnieuw alle informatie over een persoon, familie of instelling vast te leggen. Deze informatie kan een collectiebeheerder beter één keer invoeren en daarna hergebruiken. In de metadata neemt de collectiebeheerder dan een verwijzing op naar de naam die hij bedoelt. Een apart systeem bevat aanvullende informatie over de auteur, fotograaf, kunstenaar of archiefvormer, zoals hun volledige naam en geboorte- en sterfjaar.

### **Thesauri**

Een informatieprofessional legt in een thesaurus trefwoorden vast die het onderwerp van een bron kunnen aanduiden. De trefwoorden verwijzen onderling naar elkaar: een fiets is een soort voertuig, rijwiel is een ander woord voor fiets. In de inhoudelijke metadata van een foto kan een collectiebeheerder dan vastleggen dat er een “fiets” is afgebeeld. Een gebruiker die zoekt op “rijwiel” of foto’s wil van “voertuigen”, vindt dankzij de thesaurus ook de afbeeldingen die iemand heeft voorzien van de thesaurusterm “fiets”.

### **Geografische data**

Een derde vorm van stamgegevens heeft betrekking op geografische namen. Geografische specialisten kunnen deze op één plaats voorzien van coördinaten, en relaties met andere namen, zoals alternatieve spellingen of grotere geografische eenheden.

Een dergelijk systeem van stamgegevens heet in het Engels een *gazetteer*.

Voor erfgoed is het een belangrijk gegeven dat geografische namen in de loop van de tijd van betekenis veranderen: de naam heeft in verschillende periodes betrekking op een andere geografische eenheid, of een andere bestuurlijke vorm. Ook deze informatie kan centraal worden vastgelegd. En als stamgegevens worden gebruikt bij het vastleggen van metadata.

## 2.5 User Generated Content

Het is natuurlijk allang niet meer zo dat “professionals” het alleenrecht hebben op de omgang met cultureel erfgoed. Het publiek geeft reacties in de vorm van verbetervoorstellen, discussie of eigen verhalen. Dit wordt *User Generated Content* genoemd. En collectiebeheerder kan deze reacties apart opslaan of direct in de digitale informatie verwerken.

### Discussies

Het is handig om bij de presentatie van materiaal aan gebruikers de mogelijkheid aan te bieden om een reactie achter te laten. Het stelt het publiek in staat allerlei vormen van reacties achter te laten.

Discussies over een bron kunnen echter ook elders plaatsvinden. Te denken valt van Facebook, Twitter of Wikipedia.

### Verbetervoorstellen

Fouten en onvolledigheden in de verschillende soorten grondstoffen zijn onvermijdelijk. Soms hebben reacties de vorm van een verbetervoorstel van metadata (denk aan een verkeerde plaats van uitgave), van een digitale manifestatie (denk aan een typefout in een transcriptie) of van stamgegevens (denk aan een verkeerd geboortjaar van een fotograaf).

Ook hier geldt dat het publiek een voorstel tot verbetering op allerlei plaatsen kenbaar kan maken. Als dit gebeurt op de eigen website is het makkelijker hiervan kennis te nemen dan als de reactie plaatsvindt op een ander platform. Als de collectiebeheerder handige en duurzame webadressen (“url’s”) aan de objecten heeft toegekend, zijn vermeldingen op andere platformen gemakkelijker te traceren.

### Publieksverhalen

De publicatie van bronnen kan publiek inspireren om een bron te voorzien van een eigen verhaal. Of uitdagen om een eigen bron beschikbaar te stellen. Het levert nieuwe verhalen op, en vormt daarmee een nuttige aanvulling op de zoek- en vinddiensten die een erfgoedinstelling gewend is aan te bieden.

Ook hier geldt dat als de gebruiker naar het object verwijst met een webadres, de reactie gemakkelijker zijn te vinden. De collectiebeheerder kan dan deze informatie hergebruiken.

## 3. Metadata maken

Metadata is de grondstof die informatieprofessionals al tientallen, zo niet honderden jaren maken. In digitale vorm kunnen zoekmachines de verschillende soorten metadata, variërend van formele metadata tot preservation metadata, gebruiken om materiaal wereldwijd vindbaar te maken.

### 3.1 Catalogiseren/inventariseren: formeel beschrijven

Voor verschillende brontypes zijn verschillende afspraken over waaraan een formele beschrijving moet voldoen. Ook zijn er verschillende technieken om op een handige manier de formele beschrijvingen te verkrijgen.

#### Publicaties

Een collectiebeheerder van een publicatie beschrijft tegenwoordig meestal een titel niet meer zelf, maar neemt deze over uit een centrale database met een beschrijving die al door een andere instelling is gemaakt. Alleen de eigen specifieke kenmerken (zoals standplaats) voegt de collectiebeheerder dan zelf aan de beschrijving toe. OCLC biedt een dergelijke dienst aan. De aanbieder van een dergelijke dienst brengt vaak kosten in rekening. Soms biedt een collega collectiebeheerder zijn titelbeschrijvingen kosteloos als open data aan, wat het mogelijk maakt ook titels te ontlenen uit hun catalogus. Enkele grote nationale bibliotheken, zoals de KB, de DNB, de BNF en de BL, stellen hun titelbeschrijvingen op deze wijze beschikbaar.

Moderne bibliografische beschrijvingen voldoen aan de normen RDA (Joint Steering Committee for Development of RDA et al. 2010) en FRBR (IFLA 1998).

#### In de praktijk: LibraryThing.com

Op de website Librarything.com kan elke particulier (of kleine bibliotheek) een account aanmaken waarin hij zijn verzameling boeken kan catalogiseren. Dankzij de beschikbaarheid van de catalogusinformatie van verschillende grote bibliotheken (waaronder de KB) en onlineboekverkopers (waaronder Amazon), kan hij daarbij eenvoudig een titelbeschrijving aan de eigen (online) catalogus toevoegen. Dit werkt zo: met behulp van een titel, auteursnaam of ISBN-nummer kan hij een boek vinden. Door het boek te selecteren voeg je deze toe aan je eigen lijst van boeken. Alle beschikbare formele en inhoudelijke metadata wordt daar bij opgenomen.

Niet alle boeken zijn in het systeem te vinden. In dat geval kan je een eigen beschrijving opnemen. Deze titelbeschrijving kan iemand anders eventueel weer overnemen voor hetzelfde boek dat hij zijn eigen bibliotheek heeft.



## Archief

Archiefbeschrijvingen (oftewel *toegangen* of *inventarissen*) voldoen aan de norm ISAD(G) (ICA 2000). In deze norm is helder omschreven hoe een formele beschrijving van een archief er uit ziet. Een correcte beschrijving omvat informatie van het archief als geheel, dat wil zeggen van alle archiefstukken van één organisatie, persoon of familie. Vervolgens beschrijft de collectiebeheerder de afzonderlijke archiefbestanddelen waaruit een archief bestaat. Hierdoor ontstaat een hiërarchische ontsluiting. Informatie die de collectiebeheerder op een hoger niveau vastlegt, hoeft op een lager niveau niet te worden herhaald. Dit principe wordt *multi-level description* genoemd. Een archivaris kan niet alle pagina's in een archief apart beschrijven, maar zal op een bepaald niveau in de hiërarchie ophouden met ontsluiten.

## Beeld

Collectiebeheerders leggen over beeldmateriaal vaak weinig vast. Zij laten zich daarbij meestal inspireren door de metadatastandaard Dublin Core. Fotografen maken bij het fotograferen gebruik van het gestandaardiseerd metadatamodel EXIF. Een digitale camera neemt de metadata daarbij automatisch in het digitale foto-bestand op.

## Audiovisueel Materiaal

Voor het vastleggen van metadata van audiovisueel materiaal is door het Instituut van Beeld en Geluid een eigen metadatamodel ontworpen, op basis van FRBR, de metadatastandaard voor publicaties (de Jong 2007).

Voor televisieprogramma's, films en geluidsfragmenten zijn verschillende standaarden in omloop. Een overzicht is te vinden in de kennisbank van NIBG<sup>1</sup>.

## Museale objecten

Ook museale objecten hebben eigenschappen die het rechtvaardigen een specifiek metadatamodel te gebruiken. Er is behalve voor velden voor het beschrijven van de persoon, tijd en plaats van vervaardiging ook veel aandacht voor de technieken en materialen. Daarnaast is er veel aandacht voor de herkomst (oftewel *provenance*) van het object.

Het International Committee for Documentation (CIDOC) van The International Council of Museums (ICOM) schrijft acht velden voor, die vereist zijn om een museumobject te registreren en digitaal te ontsluiten<sup>2</sup>:

1. instellingsnaam;
2. inventarisnummer;
3. object trefwoord (objectnaam);
4. titel of korte beschrijving;
5. verwervingstrefwoord (methode);
6. verworven van;
7. verwervingsdatum;
8. standplaats

Een meer uitgebreide beschrijving bevat twintig velden.

---

<sup>1</sup> [http://avarchivering.nl/search/apachesolr\\_search?page=1&filters=tid%3A249](http://avarchivering.nl/search/apachesolr_search?page=1&filters=tid%3A249)

<sup>2</sup> <http://www.den.nl/standaard/189/>

## Normen

Let op dat:

1. elke soort bron wordt formeel beschreven volgens zijn eigen normen
2. bij verwijzingen naar personen, plaatsen en onderwerpsaanduiding wordt gekoppeld aan eigen of centrale stamgegevens.

## Juridische aspecten

Het ontlenen van titelbeschrijvingen uit een centrale database is alleen toegestaan als de collectiebeheerder overeenstemming heeft bereikt over het gebruik er van. Sommige partijen brengen daarvoor kosten in rekening. Andere partijen stellen hun metadata als Open Data beschikbaar, waardoor de collectiebeheerder er kosteloos gebruik van kan maken.

Het is – in beide gevallen – belangrijk na te gaan onder welke voorwaarden de collectiebeheerder de metadata weer aan derden mag doorleveren. Wellicht is het in de toekomst namelijk wenselijk de metadata op te nemen in een aggregator (zie hoofdstuk 9). Dit moet dan wel zijn toegestaan.

Omdat metadata in je database afkomstig is uit verschillende bronnen, is het ondoenlijk om bij alle informatie aan te geven welke juridische randvoorwaarden gelden. Dit probleem is alleen te voorkomen door metadata op te nemen die beschikbaar is onder de Creative Commons Licentie “CC0”<sup>3</sup>. Daarbij heeft de maker van metadata afstand gedaan van alle rechten en de metadata daarmee beschikbaar gesteld in het publieke domein.

## Effectiviteit

Collectiebeheerders slaan formele beschrijvingen voornamelijk op om een overzicht te kunnen hebben van de collectie en om bronnen terug te kunnen vinden in het depot. Het is een minimale activiteit om de informatie – zowel de analoge als de digitale manifestaties – adequaat te kunnen beheren. Zonder een formele beschrijving heeft (online) beschikbaar stellen weinig effect, omdat gebruikers niet in staat zijn om het materiaal met een zoekactie te vinden.

## 3.2 Catalogiseren: inhoudelijk beschrijven

Een effectieve ontsluiting zorgt ervoor dat publiek bronnen kan vinden op basis van persoonsnamen, onderwerpstermen, plaatsnamen en tijdsaanduidingen (“wie, wat, waar, wanneer”). De collectiebeheerder moet daarvoor in de metadata velden gebruiken die zijn gericht op de ontsluiting van resp. personen, plaatsen, onderwerpen en periodes.

De inhoudelijk beschrijving overstijgt het type bron: van elke bron, of het nu een archiefstuk, een boek of een schilderij is, kan de collectiebeheerder één of meer van de vragen “wie, wat, waar, wanneer” beantwoorden. Ook verwijst de collectiebeheerder daarbij zoveel mogelijk naar dezelfde stamgegevens, zodat het publiek gemakkelijk allerlei bronnen van en over één persoon of plaats kan combineren.

---

<sup>3</sup> <http://creativecommons.nl/>

## Aanpak

Het ontsluiten van bronnen op onderwerp is erg arbeidsintensief. Bovendien kan de keuze van een onderwerpsaanduiding erg afhankelijk zijn van de beoogde doelgroep of de informatieprofessional-van-dienst. Een handige aanpak bij het verkrijgen van een grote hoeveelheid inhoudelijk ontsluiting kan de collectiebeheerder daarom organiseren met behulp van het publiek. Bij de beschikbaarstelling op een website biedt hij daarvoor de mogelijkheid aan, om aan de bron een onderwerpsaanduiding of *tag* toe te kennen (Gerrits 2013).

De collectiebeheerder kan bij het vaststellen van een inhoudelijke beschrijving geautomatiseerde technieken inzetten. Deze worden behandeld in de volgende paragraaf.

## Normen

Let op dat:

1. de collectiebeheerder zoveel mogelijk van het viertal “wie wat waar wanneer” invult of laat invullen.
2. daarbij wordt ingevuld wat het *onderwerp* van de bron is. Er kan verwarring ontstaan over het verschil tussen de vervaardiger (bv. de fotograaf) en het onderwerp (de geportretteerde). Het is allebei “wie”, maar de eerste “wie” wordt in dit geval niet bedoeld.
3. bij het vastleggen van een onderwerp de collectiebeheerder zoveel mogelijk verwijst of laat verwijzen naar stamgegevens.

Eén en hetzelfde metadatamodel legt vaak zowel de formele als de inhoudelijke metadata vast. Sluit daarom bij de keuze van een standaard metadatamodel aan bij de keuze die is gemaakt voor formele metadata.

## Juridische aspecten

Voor de inhoudelijke metadata zijn de juridische aspecten hetzelfde als bij de formele metadata.

## Effectiviteit

Het brede publiek maakt weinig gebruik van de formele beschrijving om een bron te vinden. Een onderwerpsaanduiding spreekt veel meer tot de verbeelding. Inhoudelijke beschrijving zorgt daarom voor een groter publieksbereik. Als de collectiebeheerder de toekenning van onderwerpsaanduidingen daarnaast zoveel mogelijk overlaat aan het publiek zelf, is de investering beperkt en het effect groot.

Om het publiek te laten helpen bij het ontsluiten op onderwerp moet een collectiebeheerder investeren in een digitale omgeving waarin publiek onderwerpskenmerken aan een object kan toekennen. Begeleiding door personeel kan beperkt blijven als de collectiebeheerder ervoor kiest om het publiek zoveel mogelijk vrij te laten in haar keuze van de onderwerpsaanduidingen.

### 3.3 Inhoudelijk metadata geautomatiseerd bepalen

Indien van een object een machine-leesbare digitale manifestatie beschikbaar is, kan een inhoudelijke beschrijving geautomatiseerd worden aangevuld. Dit wordt automatische classificatie genoemd. Ook is het mogelijk de namen van personen of plaatsen geautomatiseerd te detecteren met behulp van *Named Entity Recognition* (NER). De gevonden namen geven dan een indicatie van het onderwerp van de tekst. De collectiebeheerder kan ze daarom als onderwerpskenmerken toevoegen aan de inhoudelijke metadata.

#### Werking

Bij NER maakt computersoftware gebruik van patronen om te bepalen of iets een naam van bijvoorbeeld een persoon, een plaats of een organisatie is, maar ook van specifiekere zaken als diersoorten of schepen. De NER-software verwerkt daarvoor transcripties of OCR-resultaten. Om de NER te laten werken maakt een informatieprofessional een trainingsset: een verzameling teksten met daarin door mensen gemarkeerde namen. Deze trainingsset gebruikt de NER-software om te “leren” hoe hij namen kan herkennen.

Ook geautomatiseerde classificatie begint met het maken van een trainingsset. De collectiebeheerder geeft daarbij in een representatieve steekproef van de bronnen aan welk onderwerp bij een document hoort. De software kent vervolgens de onderwerpen toe, aan de rest van de groep bronnen.

#### Aanpak

Bij verschillende universiteiten<sup>4</sup> zijn tools beschikbaar om NER mogelijk te maken. De inzet van deze tools is een technische oefening die een technisch georiënteerde medewerker kan uitvoeren. Als het academisch interessant is, zijn de medewerkers van de universiteit wellicht te interesseren in een project mee te doen.

Automatische classificatie wordt gerealiseerd door software van een aantal grote softwareleveranciers. Meer informatie is te vinden in ‘*Automatisch classificeren*’ (Aalten, van 2011)

#### Normen

De markering moet de software vastleggen op een gestandaardiseerde manier, die onafhankelijk is van het computerplatform of de software die de collectiebeheerder gebruikt voor de presentatie van de digitale informatie. ALTO – de standaard voor het opslaan van OCR-resultaten – bevat sinds de jongste versie een mogelijkheid om de Named Entity te markeren.

Onderwerpsaanduidingen maken deel uit van de inhoudelijke metadata. De onderwerpsaanduidingen die geautomatiseerd tot stand komen, moet de collectiebeheerder opnemen in de juiste velden van de standaard metadatamodellen die hij hiervoor gebruikt.

---

<sup>4</sup> <http://ilk.uvt.nl/frog/>  
<http://vanatteveldt.com/amcat/>

## Effectiviteit

Handmatig op zoek gaan naar namen in een tekst is heel arbeidsintensief en het is zeer waarschijnlijk dat de lijst met namen in dat geval niet compleet is. Een geautomatiseerde oplossing is daarom veel effectiever. De inzet van deze techniek is – vergelijkbaar met OCR – goedkoop omdat de computer het werk doet. Het verschil met OCR is dat er (nog) geen partijen zijn die het uitvoeren van NER als standaard-product aanbieden.

NER is de researchfase nog niet ontstegen: er is geen dienstverlener die het toepassen van NER als dienst aanbiedt. Er is daarnaast ook nog geen gebruikersinterface die het zoeken naar een naam mogelijk maakt. Het is namelijk nog niet zeker of het zoeken in een namenlijst waarvan de correctheid en compleetheid onduidelijk is, niet leidt tot meer frustratie dan verbetering van de zoekervaring.

### In de praktijk: Europeana

Europeana heeft in samenwerking met de Koninklijke Bibliotheek geëxperimenteerd met het herkennen van verschillende soorten namen in de gedigitaliseerde kranten. Het project gebruikte daarvoor software van Stanford University.

<http://www.europeana-newspapers.eu/named-entity-recognition-for-digitised-newspapers/>

In <http://dlib.org/dlib/july15/vanveen/07vanveen.html> beschrijven medewerkers van de KB hoe de informatie kan worden opgeslagen en gebruikt.

Automatische classificatie werkt over het algemeen het beste bij een grotere hoeveelheid tekst en een goede kwaliteit van de teksten (bv. niet te veel OCR-fouten). Het trainen van de software vergt bovendien tijd. De resultaten zijn meestal niet honderd procent correct, dus de vraag is wanneer het aantal fout toegekende onderwerpen te groot is om bruikbaar te zijn.

## 3.4 Rechten vastleggen

Het vastleggen van verschillende soorten rechten is van groot belang voor het succesvol beschikbaar stellen van informatie. Daarbij moet een collectiebeheerder denken aan:

- eigendom van de bron
- auteursrecht, naburige rechten
- portretrecht
- privacywetgeving
- wetgeving met betrekking tot beperking van openbaarheid

Belangrijke overzichten van informatie over rechten zijn:

- (Beunen and Schiphof 2006)
- <http://www.den.nl/pagina/501/de-basis-voor-auteursrechtenbeheer/>

Zodra de collectiebeheerder deze rechten op een correcte manier registreert, is het

mogelijk om bij de beschikbaarstelling geautomatiseerd te bepalen of hij het object zonder meer op internet beschikbaar mag stellen.

### **Eigenaarschap van de bron, beperkingen aan de openbaarheid**

Een erfgoedinstelling kan op verschillende manieren een object hebben verworven. De collectiebeheerder kan het object hebben aangekocht, geschonken hebben gekregen, in bruikleen hebben gekregen of hebben ontvangen in de zin van de archiefwet. Bij deze verwerving kan de vorige eigenaar contractueel beperkingen hebben gesteld aan het beschikbaar stellen van een object.

Dus, leg vast:

- een verwijzing naar het contract dat de eigendomsoverdracht van het object omschrijft.
- of beschikbaarstelling volgens dit contract is toegestaan, ja of nee.
- welke informatie – al dan niet volgens dit contract – de collectiebeheerder bij het object vermeldt, bijvoorbeeld: “in bruikleen van: ...”
- vervaldatum

Niet alleen voor beschikbaarstelling kan het van belang zijn om informatie over de verwerving vast te leggen. Dit kan bijvoorbeeld van belang zijn bij roofkunst. Het Bureau Herkomst kan meer informatie verschaffen over de informatie die de collectiebeheerder in dit verband moet kunnen vastleggen.

### **Auteursrecht**

Indien van toepassing, is het van belang de rechthebbende(n) en de copyright vermelding op te slaan. Daarnaast kan een apart contract zijn afgesloten waarin een rechthebbende aan de beherende instelling toestemming geeft voor gebruik op de website, of doorlevering aan derden. Het is van belang een verwijzing naar dit contract vast te leggen.

Dus, leg vast:

- indien van toepassing:
  - de houder van het auteursrecht
  - de copyright vermelding
  - een verwijzing naar een contract
- vervaldatum

#### **In de praktijk: Rijksmuseum**

Het Rijksmuseum beheert veel objecten waarop auteursrecht berust. Om dit materiaal beschikbaar te kunnen stellen is het noodzakelijk om na te gaan wie de auteursrechthebbende is. In het collectiebeheersysteem dat zij gebruiken hebben zij daarom extra velden gedefinieerd waarmee de rechthebbende, het copyrightstatement, en de vervaldatum worden opgeslagen. Daarnaast is een verwijzing opgenomen naar het contract dat het Rijksmuseum het recht geeft de afbeelding te publiceren.

Collectiebeheerders zijn dus in alle gevallen in staat om snel na te gaan wat de juridische status van een object. Materiaal dat niet beschikbaar mag komen op internet filtert de software er automatisch uit.

Helaas is het waarschijnlijk niet mogelijk om van alle objecten de rechthebbende te achterhalen – of anders met onredelijk veel moeite. Dit ontslaat een erfgoedinstelling echter niet van haar plicht zoveel mogelijk haar best te doen rechthebbenden te traceren en daarnaast te voorzien in een procedure waarin rechthebbenden zich kunnen melden om nadere afspraken te maken.

Voor dat deel van de collectie waarvan de rechthebbenden niet zijn te achterhalen is het gebruik van materiaal af te kopen bij een Collectieve Beheersorganisatie (CBO) voor auteurs- en naburige rechten. Hierin hebben rechthebbenden zich verenigd om hun belangen te behartigen. Belangrijke voorbeelden daarvan zijn:

- pictoright – voor beelden
- lira – voor teksten
- buma/stemra – voor muziek

De overheid heeft het College van Toezicht Auteursrechten (CvTA) ingesteld om toezicht te houden op een aantal CBO's<sup>5</sup>.

### **Naburige rechten**

De uitvoerders van een toneel- of muziekvoorstelling kunnen “naburige rechten” doen gelden. Naast de houder van het auteursrecht van tekst of muziek, is er dan iemand die de voorstelling heeft gegeven. Een opname daarvan kan dan niet zonder meer beschikbaar worden gesteld.

Een naburig recht is een vorm van auteursrecht, dus teksten in de bovenstaande paragraaf zijn ook hier van toepassing.

Dus, leg vast:

- indien van toepassing:
  - de houder van een naburig recht
  - de copyright vermelding
  - een verwijzing naar een contract
- vervaldatum

### **Portretrecht**

Naast auteursrecht kan er op een object portretrecht berusten: de geportretteerde (of zijn nakomelingen) kan verspreiding van zijn portret daarmee juridisch tegengaan tot tien jaar na zijn overlijden.

Dus, leg vast:

- indien van toepassing:
  - de houder van het portretrecht
  - een verwijzing naar een contract
- vervaldatum

---

5 <http://www.cvta.nl/over-het-cvta/op-welke-organisaties-houdt-het-college-toezicht>

## Privacywetgeving

Sommige oorlogsbronnen bevatten privacygevoelige informatie. Het is van belang dit te weten om beschikbaarstelling te voorkomen, of beschikbaarstelling alleen onder voorwaarden (bijvoorbeeld: alleen in de studiezaal) toe te staan.

Informatie is privacygevoelig als de informatie tot een persoon kan worden herleid. Dat wil zeggen als het voldoende informatie is om de betrokken persoon te achterhalen. Indien is bepaald dat een object privacygevoelige informatie bevat, mag de collectiebeheerder deze niet zonder meer online beschikbaar stellen.

Dus, leg vast:

- privacygevoelig, ja of nee
- vervaldatum

## Minimumeis

Elk van deze soorten recht kunnen de beschikbaarstelling van een object in de weg staan. Van elk object is het noodzakelijk om tenminste het rechtenstatement vast te leggen. Dit rechtenstatement vormt een samenvatting van de rechten die verschillende betrokkenen op een object kunnen doen gelden.

Voor de ontwikkeling van Europeana is daarvoor een lijst van standaard-rechtenstatements samengesteld<sup>6</sup>. In deze rechtenstatement staat wat er wel of niet met het object mag. In de metadata is dus tenminste een veld “rechtenstatement” noodzakelijk. De keuze voor het juiste rechtenstatement is afhankelijk van een aantal factoren. Met behulp van de website <http://outofcopyright.eu/> kan een collectiebeheerder nagaan welk rechtenstatement van toepassing kan zijn op een object.

Van een object moet tenminste een rechtenstatement zijn vastgelegd. Alle andere informatie vormt hiervoor de onderbouwing, en is daarom handig om vast te leggen en duurzaam te beheren.

## 3.5 Fragmenten markeren

In AV-materiaal is het handig om aan te geven op welk moment een fragment begint dat betrekking heeft op een bepaald onderwerp of een bepaalde spreker.

### Werking

Het vastleggen van het begin van een fragment kan heel eenvoudig: gekoppeld aan een tijdstip slaat een collectiebeheerder kenmerkende informatie op. Dit kan met de hand, maar dan moet iemand de gehele opname doornemen. Bij het herkennen van sprekers kan eventueel software worden ingezet om dit proces te automatiseren.

---

<sup>6</sup> <http://pro.europeana.eu/share-your-data/rights-statement-guidelines/available-rights-statements>



### In de praktijk: waisda.nl

“Waisda” is een omgeving waarin het publiek fragmenten uit een film kan voorzien van een *tag*. Met deze tag geeft het publiek aan over welk onderwerp het fragment gaat. Voorheen was deze omgeving beschikbaar op het webadres waisda.nl. Niemand onderhoudt dit platform en daarom is via deze website de computercode van het platform te downloaden zodat je het zelf kunt installeren.

### Aanpak

Het markeren van tijdstippen op basis van de sprekers kan geautomatiseerd gebeuren. Het wordt stem- of sprekersherkenning genoemd.

Een informatiesysteem kan het geluid en het beeld afspelen, terwijl iemand die dit beeld volgt, markeringen over onderwerpen of personen kan toevoegen. Een collectiebeheerder kan het publiek inzetten om deze manier van ontsluiten te realiseren.

### Normen

Informatie over fragmenten wordt – net als spraakherkenning – opgeslagen in een MPEG-7-bestand<sup>7</sup>.

### Effectiviteit

Indien een transcriptie beschikbaar is, kan het publiek een fragment vinden op basis van termen die zijn gebruikt in de tekst. Fragmenten markeren lijkt daarom vooral relevant als transcripties niet beschikbaar kunnen komen.

Het markeren van de fragmenten kan worden uitgevoerd door collectiebeheerders zelf. Indien het publiek de werkzaamheden uitvoert, door middel van *crowdsourcing*, zijn de kosten over het algemeen lager.

## 3.6 Preservation metadata vastleggen

Een digitale manifestatie is alleen bruikbaar als ook allerlei informatie *over* de digitale manifestatie beschikbaar is. Zo is het nodig om de relatie vast te leggen tussen verschillende bestanden bij één object, bijvoorbeeld de scans van een boek. Ook het digitale formaat dat gekozen is, is relevant: hiermee kan een computer bepalen welke software hij moet gebruiken om het object te kunnen laten zien. Al deze vormen van metadata scharen we onder “preservation metadata”. Ze helpen om ook in de toekomst een digitale manifestatie te kunnen begrijpen of tonen.

<sup>7</sup> <https://nl.wikipedia.org/wiki/MPEG-7>

## Aanpak

Het bedrijf of de organisatie die het maken van de digitale manifestatie voor zijn rekening neemt, is bij uitstek degene die deze metadata kan vastleggen. Het beste is daarom om goede afspraken te maken over het aanleveren van de preservation metadata bij de digitale manifestaties.

## Normen

De XML-standaard Metadata Encoding and Transmission Standard (METS) legt de volgorde en de verwijzing naar verschillende bestanden vast, behorende bij één object<sup>8</sup>. We rekenen deze metadata in rapport tot de preservation metadata, en ze wordt vaak ook structurele metadata genoemd.

In het METS-bestand staat vaak een beschrijving van de informatie die nodig is om het object ook in de toekomst te kunnen gebruiken. Deze metadatastandaard heet PREMIS<sup>9</sup>.

Indien sprake is van een afbeelding, is Metadata for Images in XML (MIX) de standaard voor het vastleggen van technische metadata zoals het software-formaat en de scan-techniek die is toegepast bij het maken van de afbeelding<sup>10</sup>. Deze informatie volgens het MIX kan ook opgenomen worden in het METS-bestand.

Ook de beschrijvende metadata kan een plaats krijgen in het METS-bestand. De collectiebeheerder kan daarbij kiezen voor de meest passende metadatastandaard (MODS, LIDO, EAD) afhankelijk van het soort object dat is beschreven.

## Effectiviteit

Het maken van deze vorm van metadata is noodzakelijk. Als de structuur niet is vastgelegd, is gebruik niet mogelijk.

## 3.7 Verbetervoorstellen van het publiek verwerken

Het maken van de verschillende soorten metadata die hier zijn omschreven is meestal de verantwoordelijkheid van de collectiebeheerder. Soms is hiervoor al omschreven hoe ook het publiek deze taak op zich kan nemen. Naast het *maken* van deze vorm van grondstof, is het mogelijk dat het publiek ook aanwijzingen geeft om metadata aan te passen of aan te vullen.

## Aanpak

Collectiebeheerders kunnen het aanpassen van de metadata op twee manieren organiseren. In sommige gevallen krijgt het publiek het recht om direct de metadata aan te passen. Maar meestal biedt een collectiebeheerder aan haar publiek de mogelijkheid om een reactie te sturen met de suggestie om informatie bij een bron te op te nemen of aan te passen. Een formulier op een website leidt dan vaak tot een e-mail aan de collectiebeheerder, die dan besluit de metadata al dan niet aan te passen.

<sup>8</sup> <http://www.loc.gov/standards/mets/>

<sup>9</sup> <http://www.loc.gov/standards/premis/>

<sup>10</sup> <http://www.loc.gov/standards/mix/>

## **Normen**

Het is belangrijk een goed proces te hebben ingericht waardoor de verbeteringen ook snel en adequaat worden doorgevoerd. Dit geeft een positieve indruk van de instelling in het bijzonder en erfgoedinstellingen in het algemeen. Een collectiebeheerder kan er belang aan hechten vast te leggen van wie hij de informatie heeft ontvangen.

## **Effectiviteit**

Het verwerken van de verbetervoorstellen is erg arbeidsintensief. Maar aanpassingen in de metadata dragen bij aan een voortdurende kwaliteitsverbetering.

## 4. Digitale manifestaties maken

Om zoeken, vinden en verrijken mogelijk te maken, selecteren collectiebeheerders de meest relevante bronnen en maken ze hiervan digitale reproducties, transcripties en/of gestructureerd data. In het geval van een waarheidsgetrouwe weergave van een bron, spreken we van digitale reproducties. De transcripties en gestructureerde data ontstaan door de inzet van OCR, data-entry of semigeautomatiseerde structurering.

Net als bij de creatie van metadata voeren collectiebeheerders deze werkzaamheden soms uit in samenwerking met het publiek.

### 4.1 Bronnen selecteren

De selectie van bronnen die als eerste in aanmerking komen voor digitaliseren, kent vier varianten:

1. het publiek bepaalt: materiaal dat in de studiezaal het meeste wordt aangevraagd komt in aanmerking om te scannen.
2. het publiek bepaalt: een gebruiker die behoefte heeft aan het scannen van een bepaald object, vraagt het scannen aan: *scanning on demand*
3. de instelling bepaalt: om kwetsbaar materiaal niet meer beschikbaar te hoeven stellen, kan de collectiebeheerder beslissen het materiaal te digitaliseren.
4. de instelling bepaalt: omdat een instelling het relevant vindt om een bepaalde bron beter onder de aandacht van het publiek te krijgen, kan de collectiebeheerder beslissen het materiaal te digitaliseren.

#### Hoeveelheid aanvragen

De collectiebeheerder registreert de aanvragen in de studiezaal meestal nauwkeurig. Dat stelt hem in staat om vast te stellen voor welke bronnen de meeste belangstelling bestaat. Deze bronnen kunnen om die reden in aanmerking komen om te worden gedigitaliseerd.

#### Scanning-on-demand: klein en groot

Bij scanning-on-demand plaatst een gebruiker een bestelling waarin hij aangeeft welke bron hij wil laten digitaliseren. De instelling digitaliseert deze bron en stelt deze vervolgens beschikbaar aan de besteller.

De instelling moet in staat zijn om binnen een redelijke termijn de gevraagde digitale reproductie te kunnen leveren. Daar moet de collectiebeheerder zijn organisatie op inrichten.

### **In de praktijk: Stadsarchief Amsterdam**

Het publiek kan bij het Stadsarchief Amsterdam opdracht geven om een inventarisnummer te laten scannen. Hierna zijn de scans van dit inventarisnummer beschikbaar om tegen kosten te bekijken en downloaden. Hiervoor heeft het SAA een afspraak met een commercieel digitaliseerbedrijf om 15.000 scans per week te maken. Niet meer, indien er meer bestellingen zijn, worden deze naar de volgende week uitgesteld.

meer info:

- [https://stadsarchief.amsterdam.nl/archieven/archiefbank/hoe\\_werkt\\_de\\_archiefbank/](https://stadsarchief.amsterdam.nl/archieven/archiefbank/hoe_werkt_de_archiefbank/)
- <http://www.slideshare.net/mholtman/scanning-on-demand-stadsarchief-amsterdam>

Scanning-on-demand kan plaats vinden op kleine en op grote schaal. Er kan behoefte zijn aan het scannen van een afzonderlijk archiefstuk voor een individueel doel. Naast deze individuele vraag, kan een bron voor digitalisering in aanmerking komen om daarmee bij te dragen aan een specifiek doel. Onderzoeksinstituten kunnen in overleg treden om gezamenlijk de digitalisering van een bepaalde bron te financieren. Een historicus kan deze bron vervolgens gebruiken voor zijn onderzoek, al dan niet met toepassing van technieken ontwikkeld in het kader van Digital Humanities.

### **Kwetsbaar materiaal eerst**

Een belangrijke reden om materiaal te willen laten scannen, is in het geval dat materiaal in een slechte staat verkeert en de collectiebeheerder het materiaal bij voorkeur niet meer in de studiezaal beschikbaar stelt. Na het scannen kan het publiek het materiaal dan meestal niet meer aanvragen, zodat het niet langer aan slijtage onderhevig is.

### **In de praktijk: Metamorfoze**

Het programma Metamorfoze subsidieert het digitaliseren van Archieven en Bijzondere Collecties (ABC) of Boeken, Kranten en Tijdschriften (BKT) met als doel om papieren erfgoed te behouden voor de toekomst. Publiek raadpleegt fysiek materiaal dat is gedigitaliseerd immers minder vaak en intensief. Subsidie wordt daarom toegekend bij de digitalisering van kwetsbaar materiaal. De website van Metamorfoze bevat veel documentatie over digitalisering. Indien je subsidie van Metamorfoze krijgt, bieden zij behalve financiële ondersteuning ook kennis en advies. Daarnaast kan een collectiebeheerder bij het verstrekken van een opdracht aan deze documenten refereren.

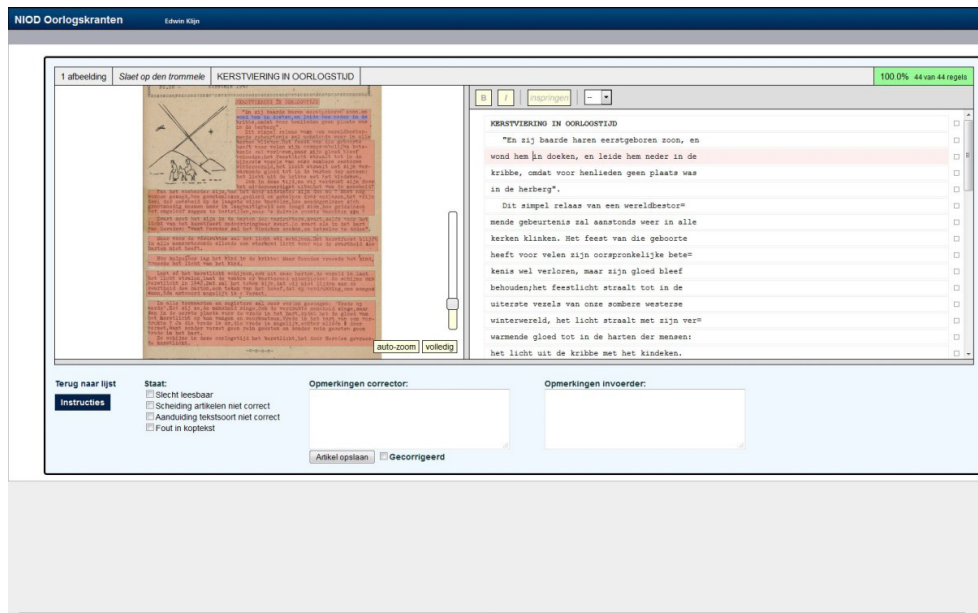
### **Op eigen initiatief**

Een collectiebeheerder kan er voor kiezen het materiaal te scannen dat het meest interessant is. Of het publiek dit materiaal ook interessant vindt, is dan de vraag. Maar het kan zijn dat de collectiebeheerder pareltjes in de collectie onder de aandacht wil brengen, waarvan het publiek het bestaan niet kent en de relevantie nog niet heeft ontdekt.

## 4.2 Digitale reproductie maken

Bronnen die alleen analoog beschikbaar zijn, zijn minder gemakkelijk te bekijken of te beluisteren dan digitaal materiaal. Het is daarom een belangrijke stap om van het materiaal een digitale reproductie te (laten) maken. Bij beeldmateriaal, archief en publicaties spreekt men daarbij van scans. Bij bewegend beeld en opnames spreekt men meer in het algemeen van digitaliseren. Voor het maken van een digitale weergave van een museaal object schakelt een collectiebeheerder een fotograaf in.

De digitale manifestatie die ontstaat kan op verschillende manieren worden gebruikt. Bij de creatie van de formele of inhoudelijke beschrijving of een transcriptie, is het bijvoorbeeld handig om een digitale weergave van het object bij de hand te hebben. Het registratiesysteem waarin de collectiebeheerder de metadata opslaat, geeft de digitale reproductie dan weer naast het invulformulier van de metadata. Bij het beschrijven van beelden zie je dit vaak gebeuren: de gescande foto's zitten dan zonder beschrijving in het systeem, waarna een collectiebeheerder of een vrijwilliger metadata toevoegt.



Afb. mock-up van een applicatie waarin een transcriptie kan worden gemaakt of verbeterd op basis van een scan.

De digitale reproductie heeft echter de meeste waarde bij het beschikbaar stellen op internet. Hierdoor kan iedereen de bron op elke plaats en op elk moment raadplegen.

### Werking

Bij het maken van een digitale reproductie meet de leesapparatuur – in het geval van bijvoorbeeld een scan met behulp van een scanner – een kleine fragment (een “sample”) van de analoge bron gemeten. Van dit gemeten fragment slaat het apparaat een digitale representatie op in een computerbestand. De bits-en-bytes vormen daarmee een neerslag van het gemeten fragment. Bij een scan is dat bijvoorbeeld de kleur van een punt (oftewel pixel) en bij een geluidsfragment de toonhoogte en het volume. Het computerbestand bevat de gecodeerde vorm van alle gemeten fragmenten van

de bron. Software kan deze codering interpreteren en omzetten in een beeld op een beeldscherm of geluid op een geluidsinstallatie.

Hoe kleinere het fragment hoe preciezer de weergave van de originele bron kan worden gemaakt.

### **Aanpak**

Het digitaliseren van een bron gebeurt meestal door een commerciële partner. Er zijn verschillende commerciële aanbieders van scandiensten. Deze presenteren zichzelf prominent op internet; een goede zoekactie levert genoeg adressen op.

Materiaal dat een collectiebeheerder heeft geselecteerd om te laten digitaliseren, moet hij goed voorbereiden:

- is het materiaal compleet?
- ligt het materiaal in de juiste volgorde?
- is het materiaal herkenbaar te identificeren, bijvoorbeeld door middel van een streepjescode?
- hoe wordt bij terugkeer gecontroleerd of het digitalisering goed is verlopen?

De belangrijkste specificatie bij digitaliseren betreft de resolutie, dat wil zeggen het aantal fragmenten dat de scanapparatuur vastlegt. De resolutie is afhankelijk van het kleinste element op de afbeelding: deze moet op de digitale reproductie ook herkenbaar blijven. Daarnaast is het opslagformaat en de compressie van belang.

De bestandsnamen kunnen betekenis hebben of niet. Een betekenisvolle naam bevat bijvoorbeeld de afkorting van de instelling, het nummer van de publicatie in de catalogus en het paginanummer. Een betekenisvolle naam stelt je in staat om geautomatiseerd een koppeling te leggen met de beschrijvingen in het collectiebeheersysteem. Het nadeel van een betekenisvolle bestandsnaam is dat in de toekomst de naam van de instelling of de nummering van de collectie-onderdelen verandert, waardoor de bestandsnamen een verouderde naam of nummering bevatten. Het is daarom aan te bevelen de context van de scan vast te leggen in aparte metadata.

De leverancier kan de digitale reproducties aanleveren op een opslagmedium, tegenwoordig vaak harde schijven of NAS-apparaten. Ook kan de leverancier het materiaal rechtstreeks via internet in een opslagsysteem kopiëren.

Het is belangrijk om een controle uit te voeren op het resultaat van de reproductieopdracht. Dat kan door bij binnenkomst een steekproef te nemen en te zien of aan de gestelde normen is voldaan. Bij hele grote opdrachten is alleen een steekproef niet voldoende om overtuigd te raken van de geleverde kwaliteit. Hoewel grote fouten, bijvoorbeeld veel zwarte scans achter elkaar, kan worden vastgesteld, is een steekproef statistisch waarschijnlijk altijd te klein. Trucs om te controleren zijn om – na compressie – het bestand van de grootste en kleinste omvang, en de eerste en de laatste van een serie te bekijken. De kans dat daarmee iets niet in orde is, is het grootste.

Een collectiebeheerder kan daarnaast allerlei geautomatiseerde controles inbouwen. Zo'n geautomatiseerde controle kan bijvoorbeeld de aantallen in de metadata en het aantal bestanden vergelijken. Dat vergroot de kans dat een collectiebeheerder op het spoor komt van de meest voorkomende fout: het overslaan van pagina's. Procedures bepalen of aan belangrijke technische voorwaarden (de kwaliteit van een XML-bestand en de correcte overdracht van bits en bytes) is voldaan. Het gehele controleproces moet voorkomen dat fouten optreden bij het opslaan van de bestanden in de omgeving waarmee de collectiebeheerder het materiaal straks beheert en publiceert.

Ten slotte is het goed om na te gaan of bij de leverancier het digitaliseringsproces goed is georganiseerd. Dit kan door zelf te gaan kijken of een audit op het digitaliseringsproces te laten uitvoeren.

## **Normen**

De meest actuele normen en best practices voor het laten uitvoeren van deze werkzaamheden zijn gepubliceerd door Metamorfoze (voor scans<sup>11</sup>) en DEN (voor scans en andere vormen van digitale reproducties<sup>12</sup>).

Maak bij een opdracht goede afspraken over:

1. wat de specificaties zijn
2. wat de bestandsnamen zijn
3. de manier waarop de leverancier de digitale reproducties aanlevert

## **Effectiviteit**

Het digitaliseren van een bron is bijzonder effectief. Een collectiebeheerder kan materiaal eenvoudiger beschikbaar stellen dan voorheen. Na het scannen zijn allerlei technieken inzetbaar om de toegankelijkheid te vergroten.

De prijs van de digitaliseringsopdracht is erg afhankelijk van de hoeveelheid materiaal dat de collectiebeheerder in één keer aan het scanbedrijf aanbiedt, evenals de kwetsbaarheid van het te scannen materiaal. Daarnaast kunnen er in een aanbesteding allerlei markttechnische redenen zijn om een prijs te bepalen. Een grove indicatie is dat het scannen ongeveer 1 euro per pagina kost. De prijs neemt af indien het volume groter en het materiaal eenduidiger is. De voorbereidingswerkzaamheden die een collectiebeheerder zelf moet uitvoeren, zijn behoorlijk tijdsintensief.

Enkele erfgoedinstellingen realiseren het scanproces zelf. Dit is pas lucratief indien er sprake is van hele grote hoeveelheden.

## **Juridische aspecten**

Het maken van een foto van een 3D-object zorgt ervoor dat de fotograaf naburige rechten krijgt van de foto. Hij voegt met de keuze van de schaduwwerking en hoek iets extra's toe. Het is dus belangrijk om deze naburige rechten met de fotograaf goed te regelen. Bij het maken van een scan is dat niet nodig, want daarbij is geen sprake van naburige rechten.

---

11 <https://www.metamorfoze.nl/>

12 <http://www.den.nl/pagina/218/>



## 4.3 OCR-en

Het maken van een volledige transcriptie van een publicatie, archiefmateriaal, be-  
wegend beeld of geluidsopnamen is heel effectief: alle gebruikte woorden kunnen  
de tekst vindbaar maken met behulp van een zoekmachine. Het is een manier van  
zoeken die we sinds de introductie van zoekmachines zoals Yahoo, Bing en Google  
gewend zijn.

### Werking

Met behulp van Optical Character Recognition (OCR) is het mogelijk om geautoma-  
tiseerd een tekst te maken van gedrukte tekst op een scan. Hierbij maakt software  
een redenering waarin de computer bepaalt welke letters op de scan staan. De soft-  
ware zoekt daarbij naar patronen in de bits-en-bytes die in een scan zijn opgeslagen.

Deze techniek is behoorlijk ver ontwikkeld voor drukwerk, maar staat voor handge-  
schreven bronnen nog in de kinderschoenen.

### Aanpak

Om OCR te kunnen uitvoeren, zijn scans nodig van de publicatie. De scans moeten  
van voldoende resolutie zijn. Meestal biedt het bedrijf dat de scan verzorgt als extra  
dienst ook het uitvoeren van de OCR-bewerking.

Belangrijk is om bij een herkend woord vast te leggen waar op de scan het woord  
zich bevindt. Bij de presentatie van de scan kan de presentatiesoftware het woord  
dan voorzien van een markering. Hierdoor kan de gebruiker snel nagaan waar de  
door hem of haar gezochte term zich bevindt in de (gescande) tekst.

#### In de praktijk: Delpher

Delpher – een samenwerkingsproject van de Koninklijke Bibliotheek en Univer-  
siteitsbibliotheken – biedt de mogelijkheid om te zoeken door een grote verza-  
meling gedigitaliseerde publicaties: kranten, tijdschriften en boeken. Dat  
zoeken is mogelijk dankzij OCR. Het resultaat daarvan is opgeslagen in het  
ALTO-formaat en doorzoekbaar met behulp van een zoekmachine.

<http://www.delpher.nl/nl/platform/pages/?title=kwaliteit+%28ocr%29>

### Normen

Het OCR-resultaat – zowel de woorden als de plaats op de scan – moet beschikbaar  
komen in een formaat dat onafhankelijk van het computerplatform kan worden ge-  
lezen. De actuele normen die daarvoor gelden zijn gepubliceerd door het IMPACT-  
project<sup>13</sup>. De standaarden om het scanresultaat in op te slaan is ALTO<sup>14</sup>. ALTO legt  
naast de gevonden letters ook de plaats vast van de letters op de scan. Hierdoor kan  
software straks een gevonden tekst op een scan markeren bij de presentatie.

<sup>13</sup> [www.impact-project.eu/](http://www.impact-project.eu/)  
<http://www.digitisation.eu/>

<sup>14</sup> <http://www.loc.gov/standards/alto/>

OCR-software kan modern drukwerk bijna foutloos herkennen. Echter: hoe slechter het drukwerk, hoe slechter het OCR-resultaat. Ook oude spelling maakt het ingewikkelder om het percentage van goed herkende letters te verhogen.

Bij controle van het OCR-resultaat dat door een leverancier wordt geleverd, kan worden nagegaan of er geen onnodige fouten zijn gemaakt waardoor het resultaat onnodig laag is. Maar een bepaald percentage eisen is weinig effectief: het is van veel factoren afhankelijk die buiten het bereik liggen van de leverancier.

### **Effectiviteit**

Het laten uitvoeren van OCR gebeurt volledig geautomatiseerd. Dat maakt dat leveranciers het bijzonder goedkoop kunnen aanbieden. De prijs is maximaal enkele centen. En de beschikbaarheid van de volledige tekst van de publicatie maakt het mogelijk dat er full-text search kan worden ingezet om een tekst vindbaar te maken. Dus: weinig moeite en kosten, voor een groot resultaat.

## **4.4 Spraakherkenning**

Ook voor geluidsopnamen van spraak kan gespecialiseerde software een transcriptie van de gesproken tekst genereren. Deze transcriptie kan worden gebruikt om een opname full-text doorzoekbaar aan te bieden.

### **Werking**

De techniek van spraakherkenning is tot op zekere hoogte vergelijkbaar met OCR: een computer bepaalt de tekst op basis van de patronen die aanwezig zijn in de bits-en-bytes waaruit de digitale reproductie bestaat.

De software slaat bij het herkende woord op, op welk moment het woord werd uitgesproken. Bij de presentatie van de bron kan dan direct het fragment worden afgespeeld waar het woord zich bevindt.

### **Aanpak**

Spraakherkenning kan alleen plaatsvinden op basis van een digitale reproductie en als deze digitale reproductie voldoende precies een weergave is van de originele bron. Wellicht levert het bedrijf dat de digitalisering organiseert ook een dienst waarbij de spraakherkenning wordt uitgevoerd.

#### **In de praktijk: wiewaswie.nl**

Archiefmateriaal waarin geboortes, huwelijken en overlijdens zijn geregistreerd wordt Burgerlijke Stand genoemd. De formele beschrijving van dit archief bestaat uit niets meer dan bv. "Geboorteregister Haarlem 1813-1823". Publiek – in het bijzonder genealogen – zijn echter geïnteresseerd in de namen van de borelingen en hun ouders.

Daarom hebben archiefdiensten zich de afgelopen twintig jaar ingespannen om deze namen uit de registers over te nemen. Publiek kan nu op zoek naar persoonsnamen, opgesplitst in voornaam, tussenvoegsel en achternaam, op de website wiewaswie.nl. De hoeveelheid informatie neemt hier nog steeds toe.

## Normen

Om spraakherkenning te kunnen toepassen is een minimum-kwaliteit noodzakelijk van de digitale reproductie. Ook heeft ruis – op dezelfde manier als oud drukwerk bij OCR – een negatieve invloed op de kwaliteit van het resultaat.

De standaard voor het vastleggen van de herkende spraak, met verwijzingen naar het fragment in de stream is mogelijk met behulp van de standaard MPEG-7<sup>15</sup>. MPEG-7 is een XML-standaard waarin de woorden en tijden aan elkaar worden verbonden.

## Effectiviteit

Het laten uitvoeren van spraakherkenning gebeurt volledig geautomatiseerd. De software is echter minder gemakkelijk in het gebruik en wordt daarom niet door alle leveranciers zonder meer aangeboden. Hiervoor bestaan gespecialiseerde bureaus. De prijs of moeite die een collectiebeheerder moet doen om spraakherkenning toe te passen is daarom groter dan bij OCR.

### In de praktijk: Radio Oranje

<http://hmi.ewi.utwente.nl/showcase/Radio%20Oranje%20demo>

Met behulp van deze demonstrator kun je zoeken in de toespraken van H.M. Koningin Wilhelmina die tijdens de Tweede Wereldoorlog door Radio Oranje zijn uitgezonden. Een computer heeft de spraak in het geluidsfragment gekoppeld aan de tekst. Na het vinden van een woord in de tekst, speelt de website het fragment uit waarin dit woord wordt uitgesproken.

De demo is gemaakt in het kader van een wetenschappelijk onderzoek en sindsdien nergens als productie-omgeving ingericht.

## 4.5 Data entry

Gestructureerde data kan worden verkregen door de relevante informatie in de juiste velden over te typen. Maar ook transcripties van geschreven en gesproken tekst (in resp. boeken/archieven en audio-visueel materiaal) kunnen door middel van overtypen worden gerealiseerd.

### Aanpak

Het overtypen van informatie uit een bron kan een collectiebeheerder eenvoudiger (laten) uitvoeren indien een digitale reproductie van de bron aanwezig is. Het invoeren kan de collectiebeheerder dan namelijk via internet organiseren en dus altijd en overal worden gedaan.

Zo is het mogelijk de scans te versturen naar een lagelonenland. Medewerkers van gespecialiseerde data-entrybureaus in bijvoorbeeld India kunnen met behulp van heldere instructies ingezet worden om de tekst of de data in te voeren.

<sup>15</sup> <https://nl.wikipedia.org/wiki/MPEG-7>

Bij het lezen van oude handschriften is de inzet van data-entrybureaus vaak niet mogelijk omdat deze meestal de expertise niet hebben om het oude schrift te lezen. Wel kunnen dan vrijwilligers worden ingeschakeld om de data over te typen. Dit wordt *crowdsourcing* genoemd. Het werven en stimuleren van vrijwilligers is daarbij een belangrijke activiteit. Gun de vrijwilligers de eer van hun werk en beloon ze met rechten voor het downloaden van scans en bijeenkomsten waar je ze voor hun inspanningen bedankt.

### **In de praktijk: Velehanden.nl**

Op de website [velehanden.nl](http://velehanden.nl) kunnen archiefdiensten een project uitvoeren waarin gebruikers van de website informatie over archiefstukken invoeren. Het gaat daarbij heel vaak om registers die het mogelijk maken om materiaal gemakkelijk op tijd of persoonsnaam terug te vinden. Voorbeelden zijn: 19<sup>e</sup> eeuwse registraties voor de dienstplicht, notarieel archief of bevolkingsregisters.

Archiefdiensten laten daarvoor eerst materiaal scannen en in het systeem inlezen. De gebruiker krijgt de scan te zien in combinatie met een invulformulier waarin hij informatie van de scan overneemt. Na de invoer controleert een andere gebruiker de ingevoerde informatie op kwaliteit.

### **Normen**

Vergelijkbaar met OCR kunnen collectiebeheerders met een leverancier afspraken maken over het aantal fouten dat per eenheid (bv. per pagina) aanwezig mag zijn. Om fouten te voorkomen, voert een invoerder data vaak twee keer in. Als twee personen onafhankelijk van elkaar dezelfde data invoeren, is de kans dat ze allebei dezelfde fout hebben gemaakt te verwaarlozen.

### **Effectiviteit**

De belangrijkste hoeveelheid werk voor het realiseren van data-entry is de voorbereiding van de bronnen, het opstellen van een heldere instructie en het werven en begeleiden van een data-entry bureau of vrijwilligers. Het resultaat is nagenoeg foutloze data.

Interessant neveneffect bij *crowdsourcing*, is dat publiek betrokken raakt bij het erfgoed waaraan zij zelf een bijdrage levert. Er vormt zich een *community* rond de erfgoedinstelling, waarop de instelling ook in andere gevallen een beroep kan doen.

## **4.6 (Semi-)automatisch gestructureerde data maken**

Collectiebeheerders kunnen met de hand structuur aanbrengen in een transcriptie om tot gestructureerde data te komen. Handiger is echter om te onderzoeken of het mogelijk is om de structuur van de data geautomatiseerd aan te brengen.

### **Aanpak**

Een oud adresboek is bijvoorbeeld vaak erg helder gestructureerd. Een naamsvermelding bestaat uit achtereenvolgens de achternaam, de voorletter, het tussenvoegsel,

het beroep en het adres. Vaak is er tussen de onderdelen een apart teken gekozen, bijvoorbeeld een punt-komma (“;”). Deze structuur kan een collectiebeheerder met zoek-en-vervang (laten) omzetten in tabelvorm, waarbij bijvoorbeeld een tab de velden onderscheidt.

Het maken van een zoek-en-vervang strategie – soms ook op basis van andere regelmatigigheden in de naamsvermelding, zoals hoofdlettergebruik – is voor elke bron weer anders. Eigenlijk wordt een specifiek computerprogramma geschreven, specifiek gericht op de bron.

### **In de praktijk: Verzetskranten uit Winkel op wikipedia**

Het is wenselijk dat er een online lijst is van de verzetskranten, zodat de informatie hierover kan worden gerelateerd aan de kranten in bijvoorbeeld Delpher. Lydia Winkel schreef daarover in de jaren '50 een standaardwerk, waarvan het NIOD de auteursrechten had. Het NIOD heeft haar recht gebruikt om de data in het boek beschikbaar te stellen onder de CC-BY licentie.

De informatie vormt de basis van het overzicht van verzetskranten op wikipedia. Een gericht project moet leiden tot een overzicht van al deze kranten, door deskundigen aangevuld en verbeterd. Omdat de data op de wikipedia-pagina's ook als aparte data kan worden benaderd, is het mogelijk de lijst in de toekomst op allerlei plaatsen te gebruiken.

<https://nl.wikipedia.org/wiki/Wikipedia:Wikiproject/Verzetskranten>

### **Effectiviteit**

Er moet een goede afweging gemaakt worden tussen de hoeveelheid tijd en expertise die nodig is om een gecompliceerd computerprogramma te schrijven en de tijd die nodig is om het structureren met de hand uit te voeren.

Indien gestructureerde data beschikbaar is, kan het publiek gericht in de bron zoeken naar de informatie van zijn of haar interesse. Bovendien is het mogelijk dat onderzoekers de data gebruiken voor hun onderzoek.

### **In de praktijk: LINKS**

Het wetenschappelijk programma LINKS gebruikt de data die is ingevoerd voor het toegankelijk maken van Burgerlijke Stand materiaal voor genealogen.

Het programma ontwikkelt software die verbanden tussen de gegevens legt en daarmee families en *life courses* reconstrueert. Historici kunnen op basis van deze informatie nieuwe inzichten krijgen over allerlei demografische ontwikkelingen in de negentiende eeuw.

<http://www.iisg.nl/hsn/projects/links.html>

## 4.7 Verbetervoorstellen van het publiek verwerken

Niet alleen metadata is voor verbetering vatbaar. Ook in digitale manifestaties zijn fouten ontstaan. Zo kunnen in een digitale reproductie fragmenten ontbreken, kan een woord verkeerd zijn herkend of overgetypt in een transcriptie of een structuurfout zijn ontstaan in de gestructureerde data.

De manier waarop collectiebeheerders aanwijzingen van het publiek voor verbeteringen doorvoeren, is hetzelfde als bij metadata. Zie daarvoor hoofdstuk 6 en paragraaf 10.4.

## 5. Stamgegevens maken en ernaar verwijzen

Stamgegevens hebben we gedefinieerd als de gegevens die de collectiebeheerder in de metadata niet steeds herhaalt, maar aan een object koppelt door middel van een verwijzing. Voorbeelden zijn *authority-files*, thesauri of gazetteers (voor plaatsnamen).

Belangrijke vraag met betrekking tot stamgegevens is: wie organiseert het beheer van de informatie? Het betreft immers lijsten waarvan het beheer op de ene plek en het gebruik op een andere plek gebeurt. Het ligt voor de om het beheer van stamgegevens daar te beleggen waar de specifieke kennis over het onderwerp beschikbaar is. Bij een archiefdienst in Utrecht zijn ze gespecialiseerd in de historische straatnamen van Utrecht, medewerkers van het NIOD zijn gespecialiseerd in trefwoorden over de Tweede Wereldoorlog en de Koninklijke Bibliotheek heeft de meeste kennis over landelijk opererende auteurs.

Het gebruik van dezelfde stamgegevens door verschillende collectiebeheerders kan niet worden afgedwongen en vaak kiezen zij terecht voor stamgegevens in hun eigen domein. Verschillende collectiebeheerders maken in de praktijk dus gebruik van verschillende stamgegevens. Er is daarom behoefte aan koppelingen: de persoon in de ene lijst, is dezelfde als een persoon in een andere lijst. De Rijksdienst voor Cultureel Erfgoed heeft voor Nederlandse plaatsnamen en locaties daarom – in samenwerking met DEN en De Waag – de website [erfgeo.nl](http://erfgeo.nl) ontwikkeld. Het koppelen van verschillende stamgegevens over hetzelfde onderwerp, persoon of plaats wordt *alignment* genoemd.

### 5.1 Authority-files gebruiken

Metadata bevat namen van onder meer personen en plaatsen. Een collectiebeheerder kiest bij het invullen van het veld auteur bij voorkeur een naam uit een lijst van reeds ingevoerde auteursnamen. Deze lijst wordt een *authority-file* genoemd. Aanvankelijk maakte elke instelling zijn eigen authority-file. Het is nog handiger als veel instellingen gebruik maken van dezelfde lijst.

#### Aanpak

Het is wenselijk dat zoveel mogelijk instellingen gebruik maken van dezelfde lijst van persoonsnamen om in te vullen bij metadatavelden voor personen van onderwerp, vervaardigers en auteurs. Ga daarom eerst na of er een lijst is die aansluit bij je behoefte.

Er zijn namelijk een aantal belangrijke bronnen voor stamgegevens voor personen. Uit deze bronnen kan een identificerende code, bijvoorbeeld een nummer of (beter nog) een gehele URL worden overgenomen, die verwijst naar de informatie over een persoon.

De belangrijkste voorbeelden zijn:

- VIAF: Virtual International Authority File (<http://www.viaf.org/>), ontwikkeld en gehost door OCLC.
- ULAN: Union List of Artist Names (<http://www.getty.edu/research/tools/vocabularies/ulan/>), ontwikkeld en gehost door Getty
- Biografisch portaal (<http://www.biografischportaal.nl>), ontwikkeld en gehost door ING-Huygens.
- RKD-artists&: <https://rkd.nl/nl/info/rkdartists>
- SNAC: <http://socialarchive.iath.virginia.edu/>

## Normen

Belangrijk is dat een collectiebeheerder op de goede manier verwijst naar de gekozen persoon. Dit kan door de gestandaardiseerde term, de identificerende code, in te vullen in de database. Het is efficiënt als de database een mogelijkheid heeft om een term te kunnen kiezen uit een selectielijst. Het is daarom noodzakelijk dat de lijst is te benaderen met behulp van een techniek die de data rechtstreeks weergeeft in het collectiebeheersysteem. Een Application Programming Interface maakt dit mogelijk.

## Effectiviteit

Alle bronnen die verwijzen naar één en dezelfde persoonsvermelding kan een zoekmachine eenvoudig combineren. Het stelt de zoekmachine in staat om een lijst aan te bieden van alle bronnen die betrekking hebben op of gemaakt zijn door één en dezelfde persoon of instelling.

Of het publiek materiaal gemakkelijker vindt, wordt bepaald door

- de kwaliteit van de stamgegevens,
- hoeveel ze zijn toegekend in de metadata,
- of veel instellingen de stamgegevens gebruiken.

Het toekennen van termen kost veel tijd. Een collectiebeheerder doet er daarom goed aan de mogelijkheid te onderzoeken of hierbij crowdsourcing kan worden toegepast, of geautomatiseerde classificatie.

## 5.2 Authority-files maken en publiceren

Aanhaken bij een bestaande authority-file is veruit het handigste. Een andere instelling neemt dan de verantwoordelijkheid om de informatie te beheren.

Maar wellicht is er geen passende authority-file beschikbaar. In dat geval kan je zelf een lijst van gestandaardiseerde namen opbouwen en – belangrijker nog – publiceren. Andere kunnen dan van dezelfde lijst gebruik maken, zodat de verwijzingen naar de personen leiden tot een complete lijst van bronnen.

## Aanpak

Indien het niet mogelijk is gebruik te maken van een authority-file van een andere organisatie, is het verstandig je eigen authority-file te publiceren en te stimuleren dat ook anderen er gebruik van maken.



Zo is het mogelijk een lijst van verzetsgroepen samen te stellen die andere collectiebeheerders kunnen gebruiken bij het ontsluiten van materiaal. Door gebruik te maken van een collaboratief platform, zoals wikipedia, is het mogelijk iedereen aan dergelijke stamgegevens te laten bijdragen.

### **In de praktijk: Joods Historisch en Westerbork werken samen**

Zowel het Joods Historisch Museum als het Herinneringscentrum kamp Westerbork beheren een informatiesysteem met informatie over Nederlandse slachtoffers van de Holocaust. Ze hebben onafhankelijk van elkaar informatie verbeterd en hebben verschillende oorlogsbronnen aan de namen gerelateerd. In de basisinformatie (zoals naam, geboortedatum en -plaats, overlijdensdatum en -plaats) waren verschillende keuzes gemaakt. Beide partijen zijn nu overeengekomen dat ze een gezamenlijke database inrichten waarin zij deze basisinformatie eenduidig vastleggen. Vanuit deze centrale database wordt met behulp van een Application Programming Interface (API) de informatie in de database van zowel het JHM als Westerbork voorzien van aanvullende informatie en bronnen.

De informatie over personen die een collectiebeheerder in zijn eigen lijst opneemt, is waarschijnlijk uniek. Dat was immers de reden om gebruik te maken van een eigen lijst. In bijvoorbeeld een overzicht van auteurs legt de Koninklijke Bibliotheek informatie vast over het literaire leven van de schrijver. Indien het NIOD een archief beheert van dezelfde schrijver, legt het NIOD informatie bij de archiefvormer vast die betrekking heeft op de oorlog. Het betreft dezelfde persoon, de informatie overlapt, maar is niet hetzelfde. Er moet daarom een verwijzing zijn van de ene naar de andere beschrijving.

### **Norm**

Er zijn verschillende normen die een collectiebeheerder kan hanteren bij het publiceren van authority-files.

Uit het bibliotheekwezen stammen MARC-21 en Metadata Authority Description Schema (MADS). MARC-21 is gemaakt als formaat voor alle vormen van formele metadata over publicaties. Ook authority-files kunnen daarin worden gepubliceerd. MADS is een alternatief van MARC-21 in XML-formaat. MADS is gemakkelijker te doorgronden dan MARC-21.

Archivarissen gebruiken Encoded Archival Context – Corporate Bodies, Persons and Families (EAC-CPF) voor het publiceren van informatie over de archiefvormers. Dit XML-formaat sluit aan bij het metadata schema ISAAR(CPF) (ICA 2004).

### **Effectiviteit**

Het gebruiken van een authority-file heeft de voorkeur boven het los invoeren van namen. Het helpt het publiek bij het eenduidig vinden van bronnen. Indien het daarbij noodzakelijk is om een eigen lijst te ontwikkelen is het verstandig deze lijst te delen, zodat ook anderen er gebruik van maken.

Het ontwikkelen en vooral het bijhouden van een authority-file is erg arbeidsintensief. Daar staat tegenover dat een beheerende instelling een prominente rol krijgt in het veld, dat zich ontwikkelt tot een relevant knooppunt in het netwerk van instellingen.

### **Juridische aspecten**

Houd er rekening mee dat stamgegevens die over levende personen gaan, vanwege privacywetgeving niet in alle gevallen mogen worden opgeslagen of beschikbaar mogen worden gesteld. Voor het maken en beschikbaar stellen van een lijst van auteurs is er geen juridische drempel, maar indien bijvoorbeeld oorlogscriminelen worden opgenomen in één systeem waardoor een collectiebeheerder naar een specifieke persoon kan verwijzen, is een nader onderzoek naar juridische aspecten op zijn plaats.

## **5.3 Thesauri gebruiken**

Een collectiebeheerder legt in de inhoudelijke metadata het onderwerp van een bron vast met behulp van één of meer woorden die het onderwerp zoveel mogelijk uitdrukken. Deze woorden (ook 'descriptor' genoemd) hebben allerlei onderlinge relaties. Zo is een "fiets" een soort voertuig en een "rijwiel" een ander woord voor fiets. Een thesaurus legt de onderlinge relaties tussen termen vast.

### **Aanpak**

Net als bij authority-files verdient het de voorkeur een bestaande thesaurus te gebruiken bij het maken van inhoudelijke metadata. Hiervoor zijn verschillende thesauri beschikbaar:

- Erfgoedthesaurus: <http://www.erfgoedthesaurus.nl/>, bevat termen relevant voor archeologie en bouwhistorie
- AAT: Art & Architecture Thesaurus, bevat termen relevant voor kunst en architectuur
- GTAA: Gemeenschappelijke Thesaurus Audiovisuele Archieven, in beheer bij het NIBG.
- Wikipedia: wikipedia kan als termenlijst worden gebruikt, waarbij het tevens mogelijk is het publiek direct achtergrondinformatie over het onderwerp te verstrekken.

### **Normen**

De open standaard die de onderlinge relaties tussen termen weergeeft heet Simple Knowledge Organization System (SKOS)<sup>16</sup>. Het registratiesysteem van de collectiebeheerder kan eenvoudiger van een thesaurus gebruik maken die beschikbaar is in SKOS-formaat.

### **Effectiviteit**

Indien een andere organisatie de trefwoordenlijst onderhoudt die je gebruikt, levert dat veel tijdsbesparing op. Daarnaast is het mogelijk verbanden te vinden tussen collecties die allemaal gebruik maken van dezelfde lijst. De toegankelijkheid van het materiaal is daardoor toegenomen.

---

<sup>16</sup> <http://www.w3.org/2004/02/skos/>

Maakt het feit dat informatie full-text doorzoekbaar is het toekennen van onderwerpstermen niet overbodig? Misschien. Maar ten eerste zijn nog lang niet alle teksten full-text beschikbaar. En ten tweede kan de relevantie die is toegekend aan een onderwerpskenmerk hoger zijn dan de aanwezigheid van een zoekterm in een tekst. Deze hogere relevantie kan een zoekmachine (zie hoofdstuk 9) gebruiken om de vindbaarheid van de bron te verbeteren.

## 5.4 Thesauri maken en publiceren

Ook bij het gebruik van een eigen lijst, is het aan te bevelen de lijst te publiceren zodat deze beschikbaar is voor anderen om te gebruiken.

### Aanpak

#### In de praktijk: WO2-trefwoordenlijst NIOD

Het NIOD beheert een lijst met termen die betrekking hebben op de Tweede Wereldoorlog. Het NIOD heeft de ambitie deze lijst naar SKOS te converteren en in OpenSKOS beschikbaar te stellen.

Er vindt alignment plaats met de erfgoedthesaurus zodat het publiek nog meer combinaties van termen kan gebruiken bij het vinden van materiaal.

Het is niet nodig een eigen omgeving in te richten waarmee een thesaurus beschikbaar wordt. OpenSKOS, beheerd bij Beeld en Geluid, is een omgeving waarin een woordenlijst in SKOS-formaat wordt aangeboden.

Bij verschillende instellingen ontstaan lijsten met trefwoorden met een verschillende doelstelling of afbakening. Vaak is er veel overlap tussen de lijsten: sommige komen in verschillende lijsten voor maar hebben dezelfde betekenis.

Ook trefwoordenlijsten kunnen om die manier overlappen. De website cultuurLINK helpt om de overlap in kaart te brengen en de relatie tussen termen waar mogelijk te leggen. <http://cultuurlink.beeldengeluid.nl/app/#/> Het koppelen van de termen in twee thesauri wordt *alignment* genoemd.

### Effectiviteit

Net als bij authority-files is het handig dat ook anderen gebruik maken van jouw lijst. De toegankelijkheid neemt daardoor toe.

Natuurlijk is het onderhoud van de lijst arbeidsintensief. Is het onderhouden van een thesaurus nog wel relevant in een tijd van zoekmachines? Een thesaurus kan ook zonder dat er trefwoorden aan zijn ontleend van belang zijn. Het is bijvoorbeeld mogelijk om in een zoekmachine zoektermen voor een full-textzoekopdracht te suggereren: indien de gebruiker zoekt op “fiets” kan de zoekmachine voorstellen ook op “rijwiel” of “voertuig” te zoeken. Deze structuren zijn vastgelegd in een thesaurus.

## 5.5 Plaatsnamenlijsten gebruiken

Plaatsnamen hebben andere eigenschappen dan namen van personen en organisaties, of trefwoorden die een onderwerp aanduiden. Net als bij personen is er extra informatie beschikbaar: geografische coördinaten bijvoorbeeld of spellingvarianten. Net als trefwoorden hebben de plaatsnamen een onderlinge relatie: Amsterdam ligt in Noord-Holland. Belangrijkste verschil is de tijd: in 1850 lag Amsterdam helemaal niet in Noord-Holland maar in Holland en was de gemeente Amsterdam bovendien veel kleiner.

### Aanpak

Er zijn veel verschillende lijsten van plaatsnamen beschikbaar:

- TGN: Thesaurus of Geographic Names (<http://www.getty.edu/research/tools/vocabularies/tgn/>), ontwikkeld en gehost door Getty
- Geonames: (<http://www.geonames.org/>), een collaboratief project
- gemeentegeschiedenis.nl, een plaats waar verschillende standaarden voor gemeentes aan elkaar zijn gerelateerd.

Al deze informatie is voor Nederland gecombineerd doorzoekbaar via erfgeo.nl. Deze omgeving is ontwikkeld en wordt gehost door Rijksdienst voor Cultureel Erfgoed. Neem in de metadata bij een verwijzing naar een plaats, de unieke verwijzing naar een plaatsnaam over uit één van deze systemen.

### Normen

Voor het koppelen van een plaatsnamenlijst aan een informatiesysteem zijn geen standaarden beschikbaar. Erfgeo.nl is gebaseerd op een API die een collectiebeheerder kan toepassen om de relatie tussen zijn collectiebeheersysteem en de geografische data te realiseren.

Wel is het handig na te gaan of de presentatie van kaartmateriaal is gemaakt met gestandaardiseerde webservices voor kaarten: Web Map Services (WMS) en/of Web Feature Services (WFS). Met behulp van deze webservices kan het kaartmateriaal worden ontsloten met behulp van een Geografische Informatie Systeem (GIS).

### Effectiviteit

Net als bij andere stamgegevens kan het publiek materiaal makkelijker en beter vinden als een collectiebeheerder gebruik maakt van gestandaardiseerde plaatsnamen. Plaatsen zijn daarnaast ook op de kaart weer te geven, waardoor een collectiebeheerder een Geografisch Informatiesysteem (GIS) kan inzetten voor het vinden en presenteren van plaatsgebonden informatie.

## 5.6 Plaatsnamenlijsten maken en publiceren

Plaatsnamen kunnen heel lokaal zijn: namen van huizen of velden, straat- en wijknamen zijn daarvan voorbeelden. Deze plaatsnamen zijn daarom meestal niet opgenomen in nationale of internationale lijsten. Een lokaal georiënteerde erfgoedinstelling

is bij uitstek deskundig als het gaat over deze lokale namen. Daarom zijn zij de aangewezen organisatie om deze informatie op te slaan en te publiceren.

Als een collectiebeheerder deze lokale geografische kennis op deze manier beschikbaar stelt, kunnen ook anderen er gebruik van maken. Helaas zijn er nog geen afspraken om het publiceren van deze geografische data te standaardiseren.

## 5.7 Verbetervoorstellen van het publiek verwerken

Net als bij metadata en digitale manifestaties bevatten ook stamgegevens fouten. Van personen en instellingen kan biografische informatie verkeerd zijn, plaatsnamen kunnen verkeerd op een kaart zijn getekend en thesaurustermen kunnen verkeerd zijn gerelateerd.

De manier waarop informatieprofessionals aanwijzingen van het publiek voor verbeteringen in stamgegevens doorvoeren, is hetzelfde als bij metadata. Zie daarvoor hoofdstuk 6 en paragraaf 10.4.

## 6. User Generated Content

Het publiek heeft in alle voorgaande hoofdstukken een rol gespeeld. Soms creëren gebruikers zelf metadata, digitale manifestaties of stamgegevens. Soms gebeurt dat onder begeleiding van een informatieprofessional, maar dat is allang niet meer vanzelfsprekend. Vaak bieden gebruikers de helpende hand door aan te geven waar fouten in de digitale informatie is geslopen. Zodra een collectiebeheerder zo'n correctie doorvoert, leidt dat tot een hogere kwaliteit van de digitale informatie en daarom tot betere vindbaarheid.

Sommige gebruikers creëren echter digitale informatie die als grondstoffen kunnen worden beschouwd: verhalen en discussies over allerlei gerelateerde onderwerpen. Deze zogenaamde *User Generated Content* (UGC) verbinden en samenbrengen met de bronnen zelf, helpt bij het vinden van bronnen. DEN biedt een uitgebreider overzicht over dit onderwerp <http://www.den.nl/thema/181/user-generated-content>.

### 6.1 Eigen bijdragen vastleggen en beschikbaar stellen

Gebruikers hebben naar aanleiding van een onderwerp of een object vaak een verhaal te delen.

#### Aanpak

Een erfgoedinstelling of een samenwerkingsverband kan op haar website de mogelijkheid bieden het verhaal van het publiek te publiceren. Veel gebruikers zijn echter al lid van Facebook, Twitter of Instagram en zullen een eigen platform kiezen om hun verhaal te kunnen doen. Hoe effectief het is om zelf een platform in het leven te roepen, dat de concurrentie aangaat met Facebook, is dus maar de vraag. Zelf het gebruik van het materiaal stimuleren op de bestaande platformen is veel effectiever. Dit kan door het publiek de mogelijkheid te geven gevonden materiaal te kunnen delen ("share") via een door hemzelf gekozen platform.

Een erg effectieve manier om erfgoed onder de aandacht te brengen is door deze beschikbaar te stellen op een wiki. Gebruikers kunnen hier interessante teksten toevoegen aan het materiaal, dat kan helpen bij het vinden van de bron. Om dit te stimuleren is het mogelijk een *wikipedian in residence* aan te stellen.

#### Normen

Het is noodzakelijk om het publiek te bewegen een verwijzing naar de bron op te nemen met behulp van een uniek internetadres. Hierdoor kan een collectiebeheerder nagaan op welke plaatsen een verwijzing naar een bron is gemaakt.

#### Effectiviteit

Het publiek maakt gebruik van zijn eigen woorden in een tekst over een bron. Het is aannemelijk dat deze woorden beter aansluiten bij de zoektermen die het brede publiek gebruikt om materiaal te vinden met een zoekmachine zoals Google. Een collectiebeheerder kan onderzoeken hoe hij in staat is om de teksten van het publiek

in zijn eigen zoekmachine beschikbaar te hebben, zodat het publiek het materiaal ook daar beter kan vinden.

## 6.2 Discussie faciliteren

Een discussie heeft een andere vorm dan de publicatie van een verhaal. Een website die de publicatie van een verhaal faciliteert, biedt vaak ook de mogelijkheid om op elkaar te reageren.

### **Aanpak**

Er zijn instellingen die digitaal aanwezig zijn op de platformen waar de discussie zich afspeelt. Het stelt de medewerker in staat om misverstanden uit de wereld te helpen over de betekenis of toegankelijkheid van bronnen. Dit wordt *webcare* genoemd.

Het volgen van de discussie kan leiden tot inzicht over de manier waarop het publiek omgaat met de bronnen en misschien wel tot de aanpassing van grondstoffen, om het vinden van de bronnen te verbeteren.

## **DEEL II: Beheer, behoud en beschikbaarstelling van grondstoffen**



## 7. Het applicatielandschap

We hebben ons in de voorgaande hoofdstukken beziggehouden met de digitale informatie – de grondstof – die bijdraagt aan zoeken, vinden en verrijken. Hoe meer van deze informatie beschikbaar is, hoe makkelijker het publiek kan zoeken, vinden en verrijken.

De kwaliteit en kwantiteit van grondstoffen alleen is een voorwaarde voor betere vindbaarheid, maar niet voldoende. Er is behoefte aan een samenspel van informatiesystemen waarin collectiebeheerders de grondstoffen opslaan, waardoor:

1. het beheer van de grondstoffen adequaat is georganiseerd,
2. het behoud van de grondstoffen kan worden gegarandeerd,
3. de beschikbaarstelling van de grondstoffen kan worden gerealiseerd.

De grondstoffen die gezamenlijk bijdragen aan toegankelijkheid zijn opgeslagen in een samenhangend landschap van informatiesystemen. Dit samenhangend landschap wordt omschreven in een “architectuur”.

### 7.1 Een drielagenmodel voor erfgoedontsluiting

Om greep te krijgen op de samenwerking tussen instellingen ontstond bij de Rijksdienst voor Cultureel Erfgoed het drielagenmodel voor erfgoedontsluiting.

Dit drielagenmodel groepeerde applicaties rondom drie functionele behoeften: (1) beheer en behoud, (2) samenbrengen en verbinden, (3) interactie met het publiek. Achtereenvolgens zijn deze lagen:

1. de data laag (voor beheer en behoud)
2. de verbindings- of referentielag (voor samenbrengen en verbinden), en
3. de presentatielaag (voor interactie met het publiek)

Collectiebeheerders kunnen met behulp van dit model de rol van softwareapplicaties die ze gebruiken eenduidig vastleggen. De applicaties kunnen variëren van applicaties voor metadatabeheer, tot digitale depots, zoekmachines en websites.

Belangrijkste doel van deze driedeling is om een heldere discussie te kunnen voeren over het samenspel van applicaties. In het verleden was het heel vaak zo dat één applicatie alle functionele rollen op zich nam. Dat betekende dat een collectiebeheerder een nieuwe website moest bouwen als hij de applicatie voor het metadatabeheer vervangde, en andersom. Dankzij het drielagenmodel is het besef ontstaan dat een applicatielandschap beter is te onderhouden indien een applicatie haar functies beperkt tot de functionaliteit van slechts één van de drie lagen.

## 7.2 De data laag beheert en behoudt

Het drielagenmodel plaatst applicaties in de data laag die zorgen voor de opslag van alle soorten grondstoffen, zoals omschreven in deel I. En opslag alleen is niet voldoende: collectiebeheerders moeten procedures inrichten waarmee geautoriseerde medewerkers de digitale informatie kunnen toevoegen, lezen, aanpassen en weggoien.

Een collectiebeheerder mag in zijn eigen applicaties kiezen voor zijn eigen datamodellen. De opslag en verwerking van informatie kan bij een collectiebeheerder namelijk heel specifiek zijn. Natuurlijk is het van belang na te gaan of aan de normen is voldaan die we op een rijtje hebben gezet in de voorgaande hoofdstukken, om de kwaliteit van de grondstoffen te waarborgen en om uitwisseling in de toekomst mogelijk te maken.

Er moeten procedures zijn om digitale informatie terug te vinden of om digitale informatie te reconstrueren in het geval van een calamiteit, bijvoorbeeld door een back-up te maken.

Voorbeelden van applicaties in de data laag zijn: een catalogus of collectiedatabase en een omgeving voor duurzame opslag. Zonder de applicaties in de data laag raken we de grondstoffen kwijt. Het NDE beleidsstuk “Nationale Strategie voor Digitaal Erfgoed” stelt daarom dat deze laag leidt tot houdbaarheid.

## 7.3 De verbindingslaag brengt samen en verbindt

De applicaties in de verbindingslaag brengen grondstoffen uit verschillende applicaties in de data laag samen. Vaak heeft het publiek namelijk de behoefte om te kunnen zoeken in verschillende collecties tegelijk. Daarnaast kan aanvullende informatie van elders, bijvoorbeeld in stamgegevens of User Generated Content, het publiek helpen om materiaal te begrijpen. Applicaties in de verbindingslaag halen daarom informatie uit verschillende bronnen op en zorgen ervoor dat de verbindingen helpen bij het vinden en begrijpen van bronnen.

### **In de praktijk: Europeana en Digitale Collectie**

[www.europeana.eu](http://www.europeana.eu) brengt metadata en afbeeldingen samen uit collecties van instellingen in heel Europa. Om te voorkomen dat Europeana contact moet onderhouden met de potentieel honderden instellingen waar ze materiaal van krijgen, roepen ze de hulp in van één organisatie per land die de grondstoffen uit dat land ophaalt. In Nederland is dat de Digitale Collectie ([www.digitalecollectie.nl](http://www.digitalecollectie.nl)). Andere verzamelpunten zijn gericht op het samenbrengen van informatie over één soort document. Archieven komen samen op het Europese platform [www.archivesportaleurope.net](http://www.archivesportaleurope.net). Museumobjecten komen samen in DiMCoN (te vinden op [www.digitalecollectienederland.nl](http://www.digitalecollectienederland.nl)).

Indien een applicatie in de verbindingslaag wegvalt zou dat niet moeten leiden tot het verlies van grondstoffen. Belangrijk is dat de applicaties in de verbindingslaag daarom géén grondstoffen bevatten die niet ook in een applicatie in de data laag aanwezig is. Applicaties in de verbindingslaag zijn niet ingericht op het behoud en beheer. Het NDE beleidsstuk “Nationale Strategie voor Digitaal Erfgoed” stelt dat de verbindingslaag leidt tot bruikbaarheid.

## 7.4 De presentatielaag zorgt voor interactie met publiek

Collectiebeheerders bereiken hun publiek via verschillende kanalen. Ze kunnen collectie-informatie beschikbaar stellen op websites gericht op verschillende doelgroepen of thema's. Maar ook de techniek kan verschillen: naast websites zijn er mobiele toepassingen. Ontwikkelingen gaan snel: wie weet welke techniek volgend jaar nodig is om het publiek te bereiken.

Het is daarom wenselijk dat de informatie die is samengebracht in de verbindingslaag en daar bruikbaar is gemaakt, op zoveel mogelijk manieren kan worden gepresenteerd. Opnieuw is de laag onafhankelijk van de onderliggende laag; applicaties in de presentatielaag kunnen worden aangepast of weggegooid, zonder dat het consequenties heeft voor de verbindingslaag. Voorbeelden van applicaties in de presentatielaag zijn verschillende websites en mobiele apps. Verschillende applicaties in de presentatielaag kunnen gebruik maken van dezelfde grondstoffen. Zo zijn oorlogskranten beschikbaar via delpher, maar ook op oorlogsbronnen.nl.

Zonder de applicaties in de presentatielaag kan het publiek de samengebrachte en verbonden bronnen niet vinden en de collecties niet zien. Het NDE beleidsstuk “Nationale Strategie voor Digitaal Erfgoed” stelt daarom dat de presentatielaag leidt tot zichtbaarheid.

## 7.5 Samenwerking tussen de lagen

Een heldere scheiding tussen de verschillende functies - beheer en behoud, samenbrengen en verbinden, en interactie met de gebruiker – leidt tot goede afspraken over de manier waarop de applicaties in de verschillende lagen met elkaar samenwerken. Hiervoor zijn vele open standaarden in het leven geroepen waarmee de data laag de grondstoffen aan de verbindingslaag aanbiedt en de verbindingslaag de grondstoffen ophaalt. De techniek van dergelijke standaarden is onafhankelijk van de techniek waarin de applicaties in de data laag en de verbindingslaag zijn gebouwd. Deze technieken worden verder uitgewerkt in de volgende hoofdstukken.

## 7.6 Toepassing bij een instelling

Het drielaagenmodel helpt om een architectuur bij één instelling in te richten. In één instelling beheert de collectiebeheerder namelijk verschillende grondstoffen vaak in verschillende collectiebeheersystemen. Om het zoeken in de gehele collectie mogelijk te maken, introduceert de collectiebeheerder daarom een applicatie in een verbindingslaag waarin hij de grondstoffen uit de verschillende systemen samenbrengt.

De website die hij vervolgens maakt op basis van de data in de verbindingslaag kan hij in de toekomst gemakkelijk vervangen. Naast deze website kan hij in de presentatielaag ook een mobiele app introduceren.

## 7.7 Toepassing in een domein

Collectiebeheerders van verschillende instellingen kunnen gezamenlijk zorgen voor betere toegankelijkheid. Er zijn de afgelopen jaren verschillende soorten samenwerkingsverbanden ontstaan, waarvan de meeste initiatieven metadata verzamelen om deze via één website doorzoekbaar aan te bieden. Voorbeelden zijn oorlogsbronnen.nl en EHRI.

### **In de praktijk: [www.ehri-project.eu](http://www.ehri-project.eu)**

De European Holocaust Research Infrastructure (EHRI) verzamelt data over de Holocaust afkomstig van instellingen door heel Europa in één centraal zoekstelsel. Dankzij dit systeem kan een Holocaust-onderzoeker op één website het materiaal vinden dat relevant is voor zijn of haar onderzoek.

De ontwikkelaars van het EHRI project maken onderscheid tussen de data-systemen die grondstoffen aanleveren en de zoekmachine die de grondstoffen combineert.

Het drielagenmodel helpt in een samenwerkingsverband om een architectuur in te richten. Collectiebeheerders kunnen in hun samenwerking afspreken wie verantwoordelijkheid neemt voor welke applicatie in welke laag. De scheiding van de verantwoordelijkheid zorgt ervoor dat ieder zich kan richten op zijn eigen taak.

## 8. Beheren en behouden

De applicaties in de data laag hebben als taak om de opslag van grondstoffen te regelen. De opslag dient zodanig te zijn dat de digitale collecties goed beheerd kunnen worden, goed behouden blijven, en goed beschikbaar zijn. De data laag vormt daarom de basis voor de functies in de verbindingslaag (samenbrengen en verbinden) en de presentatielaag (interactie met het publiek).

Adequaat beheer van de informatie zorgt voor correcte digitale informatie. De applicaties in de data laag bevatten daarom metadata van de objecten met – waar nodig – verwijzingen naar digitale manifestaties. Om de kwaliteit te waarborgen moet de collectiebeheerder de metadata en de digitale manifestaties kunnen aanpassen.

Een informatiesysteem dat het beheer van de collectie in al deze vormen ondersteunt, noemen we hier een collectiebeheersysteem (CBS). De namen voor dit systeem verschillen per documentsoort. Bij een bibliotheek spreekt men van een catalogussysteem, anderen spreken van collectie registratie.

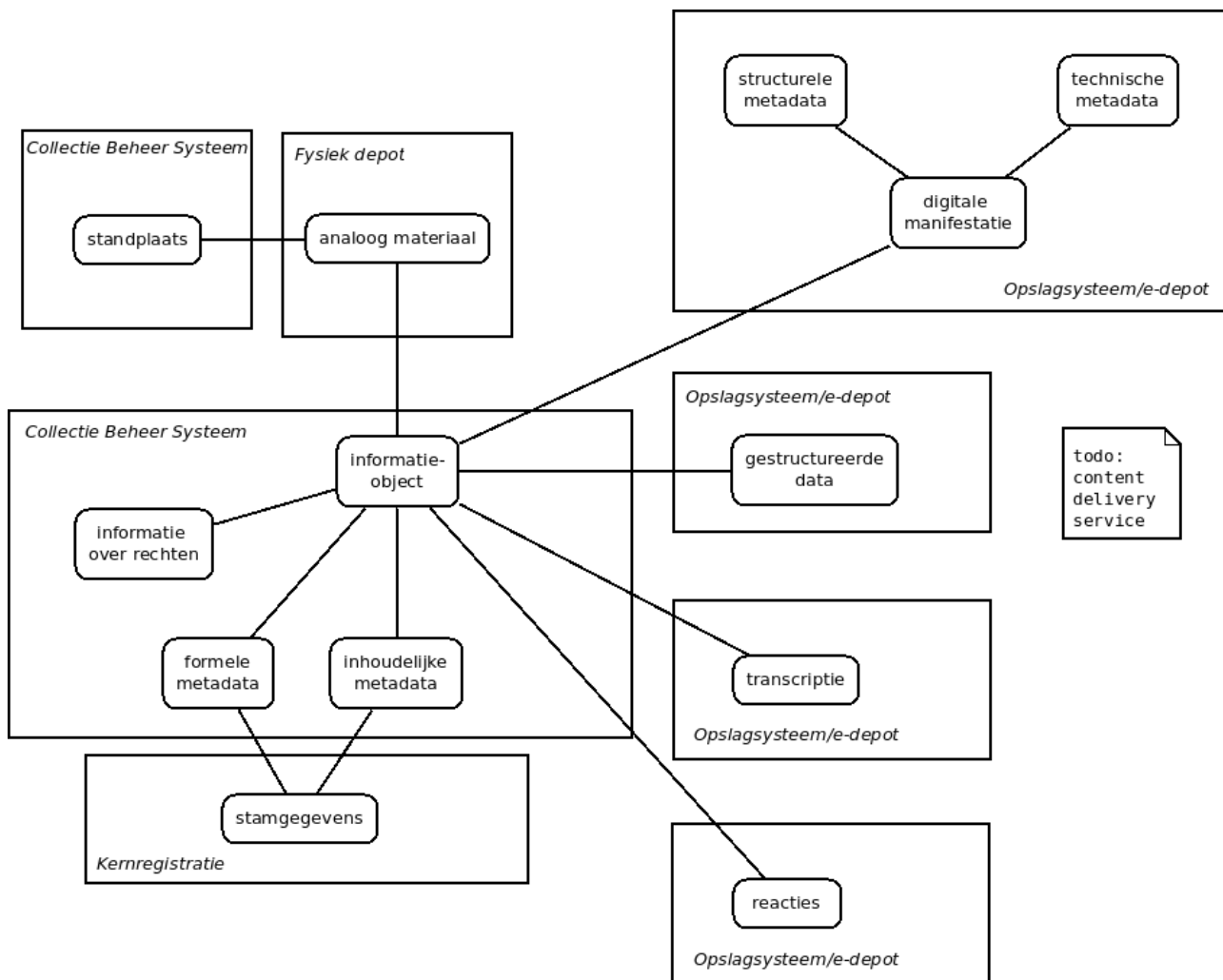
Vaak blijkt het handig om de digitale manifestaties van een object op te slaan in een apart opslagsysteem. Het collectiebeheersysteem legt dan een verwijzing vast naar de plaats waar de collectiebeheerder de digitale manifestatie heeft opgeslagen.

### 8.1 Architectuur rondom beheer en behoud

Het onderstaande schema geeft een voorbeeld van hoe de collectiebeheerder verschillende grondstoffen in informatiesystemen opslaat en duurzaam beheert. Zo'n applicatielandschap is bij een erfgoedinstelling idealiter omschreven in een architectuur. Daarnaast bepaalt een collectiebeheerder in een informatiebeleid hoe hij omgaat met het opnemen en afstoten van informatiesystemen.

Voor goed beheer van data is het belangrijk dat er slechts één plaats is waar een verantwoordelijk collectiebeheerder de data verbeteringen aanbrengt. Als de collectiebeheerder grondstoffen soms op de ene plek en soms in een kopie van de data verbetert, ontstaan onherroepelijk “meerdere versies van de waarheid”.

Soms is het nodig dezelfde grondstoffen op verschillende plaatsen op te slaan, bijvoorbeeld in een catalogus en ook in een zoekindex voor snelle doorzoekbaarheid. In zo'n geval is het wel belangrijk dat één plaats de bron van de data is en dat de andere afgeleiden zijn. De applicatie die een kopie van de data opneemt zorgt idealiter dat hij de grondstoffen kopieert via geheel automatische processen.



## 8.2 Inrichten van metadatabeheer

Het metadatabeheer geeft overzicht van de collectie die een collectiebeheerder beheert. Het systeem legt een verband tussen de metadata, de gestructureerde data, de scan en de transcriptie die aanwezig is. De applicatie zorgt daarnaast voor de beschikbaarstelling van de metadata.

Elke vorm van collectie heeft vaak een eigen applicatie voor metadatabeheer:

1. beeld – de beeldbank
2. bibliotheek – de catalogus
3. archief – het archiefbeheersysteem

### Aanbieders

Er zijn diverse leveranciers die software verkopen om de collectie mee te beheren. Er zijn veel (inter)nationale aanbieders. Daarnaast kan een collectiebeheerder er voor kiezen zelf een collectiebeheersysteem te ontwikkelen. Dit is echter vaak ingewikkelder en duurder dan een off-the-shelf pakket. Een collectiebeheerder kan daarom beter eerst onderzoeken of een off-the-shelf pakket aan de belangrijkste functionele behoeften voldoet, voordat hij een eigen systeem ontwikkelt of laat ontwikkelen.

## Normen

Belangrijk is dat de applicatie beschikt over een technische voorziening waarmee de collectiebeheerder informatie kan in- en uitlezen. Een dergelijke voorziening noemen we een interface. Voor het in- en uitlezen van informatie raden we aan om open standaarden te gebruiken. Indien de applicaties in de architectuur beschikken over interfaces op basis van open standaarden, is er sprake van Open Toegang (Anon 2015; van der Heiden and Zandhuis 2014). Het gebruik van open standaarden helpt voorkomen dat een collectiebeheerder onnodig afhankelijk wordt van één enkel (al dan niet commercieel) softwarepakket.

## Effectiviteit

Een informatiesysteem waarmee de collectiebeheerder zijn metadata beheert, is nodig om de informatie efficiënt op te slaan, aan te passen en beschikbaar te stellen. Zonder deze basisfunctie is adequate, duurzame toegankelijkheid op basis van de meest actuele informatie niet mogelijk.

## 8.3 Inrichten van opslag

Belangrijke toepassing om grondstoffen te beheren, is een omgeving waarin de collectiebeheerder scans, transcripties, gestructureerde data en reacties kan opslaan.

### Bestaande systemen

Een collectiebeheerder organiseert het opslaan van gedigitaliseerd materiaal vaak in een apart informatiesysteem. Een dergelijk systeem noemen we een Digital Asset Management (DAM) systeem, of een Content Management Systeem (CMS). De collectiebeheerder kan een dergelijk systeem inrichten in de eigen organisatie. Hij maakt dan gebruik van een groot opslagsysteem: een Storage Area Network (SAN) of Network Attached Storage (NAS). Steeds meer organisaties gaan echter geleidelijk over op het inkopen van opslagcapaciteit bij een collega-instelling of een commerciële partij.

### Normen

Aan een systeem voor de opslag stellen we verschillende eisen:

- Toegankelijk: het systeem moet de vereiste interfaces hebben, bij voorkeur op basis van open standaarden.
- Relationeel: via een dergelijke interface moet een relatie worden gelegd met de metadata in het collectiebeheersysteem.
- Schaalbaar: als het systeem vol is, moet de collectiebeheerder het gemakkelijk kunnen laten uitbreiden om de opslagcapaciteit te vergroten.

### Effectiviteit

De opslag van digitale manifestaties in een opslagsysteem stelt de collectiebeheerder in staat om sneller bronnen te vinden en beschikbaar te stellen. Veel beter dan als hij deze manifestaties op harde schijven in de kast bewaart.

## 8.4 Inrichten van duurzame opslag

De functie van de opslagsystemen die we omschreven in de voorgaande paragraaf, is beperkt tot het efficiënt beschikbaar hebben van digitale manifestaties: de collectiebeheerder kan de grondstof koppelen aan de beschrijving in het collectiebeheersysteem en daarna kan de digitale manifestatie voor collectiebeheerder en publiek gemakkelijk via het lokale netwerk of internet worden opgeroepen.

Een dergelijk opslagsysteem is niet speciaal ingericht om de digitale bestanden ook op lange termijn (5, 15, 50 jaar?) nog te kunnen bekijken en er zeker van te zijn dat we dan het document zien dat de collectiebeheerder heeft opgeslagen. De opslag in een opslagsysteem is al vele malen beter dan losse harde schijven, of – erger nog – cd-roms. Het controleren van deze losse dragers is namelijk veel te bewerkelijk om regelmatig te kunnen doen. Voor de duurzame opslag zijn aanvullende maatregelen nodig.



Afb. Doosje voor 5¼ inch floppydisks. De “5 Year Warranty” sluit aan bij een beroemde uitspraak. “Digital information lasts forever – or five years, whichever comes first”.

(Rothenberg, J., ‘Ensuring the Longevity of Digital Documents’, in: *Scientific American* (Vol. 272 (1995), Number 1, pp. 42-7)

(Collectie: IISG)

Zonder extra maatregelen zijn er een aantal risico’s aan de duurzame opslag van digitale bestanden. De drie belangrijkste volgen hieronder.

Allereerst kan er een storing optreden in het opslagmedium waardoor een deel van de opgeslagen bits verloren gaan. Het zou zomaar kunnen dat die bits zo’n grote rol spelen dat een gebruiker het document niet meer kan openen. Het verlies van bits noemen we vaak prozaïsch “bitrot”.

Ten tweede is het mogelijk dat de software waarmee een gebruiker het bestand opent niet langer beschikbaar is. Het zou kunnen dat moderne software in staat is het bestand weer te geven. Dat leidt echter niet altijd tot het gewenste resultaat. De huidige versies van Word openen een Wordperfect-document uit de jaren ‘90, maar het document ziet er niet meer uit zoals destijds.

De documentnaam die destijds is gekozen vormt het derde risico. “BRIEF.W51” zegt weinig, maar is door heel veel gebruikers destijds aan een bestand toegekend. Als een collectiebeheerder alle documenten in één omgeving opslaat, hoe vaak zou dan de bestandsnaam BRIEF.W51 voorkomen?



## Risico's verminderen

Een collectiebeheerder realiseert de digitale duurzaamheid van de opgeslagen bestanden niet met een computersysteem alleen. Sommige zaken neemt een computersysteem voor zijn rekening, maar duurzaam beheer vergt daarnaast veel mensenwerk. Dit werk is er op gericht om de risico's op verlies zo klein mogelijk te maken. Een samenhangend geheel van mensen en computers die zorgen voor beheer noemen we een e-depot.

Een collectiebeheerder voorkomt bitrot met twee maatregelen. Allereerst slaat zijn e-depot een reservekopie op, soms wel meerdere, op verschillende dragers en op verschillende geografische locaties. Het vormt het basisprincipe voor het project LOCKSS: "lots of copies keeps stuff safe"<sup>17</sup>. De kopie wordt teruggeplaatst zodra het systeem vaststelt dat er bitrot plaatsvindt. Het duurzaam opslagsysteem stelt de bitrot vast dankzij een techniek die controleert of de integriteit van het bestand nog intact is. Bij de opname in het opslagsysteem heeft het systeem daarvoor – als tweede maatregel – een getal uitgerekend op basis van de bits waaruit het opgeslagen bestand bestaat. Deze "hash-code" moet ook in de toekomst bij dezelfde berekening hetzelfde zijn. Zo niet: dan is er sprake van bitrot.

De software van het verleden kan alleen met heel veel moeite in stand worden gehouden. We noemen dit "emulatie". Meestal kiest een collectiebeheerder er voor een bestand te converteren naar een nieuw software-formaat; een formaat dat moderne software weer voor een langere periode kan lezen en weergeven. Het opslagsysteem registreert daarom bij de opname het softwareformaat. Zodra een software formaat niet langer kan worden gebruikt, ontvangt de beheerder een signaal en deze bepaalt welke actie hij onderneemt om de duurzaamheid weer voor langere tijd te kunnen garanderen. Vaak kiest hij ervoor het bestand opnieuw naar een actueel softwareformaat te converteren.

De eigenlijke bestandsnaam – zoals BRIEF.W51 – slaat de collectiebeheerder op als belangrijke informatie over de betekenis en inhoud van het bestand. Daarnaast kent het duurzame opslagsysteem een naam toe, die uniek is voor alle bestanden in het systeem of – beter nog – de hele wereld. Deze identificerende naam mag nooit meer veranderen. Hij wordt daarom een Persistent Identifier genoemd.

## Gemeenschappelijk taalgebruik

De problematiek rondom duurzame opslag heeft geleid tot een uitgebreid, nieuw jargon. Dat bleek nodig om met elkaar over de risico's, maatregelen en software te kunnen discussiëren, op zoek naar de juiste oplossing. Vaak zijn er ook leveranciers die (deel-)oplossingen leveren.

Het jargon heeft vooral vorm gekregen na de introductie van de ISO-standaard OAIS (Open Archival Information System). Het opnemen in het opslagsysteem heet daar bijvoorbeeld de "ingest"-functie. Om die opname mogelijk te maken moet een archiefvormer een Submission Information Package (SIP) aanleveren. De archivaris slaat deze als Archival Information Package (AIP) op in de Archival Storage. De

---

<sup>17</sup> <http://www.lockss.org/>

Access-functie levert aan het publiek – indien toegestaan – een Dissemination Information Package (DIP). Allemaal begrippen die een collectiebeheerder tegenkomt als hij met andere instellingen of leveranciers in het netwerk over digitale duurzaamheid gaat praten.

Voor het opslagsysteem wordt het woord e-depot gebruikt, of Trusted Digital Repository (TDR). Om te benadrukken dat een collectiebeheerder digitale duurzaamheid alleen realiseert in een samenwerkingsverband van mens en computer, rekent men tot het e-depot vaak “het geheel van organisatie, beleid, processen en procedures, financieel beheer, personeel, databeheer, databeveiliging en aanwezige hard- en software, dat duurzaam beheren en raadplegen van te bewaren digitale archiefbescheiden mogelijk maakt”.

### **Samenwerking**

Uit de bovenstaande tekst blijkt hoe veelomvattend digitale duurzaamheid is. Er zijn specialisten nodig met kennis van software-formaten, migratie-strategieën, functioneel en technisch beheer van hard- en software. De hard- en software zelf zijn grootschalig en daarom vaak duur.

Samenwerken is daarom onvermijdelijk. In het erfgoed domein wordt de samenwerking vorm gegeven door de Nationale Coalitie Digitale Duurzaamheid (NCDD). Veel publicaties op hun website geven inzicht in hoe de sector digitale duurzaamheid moet aanpakken.

Verschillende instellingen hebben de beschikking over een e-depot, elk met een eigen specialisme. Bij sommige instellingen is het mogelijk om collecties onder te brengen die niet tot de eigen collectie behoren. Het gaat daarbij om:

- Koninklijke Bibliotheek – voor publicaties
- Nationaal Archief – voor archief
- Instituut voor Beeld en Geluid – voor AV-materiaal
- DANS – voor gestructureerde data

Zelf aan de slag gaan is voor veel organisaties niet aan te bevelen. Een (Engelstalige) *whitepaper* van een project gehuisvest aan de Northern Illinois University (VS) geeft een overzicht van de functies die nodig zijn om digitale duurzaamheid te organiseren en beschikbare softwarepakketten die delen van deze functionaliteit mogelijk maken. <http://commons.lib.niu.edu/handle/10843/13610>

### **Normen**

Aan een e-depot kunnen we verschillende eisen stellen. OAIS vormt daarbij de kern van alle discussie. Deze standaard helpt om met elkaar te spreken over het systeem dat collectiebeheerders willen gebruiken.

DANS heeft een norm en een procedure ontwikkeld om een e-depot te kunnen testen. Na de procedure mag de collectiebeheerder het Data Seal of Approval voeren. Zie: <http://datasealofapproval.org/>

In samenwerking met PACKED heeft DEN een website ontwikkeld waarmee een collectiebeheerder kan toetsen welke verbeteringen hij moet doorvoeren om de digitale duurzaamheid van de organisatie te verbeteren. Dit resulteert in een rapport met aanbevelingen. Zie: <http://scoremodel.org/>

De meest uitgebreide norm voor een opslagsysteem voor digitale duurzaamheid is de ISO-standaard 16363: Audit and certification of trustworthy digital repositories. Deze relatief nieuwe standaard stelt een instelling in staat om een gecontroleerde omgeving te krijgen die de risico's op dataverlies in de toekomst zo klein mogelijk maakt.

### **Effectiviteit**

De vraag is hoeveel tijd en geld een erfgoedinstelling moet steken in het duurzaam opslaan van digitale manifestaties, waarvan de originelen meestal niet worden weggegooid. De collectiebeheerder moet een afweging maken tussen de kosten voor het duurzaam bewaren en de kosten om opnieuw te digitaliseren.

Omdat AV-materiaal op fysieke, vluchtige dragers is gemaakt en opgeslagen, geldt dat een collectiebeheerder ze digitaal – mits opgeslagen in een e-depot – duurzamer kan bewaren dan op de originele drager. Voor deze categorie is het duurzaam bewaren van de digitale manifestatie dus van groot belang. Voor kranten geldt iets vergelijkbaars. Vanwege de slechte kwaliteit van het papier waarop kranten werden gedrukt, is het de vraag hoe lang een instelling het papier duurzaam kan bewaren. De digitale manifestatie wordt, na verval van de krant, de enige manifestatie die overblijft. Ook deze moet de collectiebeheerder daarom duurzaam opslaan.

## **8.5 Persistent identifiers maken**

Het drielagenmodel heeft een belangrijke consequentie. De digitale informatie is niet meer alleen in de beheersystemen aanwezig. Ook andere informatiesystemen bevatten kopieën van dezelfde informatie. Allerlei aggregatoren brengen de digitale informatie samen en gebruiken de verbindingen die zijn gelegd met stamgegevens. Misschien stuurt deze aggregator de informatie weer door naar andere aggregatoren. Daarnaast biedt de presentatielaag allerlei functies die het nodig maken om informatie lokaal op te slaan.

De drie lagen opereren volledig onafhankelijk van elkaar. Hoe weten we nu zeker dat in een aggregator sprake is van hetzelfde object als in een website? Dat kan alleen als er sprake is van een identificatie van het object die onafhankelijk is van de laag: in de data laag krijgt het object een nummer of een code, onlosmakelijk verbonden met de digitale informatie.

Aan deze identificatiecode moeten drie voorwaarden worden gesteld:

1. hij is uniek in de hele wereld
2. hij is bruikbaar op internet
3. hij mag nooit meer wijzigen

## Uniek

Elke instelling heeft haar objecten van oudsher al eens voorzien van een nummer of een code. Afhankelijk van het soort object heet dit een signatuur, een vindplaatsnummer of een inventarisnummer. Binnen de betreffende instelling is dat een unieke code: er is in een museum maar één “object nummer 135”. Een ander museum heeft echter vaak ook een “object nummer 135”. Dus buiten de muren van het museum is het nummer niet uniek.

Speciale software is in staat om een lange, unieke code te genereren, die aan het object kan worden toegekend. Deze code wordt een (Universally Unique Identifier) UUID of Globally Unique Identifier (GUID) genoemd. Een UUID heeft bijvoorbeeld deze vorm:

```
b09056d8-78c6-102d-ba75-005056a23d00
```

Het googlen van deze code leidt naar een uniek geïdentificeerd informatieonderdeel van een groot archief in Nederland.

## Bruikbaar op internet

Internet is wereldwijd en om die reden is er behoefte om wereldwijd informatie te kunnen identificeren. Behalve het toekennen van een UUID, kan dit feit ons ook helpen om een object een unieke identificatie te geven. Door namelijk het internet-domein van de instelling toe te voegen, wordt de code uniek, bijvoorbeeld <http://www.algemeenrijksarchief.nl/nt00373/V729>. Dit heeft nog een belangrijk voordeel. Het webadres (of url) is namelijk behalve een identificatie tegelijk een adres waarop informatie over het object kan worden opgevraagd.

## Persistentie

Een webadres heeft echter ook een nadeel. Zodra een instelling van naam verandert (en vroeg of laat zal dat gebeuren), of een object naar een ander museum gaat, wil het museum ook het webadres veranderen, bijvoorbeeld naar <http://www.nationaal-archief.net/view?toegang=nt00373&nummer=V279>. Aggregatoren raken nu in de war. Ze denken dat er sprake is van een ander object. De identificerende code mag daarom niet meer veranderen. Nooit meer. Dat wordt persistentie genoemd.

Persistentie ontstaat doordat een webserver op een centrale plaats een ongewijzigde url vastlegt, die doorverwijst naar de meest actuele plaats op internet waar de informatie over het object beschikbaar is. Handle is daarvan een voorbeeld. Een handle ziet er bijvoorbeeld zo uit:

```
http://proxy.handle.net/10648/b09056d8-78c6-102d-ba75-005056a23d00
```

Als een gebruiker deze url invoert in de browser, wordt hij doorverwezen naar GaHetNA, de website van het Nationaal Archief. In de toekomst verandert wellicht de naam van deze website, maar de handle blijft dan bestaan en doorverwijzen naar de nieuwe webpagina op de nieuwe website.

[http://www.gahetna.nl/collectie/index/nt00373/b08bc244-78c6-102d-ba75-005056a23d00/sort\\_column/prs\\_tussenvoegsels/sort\\_type/asc/filtermeer/pla\\_naam\\_overlijdensplaats/filterpos/7/view/NT00373\\_oranjehotel](http://www.gahetna.nl/collectie/index/nt00373/b08bc244-78c6-102d-ba75-005056a23d00/sort_column/prs_tussenvoegsels/sort_type/asc/filtermeer/pla_naam_overlijdensplaats/filterpos/7/view/NT00373_oranjehotel)

Merk op dat het Nationaal Archief de mogelijkheid om een UUID toe te kennen combineert met een unieke handle.

### **Norm**

In Nederland heeft het Handle-protocol de meeste aanhang gekregen om objecten uniek te identificeren. Er zijn echter ook andere protocollen. Een overzicht is gepubliceerd in: [www.ica.org/download.php?id=612](http://www.ica.org/download.php?id=612)

Voor digitale publicaties gebruiken uitgevers vaak een Digital Object Identifier (DOI). Een voorbeeld hiervan ziet er als volgt uit <http://dx.doi.org/10.17026/dans-2ah-rtay>

Er is veel voor te zeggen om alle objecten in een instelling te nummeren vanaf 1. Maar vaak is er een keer behoefte om er een object tussen te voegen. Is dat nieuwe object dan object “52a”? Of hernoemen we object “52” dan naar “52a” en wordt het nieuwe object dan “52b”? Dat is verwarrend en leidt tot een situatie die een uniek toegewezen naam in gevaar kan brengen. Dit komt omdat op deze manier het nummer een betekenis heeft gekregen: blijkbaar is de volgorde van de objecten van belang. Het voordeel van de handle is dat deze (vrijwel) betekenisloos is. Alleen het nummer “10648” betekent “Nationaal Archief”. Er is dus niet de neiging om veranderingen te willen aanbrengen in de naamgeving om de betekenis te kunnen behouden.

### **Efficiëntie**

Het up-to-date houden van het register waarin op basis van een handle wordt doorverwezen naar de juiste website kan heel arbeidsintensief zijn. De hoeveelheid werk hiervoor is het grootst bij de introductie van nieuwe url's op de website van de instelling. De hoeveelheid werk die gepaard gaat met het afhandelen van allerlei vragen over vreemde verwijzingen weegt ruimschoots op tegen deze extra werkzaamheden rondom een persistent identifier.

## **8.6 Reacties opslaan**

Nadat materiaal beschikbaar is gesteld aan het publiek via een website wil het publiek vaak graag op het materiaal reageren. Natuurlijk kan een geïnteresseerde gebruiker hiervoor gebruik maken van zijn Facebook of Twitter account. Indien een collectiebeheerder deze functionaliteit aanbiedt op zijn eigen website, kan hij eenvoudig de reactie verwerken in zijn eigen collectiebeheersysteem.

In eerdere paragrafen hebben we beschreven hoe het publiek betrokken kan zijn bij het maken van grondstoffen. Dat noemen we crowdsourcing. Een apart hoofdstuk is gewijd aan User Generated Content: een situatie waarbij gebruikers zelf bronnen gebruiken in hun blogs of verhalen vertellen naar aanleiding van hun eigen archiefmateriaal.

Deze paragraaf gaat in op de reacties die het publiek maakt over een gepubliceerde bron.

### Aanpak

Zoals bij *User Generated Content* kan een collectiebeheerder de reacties vinden op verschillende platformen: Facebook, Twitter, Instagram. Hij kan ook op zijn eigen website een reactiemogelijkheid aanbieden. In alle gevallen kan het handig zijn om de reactie als erfgoedinstelling zelf op te slaan en te relateren aan het object. Misschien leidt de reactie zelfs tot een aanpassing van de metadata of de digitale manifestatie.

### Normen

Er zijn geen formele normen waaraan het opslaan van een reactie moet voldoen. Het is netjes als de collectiebeheerder alle reacties gelijkwaardig behandelt.

### Effectiviteit

Een klein deel van de investering bestaat uit het maken van een forum en het inrichten van een reactieformulier. Het grootste deel van de investering bestaat uit het opbouwen van een gemeenschap die zich met de erfgoedinstelling verbonden voelt.

## 8.7 Aanbieden aan website of aggregator

Alle grondstoffen aanwezig in de data laag zijn blijkbaar het bewaren waard. Het is daarom ook de moeite waard om deze beschikbaar te hebben om te gebruiken. De applicaties in de data laag moeten om die reden allemaal zijn voorzien van een faciliteit die zorg draagt voor beschikbaarstelling.

### Aanpak

Deze beschikbaarstelling moet op geen enkele manier een belemmering zijn voor de manier waarop iemand het wil gebruiken. De collectiebeheerder stelt de grondstoffen daarom in zijn meest pure en complete vorm beschikbaar.

### Normen

De meest pure en complete vorm waarin de grondstoffen beschikbaar zijn, wordt bepaald door het uitwisselformaat waarin de data beschikbaar is. Elke soort grondstof heeft daarvoor een eigen formaat, meestal gebaseerd op XML:

Soort grondstof	Uitwisselformaat (afkorting)	Afkorting staat voor:
Metadata van publicaties	MARC, MODS	MAchine-Readable Cataloging Metadata Object Description Schema
Metadata van archieven	EAD	Encoded Archival Description
Metadata van museumobjecten	LIDO	Lightweight Information Describing Objects
Transcripties	TEI	Text Encoding Initiative
OCR-resultaten	ALTO	Analyzed Layout and Text Object

Metadata in het algemeen (minimumset)	Dublin Core	
Authority File voor auteurs	MADS	Metadata Authority Description Schema
Authority File voor archiefvormers	EAC-CPF	Encoded Archival Context for Corporate Bodies, Persons and Families
Thesaurus	SKOS	Simple Knowledge Organization System

De collectiebeheerder moet de grondstoffen in de meest passende formaat beschikbaar stellen met behulp van een handig protocol. In dat protocol spreken we af hoe de communicatie tussen de computers die de informatie uitwisselen verloopt.

De meest toegepaste techniek hiervoor is Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH). Dit protocol schrijft voor: welke metadata wordt overgedragen, in welk (XML-)formaat en met behulp van welke commando's. Het feit dat is voorgeschreven hoe dit werkt, maakt dat een collectiebeheerder er op kan vertrouwen dat een aggregator een collectiebeheersysteem op een afgesproken manier kan aanspreken. Door het juiste commando te versturen, weet het collectiebeheersysteem welke informatie moet worden teruggestuurd.

Het Open Archives Initiative heeft een standaard ontwikkeld – ResourceSync genaamd – waarmee een collectiebeheerder naast de metadata ook de bronnen zelf kan aanbieden. Indien deze – betere – standaard ook op brede schaal wordt geadopteerd, kan een collectiebeheerder beter kiezen voor ResourceSync dan voor OAI-PMH.

Een collectiebeheerder kan de mogelijkheid bieden dat een gebruiker een selectie van de grondstoffen ophaalt en verwerkt. Veel gebruikte standaarden daarbij zijn SRU (Search and Retrieve via URL) of OpenSearch. SRU wordt voor deze functie toegepast bij de Digitale Collectie Nederland. In het algemeen kan een Application Programming Interface (API) voorzien in de behoefte om data uit de data laag beschikbaar te stellen. Data komt dan vaak beschikbaar in het formaat JSON (JavaScript Object Notation).

## 8.8 Enkele juridische aspecten

Een collectiebeheerder heeft te maken met enkele juridische aspecten bij:

1. het opnemen van data in de informatiesystemen
2. het beschikbaar stellen van data

### Opname van data

Bij de opname van data is het van belang na te gaan of de opslag van de data is toegestaan. Het opslaan van informatie over personen mag bijvoorbeeld alleen onder bepaalde voorwaarden.

Bij de opname van data moet de collectiebeheerder nagaan of wetten het gebruik in de toekomst uitsluiten. Als er auteursrechtelijke beperkingen zijn die de beschikbaarstelling beïnvloeden, moet de collectiebeheerder deze beperkingen bij de opname uitzoeken en vastleggen. Misschien kan hij zelfs beslissen de informatie niet in beheer te nemen als beschikbaarstelling niet is toegestaan.

### **Beschikbaarstelling van data**

De beschikbaarstelling van data kan aan regels zijn gebonden. Als het goed is, kan de collectiebeheerder op basis van de vastgelegde metadata bepalen of iemand de grondstoffen mag ophalen. Bij de beschikbaarstelling moet de collectiebeheerder deze regels toepassen.

#### **In de praktijk: Koninklijke Bibliotheek**

De Koninklijke Bibliotheek bepaalt op basis van de metadata of een werk online beschikbaar kan worden gesteld. Dit systeem, SeSAM genoemd, bevat de regels (“business rules”) en past deze toe. Indien bijvoorbeeld het jaar van uitgave langer dan 140 jaar geleden is, is toegang mogelijk. Of: indien de gebruiker in de studiezaal is, is toegang mogelijk.

De toegang kan worden bepaald op basis van:

- metadata (publicatiejaar, sterfjaar auteur ...)
- context van het verzoek (wie?, waar vandaan?)
- uitzonderingslijsten (lijsten van werken met een beperking)



## 9. Samenbrengen en verbinden

Het publiek is meestal pas in tweede instantie geïnteresseerd in de instelling waar hij of zij oorlogsbronnen kan vinden. Het publiek stelt zich allereerst de vraag: “welke oorlogsbron geeft antwoord op de vragen waarin ik ben geïnteresseerd?”

Daarom is het handig om informatie over oorlogsbronnen aan te bieden via één centrale plaats. Dit biedt bovendien de mogelijkheid om informatie uit verschillende collecties te combineren. Het wordt dan mogelijk om gemakkelijk van de ene naar de andere bron door te verwijzen, zonder onderscheid te maken tussen instellingen.

Dit bijeenbrengen van informatie noemen we *aggregatie*: “het op één plaats presenteren van inhoud uit meerdere online databases of websites”. Dankzij de aggregatie kunnen deze bronnen in één overzicht, bijvoorbeeld een mash-up site, gepresenteerd of met één vraag doorzocht worden. [<http://www.den.nl/abc/Aggregatie/>]

Om aggregatie mogelijk te maken, haalt (of “harvest”, “oogst”) de aggregator allereerst via internet de beschikbare grondstoffen op van verschillende instellingen. Daarna kan de aggregator informatie combineren of nieuwe informatie maken, die aanvullend is op de metadata die al is opgehaald.

### In de praktijk: oorlogsbronnen.nl

Via de website [oorlogsbronnen.nl](http://oorlogsbronnen.nl) zijn verschillende collecties te doorzoeken die verband houden met de Tweede Wereldoorlog. De website biedt de mogelijkheid om een zoekterm in te voeren, waarna hij een zoekresultaat presenteert met daarin verwijzingen naar informatie over collecties bij verschillende instellingen: oorlogsgraven geregistreerd bij de oorlogsgravenstichting, archiefbeschrijvingen en beeldbanken bij archiefdiensten, boeken en kranten bij bibliotheken.

De aggregator kopieert metadata over materiaal met behulp van OAI-PMH geautomatiseerd van de collectiebeheerder naar [oorlogsbronnen.nl](http://oorlogsbronnen.nl), waardoor het mogelijk wordt de informatie in één zoekmachine op te nemen en van daaruit aan te bieden.

Behalve informatie over de collectie kan allerlei informatie die gebruikers op internet maken in de aggregator worden meegenomen. Deze *User Generated Content* levert een extra verrijking op die het publiek helpt materiaal te vinden, maar ook collectiebeheerders nieuwe relevante informatie kan opleveren.

### 9.1 Redactie voeren

Als de eigenaar van een aggregator grondstoffen verzamelt rond een thema, zoals bij Oorlogsbronnen, is het belangrijk te bepalen welke bronnen hij wel en welke bronnen hij niet in de aggregator opneemt. Is het materiaal relevant voor het gekozen

thema? Biedt de collectie een interessant meerwaarde?

Na deze inhoudelijke afweging, is het noodzakelijk na te gaan:

- of de collectie-informatie van voldoende kwaliteit is, en
- of de collectie-informatie op een eenvoudige manier kan worden opgehaald.

Een permanente redactie die de informatie op een goede manier waardeert en de opname organiseert, is daarbij van belang. Deze redactie is in staat om te reageren op behoeftes van het publiek, zowel wat betreft de informatie die wordt opgenomen als wat betreft de functionaliteit die wordt aangeboden om de informatie te doorzoeken.

## 9.2 Ophalen van grondstoffen

Een aggregator kan het ophalen van de informatie over of van een collectie op verschillende manieren organiseren: niet geautomatiseerd, geautomatiseerd zonder selectie en geautomatiseerd met selectie.

### Niet geautomatiseerd

Een collectiebeheerder kan de grondstoffen bij een aggregator aanleveren door middel van een cd-rom, een usb-stick of een e-mail. De beheerder van de aggregator kan de aangeleverde informatie inlezen. In uitzonderlijke gevallen is er geen mogelijkheid om het anders te organiseren. Bijvoorbeeld omdat de collectie-informatie niet is vastgelegd in een applicatie voor metadatabeheer of een applicatie die de metadata niet via een netwerk aanbiedt.

#### In de praktijk: archievenwo2.nl

Aan de website archievenwo2.nl is in 2008 eenmalig informatie aangeleverd over archiefstukken die betrekking hebben op de Tweede Wereldoorlog. Sindsdien zijn sporadisch wijzigingen doorgevoerd in de informatie.

Gelukkig veranderen de archiefbeschrijvingen hoogst zelden, maar de kans bestaat dus dat archievenwo2.nl verouderde informatie biedt. De informatie is overgenomen door oorlogsbronnen.nl, zodat ook daar deze niet actuele informatie wordt gepresenteerd.

Archievenwo2.nl actualiseert de informatie regelmatig, maar beheerders kunnen een efficiëntere techniek inzetten.

Deze manier van werken heeft een aantal belangrijke nadelen en is om die reden niet aan te bevelen. Een collectiebeheerder kan de wijzigingen in de aggregator niet eenvoudig doorvoeren. Het verkrijgen van verbeterde informatie van de collectiebeheerder, omzetten tot een digitaal formaat dat de aggregator kan inlezen en vervolgens het inlezen zelf, is een arbeidsintensief proces. Het is de vraag of een organisatie deze taak duurzaam kan beleggen. Vroeg of laat leidt het tot informatie die niet actueel is, en daarom bij het publiek tot verwarring leidt.

### **Geautomatiseerd ophalen zonder selectie**

Het is wenselijk dat een kopie van de informatie geautomatiseerd vanuit een applicatie voor metadata-beheer wordt overgebracht naar de aggregator. Zonder standaarden moeten collectiebeheerders en beheerders van een aggregator één-op-één afspraken maken per collectiebeheersysteem over de manier waarop informatie wordt overgedragen. De collectiebeheerder stelt daarvoor de grondstoffen beschikbaar met behulp van standaarden (zoals OAI-PMH en ResourceSync) zoals omschreven in paragraaf 8.7.

### **Geautomatiseerd ophalen met selectie**

OAI-PMH en ResourceSync zijn gericht op het ophalen en up-to-date houden van een volledige verzameling van allerlei vormen van digitale informatie. Een thematische aggregator kan daarom collecties ophalen die als geheel betrekking hebben op het thema.

Collectiebeheerders bieden hun collecties echter niet per definitie thematisch aan. Daarom is het soms nodig om een selectie te kunnen oogsten om op te nemen in de aggregator. De aggregator kan dan een selectie maken met behulp van een zoekvraag. Om dit mogelijk te maken, is het nodig dat de aanbiedende partij de informatie zodanig beschikbaar stelt dat een aggregator een zoekvraag kan opsturen en er een antwoord terugkomt. De standaarden die daarvoor zijn ontwikkeld – variërend van SRU tot specifieke API's – zijn omschreven in paragraaf 8.7.

## **9.3 Verbeteren van grondstoffen**

### **Normaliseren**

Informatie die binnenkomt, is lang niet altijd op dezelfde manier gestructureerd of op hetzelfde kwaliteitsniveau ingevuld. Bij binnenkomst is het dan mogelijk om de informatie te normaliseren. Daarmee bedoelen we dat de aggregator deze informatie zoveel mogelijk tot eenzelfde norm opwerkt.

Een aggregator kan bijvoorbeeld de datums normaliseren die in de metadata zijn opgenomen. Indien de aggregator deze allemaal op dezelfde manier opslaat, is het mogelijk om de zoekmachine sneller materiaal te laten selecteren die betrekking heeft over een bepaalde periode.

- “3 juli 1944” wordt “19440703”
- “1943” wordt “19430101-19431231”

Ook velden waarin persoonsnamen staan kan een aggregator opknippen in voornamen, tussenvoegsel en familienaam, indien het nodig is voor het doel van de aggregator.

### **Combineren**

In de aangeleverde metadata zijn wellicht verwijzingen opgenomen naar andere informatiebronnen. De collectiebeheerder heeft in de metadata hopelijk een verwijzing gemaakt naar stamgegevens, zodat in de beschrijving van een boek bijvoorbeeld een

verwijzing is opgenomen naar een centraal geregistreerde auteursnaam. De metadata over het boek en de stamgegevens over de auteur kan de aggregator combineren.

Helaas is niet alle data even compleet en correct. Men name als het gaat over het gebruik van stamgegevens kunnen collectiebeheerders hun grondstoffen op verschillende manieren hebben ingevuld. De ontwikkelaar van de aggregator moet zo helder mogelijk communiceren over de tekortkomingen van de data, zodat de ontwikkelaar van de applicatie in de presentatielaag hiermee rekening kan houden.

### **Aanvullen**

Een aggregator kan misschien op basis van de aangeleverde grondstoffen extra informatie toevoegen. Aan een plaatsnaam of adres kan hij bijvoorbeeld coördinaten toevoegen, zodat een website in de presentatielaag een kaart kan tonen. De informatie in de aggregator is dan rijker geworden dan hij was in het collectiebeheersysteem.

Een ander voorbeeld. Niet alle collectiebeheerders bieden metadata aan die verwijst naar dezelfde stamgegevens. Een aggregator kan een extra verwijzing toevoegen aan een metadataveld waarin een trefwoord is opgenomen. In het trefwoordenveld waarin het woord “fiets” is ingevuld kan de aggregator het woord fiets koppelen aan de descriptor “fiets” in een thesaurus. Dankzij deze extra verwijzing wordt het mogelijk om in een zoekmachine te zoeken op “fiets”, “rijwiel” of “voertuig”. In alle gevallen kan het publiek het object nu vinden.

### **Verbeteringen apart opslaan**

De aggregator moet de aanvullende informatie die zij maakt, apart opslaan. De gebruiker van de aggregator moet namelijk (in de presentatielaag) duidelijkheid hebben over welke grondstoffen afkomstig zijn van de collectiebeheerder en welke door de aggregator is toegevoegd.

### **Verbeteringen teruggeven**

Een collectiebeheerder kan de behoefte hebben om de aanvullende informatie die over zijn eigen grondstoffen is gemaakt op te halen en te bewaren als een nieuwe grondstof in zijn eigen collectiebeheersysteem. In het volgende hoofdstuk beschrijven we de technieken die de collectiebeheerder daarvoor kan inzetten.

## **9.4 Diensten ontwikkelen en aanbieden**

De aggregator heeft als taak de verkregen grondstoffen met haar verbindingen en verbeteringen weer aan te bieden aan een gebruiker. Meestal zal de gebruiker in de presentatielaag een website bouwen waarmee het publiek informatie kan vinden.

Nadat de aggregator alle relevante grondstoffen heeft verzameld en verbeterd, kan zij allerlei digitale diensten ontwikkelen. De dienst die de meeste aggregators aanbieden is het doorzoeken van de data. Een gebruiker kan dan een zoekvraag stellen en het antwoord verwerken tot een interessante presentatie op een bepaald medium (bijvoorbeeld een app), gericht op een specifiek publiek (bijvoorbeeld toeristen).

Net als de collectiebeheerder, moet ook de beheerder van de aggregator zijn data en de zoekdiensten daarvoor aanbieden met behulp van open standaarden. Daarbij kan hij kiezen uit dezelfde standaarden als de collectiebeheerder gebruikt bij het aanbieden van zijn grondstoffen.

De diensten komen tot stand met behulp van technieken die een zoekvraag ontvangen en beantwoorden. Dit zijn SRU en Open Search. Een specifieke interface noemen we een Application Programming Interface (API). De aggregator kan ook de data als geheel weer doorsturen aan een ander. Hierbij maakt hij gebruik van OAI-PMH of ResourceSync.

## 9.5 Enkele juridische aspecten

De aanbieder van de grondstoffen, dat wil zeggen de collectiebeheerder, is er verantwoordelijk voor dat hij de grondstoffen aanbiedt binnen de juiste juridische kaders. Een aggregator kan er voor kiezen alleen grondstoffen op te nemen die voldoen aan een bepaald profiel, bijvoorbeeld alleen materiaal dat in het publieke domein valt. In dat geval hoeft de aggregator weinig te doen om te bewaken of hij het materiaal beschikbaar mag stellen. Zodra ook andere informatie is opgenomen moet de aggregator bewaken dat zij zich aan de geboden kaders houdt.

## 10. Interactie met publiek

Uiteindelijk is al het werk dat collectiebeheerders en aggregatoren uitvoeren bedoeld om het publiek te voorzien van de juiste informatie: collectiebeheerders hebben grondstoffen geproduceerd, al deze digitale informatie gestructureerd opgeslagen en in beheer genomen. De aggregator heeft de digitale informatie gecombineerd vastgelegd en doorzoekbaar aangeboden aan het publiek. Om de toegankelijkheid af te ronden, ontbreekt een handige website waarin het publiek de zoekvraag kan stellen, het resultaat kan bestuderen en de grondstoffen kan gebruiken. En: een juiste visie op toegankelijkheid leidt niet alleen tot beschikbaarstelling, maar ook tot een wederkerige relatie met het publiek. Van een instelling uit de 21ste eeuw mag worden verwacht dat zij via internet een band opbouwt waarbij het publiek reageert en de instelling haar publiek daarvoor iets teruggeeft.

De interactie met het publiek delen we op in de vier activiteiten die het publiek uitvoert:

1. vinden
2. bekijken
3. gebruiken
4. verrijken

Belangrijkste functionaliteit die een website realiseert, draagt bij aan het vinden en bekijken van collectiemateriaal. Het *vinden* maakt een website mogelijk dankzij een zoekformulier waarin een gebruiker een zoekvraag kan invoeren, die de website vervolgens doorstuurt aan de zoekmachine. De zoekmachine geeft een resultaatlijst, die de website omzet in een webpagina die de gebruiker kan bestuderen. Zodra een gebruiker in de lijst op een resultaat klikt, toont de website de informatie over het materiaal of de digitale representatie van het materiaal zelf aan de gebruiker zodat deze het kan *bekijken*. Indien een digitale representatie beschikbaar is, kan deze in een viewer worden getoond. Als dat niet het geval is, komt een gebruiker naar een studiezaal om het materiaal te bestuderen.

Nadat het materiaal is gevonden en bekeken, zal het publiek het materiaal willen *gebruiken*. Gebruikers downloaden informatie of slaan verwijzingen op naar voor hen relevant materiaal en ordenen het materiaal. Ze zijn curator van hun eigen deelcollectie. Ze willen hun kennis over en affiniteit met hun vondst graag delen via sociale en collaboratieve media. Denk daarbij aan Facebook, Twitter en wikipedia.

Het publiek heeft veel verstand van de collectie en wil de informatie die over het materiaal beschikbaar is, vaak aanvullen of verbeteren. Meestal zijn ze bereid om informatie over collectieonderdelen met elkaar of met de collectiebeheerder te delen. Deze activiteiten noemen we *verrijken*.

## 10.1 Vinden

Belangrijke taak die een collectiebeheerder heeft, is om het mogelijk te maken dat het publiek in staat is om in de collectie materiaal te vinden dat antwoord geeft op de (onderzoeks-)vraag die de gebruiker zich gesteld heeft. Natuurlijk gebeurt dit via internet.

De gebruiker heeft daarvoor op een website de beschikking over een formulier waarmee een zoekvraag kan worden uitgedrukt. Veel gebruikers zijn intussen gewend om dit te doen op de manier die Google aanbiedt: door middel van één veld waarin een zoekterm kan worden ingevoerd. Er kunnen ook twee of meer termen worden ingevoerd, waarna Google op zoek gaat naar resultaten waarin beide termen voorkomen. Indien de gebruiker de zoektermen tussen aanhalingstekens plaatst, gaat Google op zoek naar precies deze combinatie. Indien een website een zoekinterface voor erfgoed aanbiedt, is het aan te bevelen deze *de facto* afspraken te volgen om de gebruiker niet in verwarring te brengen over de manier waarop zij een zoekvraag kan stellen.

Weinig gebruikers maken gebruik van een formulier voor “uitgebreid” of “geavanceerd” waarin zij een gedetailleerde vraag kunnen stellen. Verfijning van een zoekresultaat gebeurt meestal achteraf, dankzij de weergave van “facetten”. Met deze facetten kan de gebruiker meestal documentsoorten, periodisering of vervaardigers kiezen, waardoor in een nieuw zoekresultaat alleen de resultaten worden getoond die betrekking hebben op de gekozen documentsoort, periode of vervaardiger.

In de weergave van het zoekresultaat is het belangrijk aan te geven waarom de zoekmachine het collectiemateriaal in het zoekresultaat heeft opgenomen. Google doet dit door een fragment van de gevonden tekst in het zoekresultaat weer te geven, waarbij de gevonden zoekterm gemarkeerd is.

De zoekmachine kan de gebruiker helpen bij het kiezen van een relevante zoekterm. Ook hier is Google het grote voorbeeld: de zoekmachine suggereert andere of aanvullende zoektermen die op de ingevoerde zoekterm lijken, of daarmee in verband staan. Een thesaurus kan hierbij een rol spelen. Indien de gebruiker “fiets” als zoekterm heeft gekozen, kan de zoekmachine op basis van een thesaurus suggereren dat de gebruiker ook op “rijwiel” kan zoeken.

## 10.2 Bekijken

De gebruiker kan het gevonden materiaal vervolgens op verschillende manieren bekijken. Een archiefstuk heeft een eigen presentatiepagina of *landingpage*. In de meest gebruiksvriendelijke vorm biedt de website – indien beschikbaar – een digitale representatie van het materiaal aan. In de viewer die daarvoor nodig is, kan de gebruiker bijvoorbeeld op de afbeelding inzoomen, de afbeelding kantelen, of het contrast aanpassen.

Indien de gebruiker materiaal heeft gevonden waarvan geen digitale representatie beschikbaar is, kan hij het origineel komen bekijken in de studiezaal. De website kan daarbij de mogelijkheid bieden om het materiaal te reserveren. Dit heeft als voordeel dat de gebruiker er zeker van kan zijn dat het materiaal beschikbaar is (en niet door een ander wordt bekeken, of voor restauratie niet raadpleegbaar is), maar ook dat het personeel het materiaal klaar kan leggen.

Ook is het mogelijk dat de gebruiker een bestelling plaatst om een digitale representatie te laten maken, al dan niet tegen een vergoeding. Dit noemen we scanning on demand (zie paragraaf 4.2).

### 10.3 Gebruiken

Gevonden materiaal kan voor de gebruiker zo relevant zijn dat deze het wil bewaren. Voor dat bewaren zijn twee mogelijkheden: ofwel de gebruiker kan de digitale informatie (zoals de transcriptie, de metadata of de scans) downloaden, zodat hij het kan opslaan in elke gewenste omgeving, ofwel hij kan een verwijzing (een *bookmark*) opslaan.

#### Downloaden

Het downloaden van een digitale manifestatie, zoals de scans of de transcriptie, is mogelijk door het weergeven van een link op de presentatiepagina. Daarnaast kan de gebruiker ook een link volgen naar een weergave van de metadata in bijvoorbeeld XML-formaat.

Het downloaden van een digitale manifestatie stelt een gebruiker in staat een bewerking te doen op de grondstof die relevant is voor zijn onderzoek met de tool die hij nodig heeft om de analyse uit te voeren. Door data in te lezen in een database, kan hij bijvoorbeeld bewerkingen uitvoeren op de data die hem de benodigde inzichten geven voor zijn onderzoek. Allerlei soorten tools kunnen op basis van de data allerlei soorten analyses uitvoeren. Denk aan statistische pakketten, Geografische Informatiesystemen, software voor Netwerk Analyses of Natural Language Processing.

#### Bookmarks

Natuurlijk is het niet altijd nodig dat de gebruiker de informatie downloadt om het te gebruiken. Een verwijzing is in veel situaties voldoende. Dit werkt alleen als het webadres dat de gebruiker opslaat niet meer wijzigt. Er moet sprake zijn van een zogenaamde persistent identifier (zie paragraaf 8.5). Het is namelijk onwenselijk dat gebruikers in de toekomst een verwijzing willen volgen naar materiaal dat niet meer op dat webadres beschikbaar is.

Daarnaast is het prettig als de url beknopt en leesbaar is. Een gebruiker kan dan – als dat nodig is – de url onthouden en overtypen. Dit is in tegenspraak met ideeën die we hebben besproken in paragraaf 8.5. Veel persistent identifiers hebben namelijk (bewust) een onleesbare opbouw. Een heldere keuze en een oplossing voor de consequenties is daarom belangrijk.



## Embedded metadata

In de presentatiepagina van het materiaal kan de ontwikkelaar van de website ook embedded metadata opnemen. De metadata is dan voor mensen onzichtbaar in de pagina vastgelegd, maar voor software beschikbaar om te kunnen gebruiken. Het is daarbij van belang dat de namen van de velden hierbij zijn meegegeven.

Dankzij deze embedded metadata kunnen bijvoorbeeld algemene zoekmachines, zoals Google, er voor zorgen dat het publiek gericht op het materiaal kan zoeken. Het vormt daardoor een aanvulling op de zoekmogelijkheid op basis van de volledige tekst die op de webpagina beschikbaar is.

Andere software die gebruik maakt van de embedded metadata, is Reference Management Software<sup>18</sup>. Met behulp van deze software kan een gebruiker zijn eigen verzameling van verwijzingen naar bronnen vastleggen en deze verwijzingen hergebruiken in het notenapparaat van zijn publicaties.

De embedded metadata moet gecodeerd zijn volgens een internationale norm: iedere toepassing moet de namen van de velden kunnen begrijpen. Er zijn veel verschillende standaarden beschikbaar om embedded metadata te realiseren:

- COinS
- RDFa
- Microdata en Microformats

## Opslaan van een zoekvraag

De toename en verbetering van de informatie in een aggregator houdt nooit op. Het publiek kan de behoefte hebben om een zoekvraag na een aantal weken of maanden opnieuw te stellen om te zien of er nieuwe grondstoffen zijn toegevoegd, die relevant zijn voor zijn zoekvraag. De ontwikkelaar van de website kan daarvoor een dienst aanbieden waarmee de gebruiker de zoekvraag opslaat en na een bepaalde periode opnieuw kan opvragen en uitvoeren.

## 10.4 Verrijken

Het publiek kan op verschillende wijzen betrokken zijn bij de collectie. In deel 1 is op verschillende plaatsen aangegeven hoe het publiek kan bijdragen aan het maken van de grondstoffen. Dit noemen we crowdsourcing. In hoofdstuk 6 ging het over de mogelijkheid een eigen bijdrage te leveren, die aan een bron is gerelateerd. Collectiebeheerders en beheerders van aggregatoren kunnen het publiek nog op een derde manier betrekken: door het publiek rechtstreeks opmerkingen bij een object te laten maken. Hiervoor bieden zij een functie aan op de website waarmee een gebruiker informatie over een bron kenbaar kan maken.

---

18 [https://en.wikipedia.org/wiki/Reference\\_management\\_software](https://en.wikipedia.org/wiki/Reference_management_software)

### **In de praktijk: Brabants Historisch Informatie Centrum (BHIC)**

Het BHIC is ervaren in het vormen van een gemeenschap van publiek. Het BHIC werkt intensief samen met die gemeenschap om de toegankelijkheid van het materiaal te verbeteren en kennis over het verleden van Brabant vast te leggen.

De website beschikt daarvoor over mogelijkheden voor het publiek om verhalen vast te leggen en discussies te voeren over het onderzoek dat wordt uitgevoerd.

Bij elke presentatiepagina (of “landingpage”) kan de ontwikkelaar het publiek de mogelijkheid bieden om een reactie kwijt te kunnen. De reactie kan allerlei vormen aannemen: soms is er sprake van verbetervoorstellen of aanvullingen voor de geboden informatie of een persoonlijke opmerking.

### **Verbetervoorstellen of aanvullingen**

Een collectiebeheerder moet de reacties in de gaten houden om vast te stellen of de reactie aanleiding geeft tot het wijzigen van de grondstoffen. Vervolgens kan hij, als dat wenselijk is, de reactie van de website verwijderen.

Technisch is het mogelijk om het publiek rechtstreeks te laten corrigeren in de grondstoffen zoals deze in het collectiebeheer- of het opslagsysteem zijn opgenomen. De website moet dan voorzien in deze mogelijkheid. Een collectiebeheerder moet op deze manier de controle over zijn grondstoffen uit handen willen geven.

### **Persoonlijke reacties**

Persoonlijke reacties geven vaak verhalen en informatie over collectiemateriaal die voornamelijk relevant zijn voor degene die reageert en zijn directe omgeving. De informatie is interessant om te laten zien en de discussie mogelijk te maken.

Het is belangrijk na te denken over de vraag of deze reacties moeten worden bewaard op de lange termijn en zo ja hoe dan? Ooit zal de website veranderen en het is dan de vraag of de collectiebeheerder deze persoonlijke reacties ook kan overnemen in een collectiebeheer- of opslagsysteem.

### **Tags**

Op de presentatiepagina kan de ontwikkelaar van de website de mogelijkheid bieden om een trefwoord (oftewel een “tag”) aan een presentatiepagina te koppelen. Andere gebruikers kunnen dan in de toekomst het materiaal met behulp van de gekozen term terugvinden. Meestal zullen gebruikers een term kiezen die hen persoonlijk in staat stelt om het materiaal met deze term op te roepen.

## **10.5 Gebruiksgegevens vangen**

In de interactie tussen gebruikers en de website laat het publiek allerlei sporen na. Zo kan de website bijvoorbeeld de gekozen zoekterm vastleggen en koppelen aan de collectie-onderdelen die daarna uitgebreid zijn bestudeerd. Blijkbaar zijn de bewuste zoekterm en het collectieonderdeel aan elkaar gerelateerd. Een ander voorbeeld is

dat een gebruiker verwijzingen van verschillende objecten bij elkaar opslaat, waardoor materiaal met elkaar wordt verbonden. Dit soort informatie kan een aggregator in zijn zoekmachine gebruiken om een gebruiker te helpen haar zoekvraag nog beter te beantwoorden.

We maken onderscheid tussen gebruiksinformatie die een gebruiker achterlaat tijdens zijn gedrag (we noemen deze “passief”) en gebruiksinformatie die hij bewust (“actief”) toevoegt door bijvoorbeeld materiaal of een zoekvraag op te slaan.

### **Passieve gebruiksinformatie vangen**

De passieve gebruiksinformatie laat een gebruiker achter tijdens zijn zoekgedrag. Objecten die een gebruiker in dezelfde sessie heeft aangeklikt en bestudeerd horen blijkbaar bij elkaar. Ook de reservering van een boek of archiefstuk is in dit opzicht relevante gebruiksinformatie. Dit maakt het mogelijk om een functie te maken vergelijkbaar met wat een webwinkel aanbiedt: “andere bezoekers bestudeerden ook ...”

De passieve gebruiksinformatie kan een website verzamelen zonder dat de gebruiker heeft moeten inloggen.

### **Actieve gebruiksinformatie vangen**

Een gebruiker kan op een website een functionaliteit gebruiken waarmee hij de objecten opslaat die hij relevant vindt. Het kan zijn dat hij daarvoor een profiel heeft aangemaakt op de website zelf, maar hij kan deze acties ook uitvoeren in een andere omgeving. Een belangrijk voorbeeld van dit soort platformen is Reference Management Software.

Dit soort systemen biedt de mogelijkheid om trefwoorden toe te kennen en objecten in mapjes bij elkaar te plaatsen. De structuren die daardoor ontstaan zijn allemaal extra aanwijzingen over de verbanden tussen het materiaal. Ook andere gebruikers kunnen in deze verbanden zijn geïnteresseerd.

## **10.6 Enkele juridische aspecten**

Alle juridische onzekerheden zijn als het goed is allemaal opgevangen in de data-laag en de verbindingslaag. Bij de presentatie van materiaal is het vervolgens van belang dat de ontwikkelaar de juridische mogelijkheden of beperkingen op een heldere manier laat zien. Hierbij kan hij bijvoorbeeld gebruik maken van de logo's die voor Creative Commons licenties zijn ontworpen<sup>19</sup>.

Als een website een gebruiker in staat stelt om materiaal te verrijken is het belangrijk om van tevoren met de gebruiker van de website af te spreken onder welke voorwaarden de verrijking mag worden gebruikt. Een website ontwikkelaar neemt alleen bij hoge uitzondering materiaal op waaraan een websitegebruiker voorwaarden stelt voor hergebruik.

---

<sup>19</sup> <https://creativecommons.org/about/downloads>

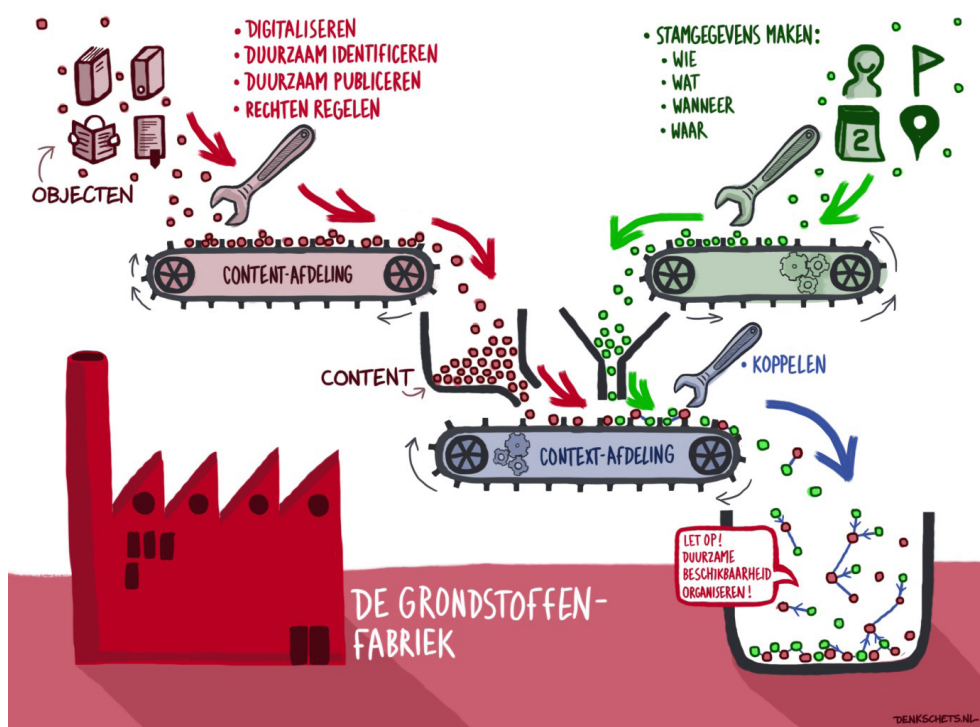
Bij het vastleggen van het gedrag van een gebruiker of bij het hergebruik van zijn actieve gebruikersinformatie moet de gebruiker nadrukkelijk toestemming hebben verleend.

## 11. Conclusie

Het zoeken, vinden en verrijken van bronnenmateriaal is voor een breed publiek van belang om een historische zoekvraag te beantwoorden. Verschillende partijen in een breed pallet van personen, instellingen en leveranciers dragen bij om dit mogelijk te maken.

Het eerste deel van dit rapport gaat in op het maken van de “grondstoffen” van zoeken, vinden en verrijken. In deel 2 omschrijven we welke technieken nodig zijn voor het online beschikbaar krijgen en houden van de gemaakte grondstoffen.

Van objecten die alleen fysiek bestaan, is het belangrijk om als grondstof allerlei digitale informatie (“metadata”) vast te leggen over het object en haar onderwerp. De metadata verwijst in veel velden naar informatie die op een centrale plaats wordt bijgehouden: stamgegevens. Het gaat daarbij bijvoorbeeld om informatie over auteurs of plaatsen. Naast de metadata en de stamgegevens, speelt een digitale manifestatie van een fysieke object (bijvoorbeeld een scan) een belangrijke rol. Het maakt het mogelijk dat het publiek een object op afstand en op elk moment van de dag kan bestuderen. De vierde soort grondstof die we in dit rapport onderscheiden wordt gemaakt door het publiek: “User Generated Content”. De grondstoffen vormen gezamenlijk een datanetwerk dat op allerlei manieren kan worden geëxploiteerd.



Figuur: De “grondstoffenfabriek” naar concept van Ivo Zandhuis, getekend door Elco van Staveren | Denkschets.nl. In de content-afdeling wordt digitale informatie over de collectie gemaakt door het vastleggen van data of het scannen van materiaal. In de context-afdeling worden de onderwerpen (plaatsen, personen, tijden en onderwerpen) aan de objecten gekoppeld. Rechtsonder de eindproducten; bronnen die verbonden zijn tot zinvolle semantische netwerken.

De exploitatie vindt plaats door een intensief samenwerken van instellingen en software. Dit is het onderwerp van deel 2. In een applicatielandschap werken de diverse soorten softwarecomponenten samen om de grondstoffen duurzaam aan de man te brengen. Het drie-lagen-model geeft vorm en structuur aan deze samenwerking. In elke laag van dit model zijn de applicaties gegroepeerd rondom een specifieke functionele behoefte. In de data-laag staan beheren en behouden centraal, een functionele behoefte van collectiebeheerders. In de verbindingslaag wordt informatie samengebracht en verbonden tot bruikbare eenheden. Deze eenheden kunnen in de presentatielaag in interactie met het publiek worden gepresenteerd.

Standaarden en normen spelen een belangrijke rol in dit rapport. Bij de creatie van grondstoffen moet een collectiebeheerder nagaan welke normen hij moet hanteren. Grondstoffen die aan de norm voldoen komen beter tot hun recht in de verbindings- en presentatielaag. De samenwerking tussen de applicaties in de verschillende lagen wordt bepaald door standaarden in uitwisseling van informatie. Op die manier kunnen applicaties en lagen onafhankelijk van elkaar evolueren.

Dit rapport geeft overzicht in de mogelijkheden en de normen die leiden tot verbetering van zoeken, vinden en verrijken. Bestudering leidt tot inzicht in de keuzes die een collectiebeheerder moet maken om zijn doel te bereiken. Collectieontsluiting is een meerjarenplan, waarin collectiebeheerders geleidelijk kennis en ervaring opdoen over de technieken en hoe zij deze het beste kunnen inzetten. Technologie ontwikkelt zich ondertussen op onvoorspelbare wijze. Standaarden, samenwerking in een infrastructuur, kennisdeling en -uitwisseling bereidt instellingen het beste voor op de toekomst.

# Bijlages

# Bijlage A: met dank aan ...

Aanspreekpunt bij het NIOD, opdrachtgever: Edwin Klijn

Werkgroep:

- Erik Groeneveld, SEECR
- Jesse de Vos, wikimedia/Beeld en Geluid
- Linda Reijnhoudt, DANS
- Lisette Kalshoven en Tessa Askamp, Kennisland
- Marian Hellema, KB
- Pieter Vijn, Digitale Collectie
- Robert Gillesse, DEN
- Tico Koopmans, Nationaal Archief
- Tim Veken, NIOD/KB
- Annelies van Nispen, EYE/NIOD
- Joop Vanderheijden, RCE

Collega's in het KCO-programma:

- Wilbert Helmus, DEN
- Marjo Bakker, NIOD
- Harco Gijsbers, NIOD
- Margreet Windhorst, Raamwerk Advies

Informatie ingewonnen bij:

- Ellen van der Waerden, Herinneringscentrum kamp Westerbork
- Lizzy Jongma, Rijksmuseum
- Ed de Heer, projectleider referentie-architectuur, DEN



## Bijlage B: relatie met DE BASIS

Veel van de informatie in dit rapport is ook te vinden in DE BASIS van DEN. DE BASIS staat voor Digitaal Erfgoed: Bouwen Aan Succesvolle ICT-Strategie. Deze is te vinden op [www.den.nl/debasis](http://www.den.nl/debasis).

DE BASIS is verdeeld in zes onderwerpen. De onderstaande tabel geeft een overzicht van de plaats in dit rapport waar hetzelfde onderwerp ter sprake komt. Vaak is de informatie in DE BASIS uitgebreider, bevat nog meer verwijzingen en bovendien wordt de informatie geactualiseerd. Het is een goede aanvulling op deze verkenning.

DE BASIS voor ...	url	Deze verkenning
auteursrechtenbeheer	<a href="http://www.den.nl/pagina/501/">http://www.den.nl/pagina/501/</a>	§3.4, §8.8, §9.5, §10.6
presentatie	<a href="http://www.den.nl/pagina/220/">http://www.den.nl/pagina/220/</a>	Hoofdstuk 10
vervaardiging	<a href="http://www.den.nl/pagina/218/">http://www.den.nl/pagina/218/</a>	Hoofdstuk 4
vindbaarheid	<a href="http://www.den.nl/pagina/219/">http://www.den.nl/pagina/219/</a>	§8.5, §8.7
beschrijving	<a href="http://www.den.nl/pagina/307/">http://www.den.nl/pagina/307/</a>	Hoofdstukken 3 en 5
duurzaamheid	<a href="http://www.den.nl/pagina/217/">http://www.den.nl/pagina/217/</a>	§8.4 en §8.5

De omgang met het publiek in de vorm van crowdsourcing of bij de creatie van User Generated Content (hoofdstuk 6, §8.6, §10.4) maakt (nog) geen deel uit van DE BASIS. Wel heeft DEN hierover een overzicht gepubliceerd. Deze is hier te vinden: <http://www.den.nl/thema/181/>

## Bijlage C: literatuur

Verwijzingen naar relevante websites zijn in de tekst op de betreffende plek opgenomen in een voetnoot. De meer formele literatuurverwijzingen staan hieronder.

- Aalten, van, J. 2011. 'Automatisch classificeren'. *Informatiewetenschap*. Retrieved (<http://www.invenier.nl/wp-content/uploads/2011/02/Automatisch-classificeren-Joyce-van-Aalten.pdf>).
- Anon. 2015. *Open Toegang - Archief2020*. Retrieved (<https://www.youtube.com/watch?v=BXHkKgOMeKc>).
- Beunen, Annemarie and Tjeerd Schiphof. 2006. *Juridische Wegwijzer Archieven en Musea online*. Den Haag: Taskforce Archieven/Museumvereniging. Retrieved ([http://www.nationaalarchief.nl/sites/default/files/docs/juridische\\_wegwijzer\\_archieven\\_en\\_musea\\_online\\_0\\_0.pdf](http://www.nationaalarchief.nl/sites/default/files/docs/juridische_wegwijzer_archieven_en_musea_online_0_0.pdf)).
- Foskett, A. C. 1969. *The Subject Approach to Information*. London: Clive Bingley.
- Gerrits, Boy. 2013. 'Archief in bewerking: een constante factor in de netwerkmaatschappij'. Masterthesis, Universiteit van Amsterdam, Amsterdam.
- Van der Heiden, Karin and Ivo Zandhuis. 2014. *Open toegang tot archiefcollectie Nederland*. Den Haag: Archief2020. Retrieved (<http://www.archief2020.nl/downloads/open-toegang-tot-archiefcollectie-nederland>).
- Helwig, Petra and Roland Bisscheroux. 2015. 'Landelijke Archieven Referentie Architectuur : Architectuur voor duurzame toegankelijkheid van informatie'. *Archievenblad* (1):16–19.
- ICA. 2000. *ISAD(G): General International Standard Archival Description*. Ottawa.
- ICA. 2004. *ISAAR (CPF), International Standard Archival Authority Record for Corporate Bodies, Persons and Families*.
- IFLA. 1998. *Functional Requirements for Bibliographic Records: Final Report*. Retrieved April 8, 2015 (<http://alltitles.ebrary.com/Doc?id=10831486>).
- Joint Steering Committee for Development of RDA et al. 2010. *Resource Description & Access: RDA*. Chicago: American Library Association.
- De Jong, Annemieke. 2007. 'Metadatamodel Beeld en Geluid biedt gebruiker van digitale content diepte en structuur'. *Informatie Professional: vakblad voor informatiewerkers* (6):24–29.
- Netwerk Digitaal Erfgoed. 2015. *Nationale strategie digitaal erfgoed*. [Den Haag]. Retrieved ([http://www.den.nl/art/uploads/files/Publicaties/Nationale\\_Strategie\\_Digitaal\\_Erfgoed\\_MinOCW.pdf](http://www.den.nl/art/uploads/files/Publicaties/Nationale_Strategie_Digitaal_Erfgoed_MinOCW.pdf)).
- Schneiders, P. 1984. *Bibliotheek en documentatie: handboek ten dienste van de opleidingen*. Deventer: Van Loghum Slaterus.