

NETWERK DIGITAAL ERFGOED

Een *knowledge graph* voor de integratie van erfgoedinformatie

Februari 2016



Commissioned by Network Digital Heritage

Authors:
Michiel Hildebrand
Wouter Alink

Inhoudsopgave

| | |
|---|-----------|
| Samenvatting | 3 |
| 1. Introductie | 4 |
| 1.1 Achtergrond..... | 4 |
| 1.2 Taak | 5 |
| 1.3 Aanpak..... | 5 |
| 1.4 Structuur..... | 6 |
| Versies | 6 |
| 2. Eisen aan het termennetwerk | 7 |
| 2.1 Huidige ontwikkelingen: integratie binnen de erfgoed domeinen | 7 |
| 2.2 Termennetwerk: gedistribueerd beheer, uniforme toegang..... | 8 |
| Metadata schema's linken | 8 |
| Gecontroleerde vocabulaires linken | 9 |
| 2.3 Toegang tot het termennetwerk voor annotatie | 10 |
| 2.4 Toegang tot collecties via het termennetwerk | 10 |
| 3. Het termennetwerk als knowledge graph | 12 |
| Rollen en taken..... | 13 |
| Collectiebeheerder | 14 |
| Domeinexpert..... | 14 |
| Knooppunt..... | 15 |
| Netwerk Digitaal Erfgoed | 15 |
| Het bepalen van rollen | 15 |
| Werken en realisaties..... | 15 |
| 4. Dienstenconcept | 17 |
| 4.1 Collectie-objecten beschrijven (annoteren)..... | 17 |
| 4.2 Aandragen kandidaat-termen | 17 |
| Domeinexpert..... | 18 |
| Beheer domeintermen | 18 |
| Linken van gecontroleerde vocabulaires..... | 18 |
| Knooppunt..... | 18 |
| Publiceren gecontroleerde vocabulaires..... | 18 |
| Beheer en integratie van schema..... | 19 |
| Registratie knooppunten..... | 19 |
| Netwerk Digitaal Erfgoed | 19 |
| Kwaliteitscontrole..... | 19 |
| Toegang tot de knowledge graph..... | 19 |
| 5. Tot slot | 21 |
| 5.1 Realisatie dienstenconcept | 21 |
| 5.2 Toegang tot collecties | 21 |

Samenvatting

De Nationale Strategie voor Digitaal Erfgoed draagt bij aan de grenzeloze toegang tot erfgoedcollecties. Hiervoor moeten de collecties betekenisvol met elkaar worden verbonden. De gecontroleerde vocabulaires die worden gebruikt door erfgoedinstellingen om hun collectie-objecten mee te beschrijven spelen hierin een belangrijke rol. In dit rapport geven wij advies over het verbinden van gecontroleerde vocabulaires tot een duurzame en bruikbare verbindingslaag. Wij stellen voor om de verbindingslaag te realiseren als een knowledge graph voor erfgoed informatie. Naar buiten toe biedt deze knowledge graph een uniform beeld van de personen, plaatsen, gebeurtenissen en concepten die voorkomen in de collecties. Naar binnen toe wordt de knowledge graph op een gedistribueerde manier gerealiseerd en beheerd. Erfgoedinstellingen dragen bij door het delen en onderling verbinden van hun gecontroleerde vocabulaires. Daarnaast zijn bronnen van buiten de erfgoedsector relevant.

Om een knowledge graph van erfgoed informatie op een duurzame manier te realiseren stellen we vier voorwaarden:

1. iedereen kan een knooppunt worden en een gecontroleerd vocabulaire registreren;
2. knooppunten garanderen dat een gecontroleerd vocabulaire openbaar beschikbaar is en blijft en dat de bron interoperabel is met andere bronnen door de schema's te integreren;
3. beheerders van een gecontroleerd vocabulaire zijn verantwoordelijk voor de integratie met andere vocabulaires in het netwerk door de termen te linken;
4. collectiebeheerders beschrijven hun collecties met termen uit het netwerk.

We beschrijven de vier rollen die de genoemde voorwaarden blijvend borgen: het Netwerk Digitaal Erfgoed, de knooppunten, de domeinexperts en de collectiebeheerders. Voor de taken die bij deze rollen horen beschrijven we de eisen aan de diensten die nodig zijn om deze taken uit te voeren.

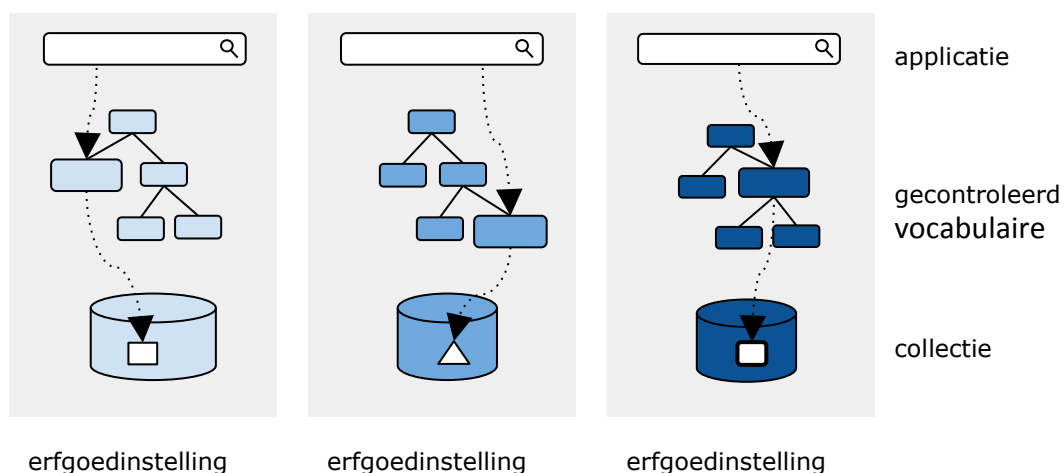
Met dit rapport kan de stuurgroep van het werkpakket Digitaal Erfgoed Bruikbaar een besluit voorbereiden voor de uitgangspunten, globale opzet en realisering van de verbindingslaag als onderdeel van de nationale infrastructuur voor digitaal erfgoed.

1 Introductie

1.1 Achtergrond

Met de Nationale Strategie voor Digitaal Erfgoed heeft het Netwerk Digitaal Erfgoed (NDE) de ambitie om de digitale voorzieningen en diensten voor erfgoedinstellingen te verbeteren. Deze voorzieningen en diensten moeten bijdragen aan een betere houdbaarheid, bruikbaarheid en zichtbaarheid van het Nederlandse erfgoed. Het initiatief beslaat meerdere erfgoed domeinen, die elk worden gerepresenteerd door een landelijk knooppunt. Dit advies heeft betrekking op *werkpakket bruikbaar*. Dit werkpakket heeft als doel de toegankelijkheid van collecties te vergroten en het gericht doorzoeken van de collecties te verbeteren, door alle erfgoed domeinen heen. Om dit doel te realiseren schetst de nationale strategie een verbindingslaag over de individuele collecties heen.

Figuur 1 schetst de traditionele situatie in de erfgoedsector. Elke instelling beheert een *collectie* en beschrijft de collectie-objecten met *termen* (ofwel concepten) uit haar eigen termenlijst, thesaurus of register. In dit rapport spreken we in zijn algemeenheid over een *gecontroleerd vocabulaire*. In sommige gevallen spreken we ook simpelweg over een *bron*, in de context van dit rapport bedoelen we dan een gecontroleerd vocabulaire en niet een collectie. In de huidige situatie bieden instellingen zelf toegang tot de collectie, met behulp van het eigen vocabulaire.



Figuur 1. Traditionele situatie in de erfgoedsector. Elke erfgoedinstelling beschrijft collectie-objecten op zijn eigen manier, met haar eigen metadataschema en met termen uit haar eigen, intern, vocabulaire. De erfgoedinstelling beheert zelf het gecontroleerde vocabulaire. Gebruikers kunnen zoeken in een collectie met behulp van de termen uit het vocabulaire. De collecties worden individueel ontsloten.

Wanneer deze collecties met de daarbij behorende vocabulaires worden samengevoegd (aggregatie), doet zich een probleem voor. Vocabulaires van de verschillende collecties kunnen vergelijkbare (of gerelateerde) concepten bevatten, maar deze overeenkomsten en relaties zijn nergens expliciet vastgelegd. Een systeem dat iets met de samengevoegde collecties wil doen, heeft dus geen kennis over welke objecten over bepaalde onderwerpen gaan en welke niet. Het doel van de verbindingslaag is om de relaties tussen termen in de vocabulaires expliciet te maken. De termen in de vocabulaires worden met elkaar verbonden tot een termennetwerk; overeenkomsten tussen objecten worden hierdoor bekend en indirect worden hiermee de collecties geïntegreerd.

Individueel beheer van een eigen vocabulaire heeft nog een ander nadeel. Het beheer van een vocabulaire kost instellingen veel tijd, terwijl dit doorgaans niet tot de kerntaken van een instelling behoort. Hierdoor is de dekking en rijkdom van het vocabulaire vaak beperkt tot wat noodzakelijk is voor de collectie. Daarnaast hebben verschillende publieke initiatieven inmiddels nieuwe bronnen opgezet, veelal actief beheerd, en met grote dekking en rijke inhoud. Voorbeelden zijn de gecontroleerde vocabulaires beheerd door de knooppunten. Ook bronnen die beheerd worden buiten de erfgoedsector kunnen relevant zijn, zoals Geonames, DBPedia en Wikidata. Binnen de verbindingslaag zouden al deze grote publieke bronnen een belangrijke rol moeten gaan spelen.

1.2 Taak

Spinque is door de stuurgroep van NDE werkpakket bruikbaar gevraagd om de doelstellingen in dit werkpakket te vertalen naar een bruikbaar dienstenconcept, met specifieke aandacht voor:

- de rol van gecontroleerde vocabulaires en de bijbehorende diensten die nodig zijn om deze effectief in te zetten voor de integratie en ontsluiting van erfgoed data,
- de beoogde technische componenten en in welke mate deze al beschikbaar zijn, dan wel in ontwikkeling zijn,
- op welke manier de erfgoedsector kan aansluiten bij de ontwikkelingen buiten het erfgoeddoelgebied op het gebied van linked data en kennisnetwerken, en
- een gebruikersperspectief waarmee de beoogde resultaten van het werkpakket effectief kunnen worden gecommuniceerd.

De focus van dit rapport ligt op de verbindingslaag door middel van een duurzaam en bruikbaar termennetwerk. Om de werking van de verbindingslaag te illustreren kunnen raakvlakken met bijvoorbeeld het beheren van collecties of het aggregeren van collectiedata worden beschreven. De wijze van beheren van collecties of aggregeren van collectiedata ligt echter buiten de scope van dit rapport. Aggregatie van collectiedata wordt behandeld in het position paper geschreven binnen werkpakket bruikbaar door Johan Oomen, Wilbert Helmus en Enno Meijers.

Met het advies in dit rapport kan de stuurgroep van het werkpakket Digitaal Erfgoed Bruikbaar een besluit voorbereiden voor de uitgangspunten, globale opzet en realisering van het termennetwerk met het bijbehorende dienstenconcept.

1.3 Aanpak

Dit advies is tot stand gekomen middels (bureau)onderzoek, interviews met de stuurgroepleden (of vertegenwoordigers van de stuurgroepleden) en met de voorzitter van de stuurgroep. De stuurgroepleden vertegenwoordigen de landelijke knooppunten van het Netwerk Digitaal Erfgoed. Wij hebben interviews afgenomen met:

- Dirk Houtgraaf en Joop VanderHeiden van de Rijksdienst Cultureel Erfgoed (RCE)
- Enno Meijers van de Koninklijke Bibliotheek (KB)
- Vincent Huis in 't Veld en Willem Melder van het Nederlands Instituut voor Beeld en Geluid (B&G)
- Pieter Koenders van het Nationaal Archief (NA)
- Annette Gaalman van Erfgoed Brabant (EB)

1.4 Structuur

Op basis van de interviews beschrijven we in sectie 2 observaties van de huidige ontwikkelingen, de wensen en uitdagingen. De observaties vertalen we naar eisen die worden gesteld aan het termen-netwerk. In sectie 3 vertalen we de eisen in combinatie met ons onderzoek tot een voorstel om te komen tot een *knowledge graph voor erfgoedinformatie*. We definiëren de rollen en taken voor het beheer en gebruik van deze knowledge graph, waarin we alle stakeholders betrekken: het Netwerk Digitaal Erfgoed, de knooppunten, domeinexperts en collectiebeheerders. In sectie 4 geven we een overzicht van de diensten die nodig zijn om de geïdentificeerde taken uit te voeren. In de afsluitende sectie bediscussiëren we wat de consequenties en mogelijkheden zijn van het termen-netwerk op aggregatie en toegang tot de collecties.

Versies

De eerste opzet van het advies is gepresenteerd en besproken bij de stuurgroepvergadering op 20 januari 2016. Het eerste concept is 8 februari beschikbaar gesteld aan de stuurgroep en aan externe adviseur Victor de Boer. Na deze feedback ronde is de eindversie beschikbaar gemaakt op 18 februari. We danken alle geïnterviewden voor hun medewerking en enthousiasme, en de stuurgroep voor hun feedback. We danken Victor de Boer (VU), Chi Shing Chang (Spinque) en Arjen de Vries (Spinque) voor hun discussies en bijdragen aan dit document.

2 Eisen aan het termennetwerk

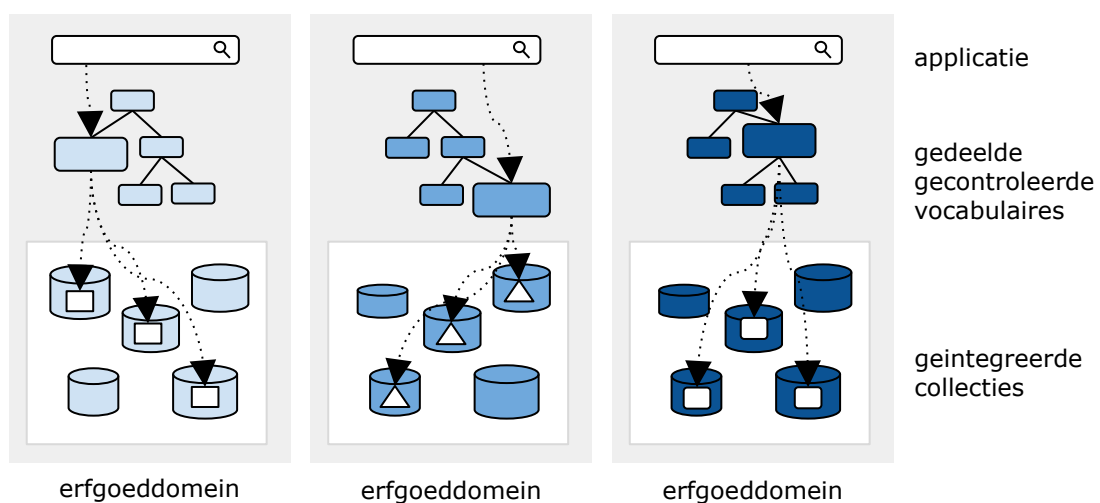
In de interviews heeft de focus gelegen op (i) de huidige ontwikkelingen op het gebied van integratie bij de knooppunten, (ii) de eisen die de knooppunten en de erfgoedsector in zijn algemeenheid stellen aan een termennetwerk, en (iii) de beoogde manieren om een termennetwerk toe te passen voor geïntegreerde toegang.

2.1 Huidige ontwikkelingen: integratie binnen de erfgoed domeinen

Binnen de erfgoed domeinen zijn er volop ontwikkelingen op het gebied van integratie. Op het niveau van de collecties wordt binnen verschillende erfgoed domeinen gewerkt aan een gedeelde infrastructuur. Binnen een gedeelde infrastructuur kunnen aangesloten instellingen gemakkelijk onderling informatie uitwisselen en toegang verschaffen tot informatie uit meerdere collecties. Voorbeelden zijn de Nationale Bibliotheek Catalogus (NBC), de gedeelde infrastructuur voor bibliotheken en de Archieven Portaal Europa (APE), de centrale aggregatie van archiefmateriaal.

Wat betreft gecontroleerde vocabulaires worden er binnen verschillende erfgoed domeinen al grote bronnen hergebruikt door de aangesloten instellingen. Voorbeelden zijn de Gemeenschappelijke Thesaurus Audiovisuele Archieven (GTAA) die gebruikt wordt door de publieke omroepen, de Erfgoed Thesaurus die wordt gebruikt voor onroerend erfgoed, en AAT-NL en RKD Artist die worden gebruikt door instellingen in de kunsten en architectuur. Andere voorbeelden zijn de thesauri binnen het Gemeenschappelijk Geautomatiseerd Catalogiseersysteemcentrale (GGC) voor bibliotheken, en de registratie van actoren bij het Nationaal Archief.

Figuur 2 geeft een schematische weergave van de integratie binnen de erfgoed domeinen. Ten opzichte van de oorspronkelijke situatie in Figuur 1 maken de instellingen binnen een erfgoed domein gebruik van dezelfde vocabulaires. Via het gedeelde vocabulaire kunnen objecten worden gevonden uit verschillende collecties.



Figuur 2. Integratie binnen erfgoed domeinen. In de huidige ontwikkelingen worden collecties centraal georganiseerd binnen een erfgoed domein. Binnen een erfgoed domein worden dezelfde gecontroleerde vocabulaires gebruikt. Gebruikers kunnen zoeken over meerdere collecties binnen één erfgoed domein.

Het doel van de nationale strategie is om, naast de integratie binnen ieder erfgoeddomein, (*intra*-domein) ook de integratie tussen de erfgoeddomeinen (*inter*-domein) te realiseren. Hiermee kunnen collecties op nationaal niveau als één geheel aangeboden worden, zodat bijvoorbeeld alle werken die zijn gemaakt door één en dezelfde persoon gevonden kunnen worden.

Het simpelweg samenvoegen van collecties is echter niet voldoende om uniforme toegang te realiseren. Omdat erfgoedinstellingen op verschillende manieren hun collectie-objecten beschrijven, kunnen we onze zoekvraag niet eenduidig mappen op de onderliggende data uit meerdere collecties. We onderscheiden twee soorten verschillen: (i) het *metadata schema* (bijv. “maker” vs “auteur”) en (ii) de *termen in de metadata velden* (bijv. “Johannes Vermeer” vs. “J. Vermeer”). Om alle werken van één en dezelfde persoon te vinden in de meerdere collecties moet er kennis zijn van alle beschikbare metadata-schema’s, alsmede alle verschillende representaties van die persoon in het desbetreffende veld.

Zowel binnen een erfgoeddomein als tussen erfgoeddomeinen treffen we diversiteit aan in de metadata-schema’s en de termenlijsten van verschillende instellingen. Binnen hetzelfde erfgoeddomein kunnen deze verschillen echter vaak worden opgelost door gebruik te maken van één gezamenlijk metadata schema en één (of enkele) gecontroleerde vocabulaires. Dit is ook precies de aard van de huidige ontwikkelingen binnen de erfgoeddomeinen, convergentie naar één homogeen systeem. Echter, over de erfgoeddomeinen heen is deze oplossing lastiger te realiseren.

De heterogeniteit tussen de erfgoeddomeinen is vele malen groter en bovendien verschilt de dagelijkse praktijk waarin de collectie gebruikt wordt. Voor de integratie van collecties uit verschillende erfgoeddomeinen is een oplossing nodig waarin de verschillende wereldbeelden en bijbehorende dagelijkse manier van werken naast elkaar kunnen blijven bestaan. Bovendien ontstaat dan de mogelijkheid om de rijke kennisbronnen van buiten de erfgoedsector eveneens te integreren.

- Wij zien de op natuurlijke wijze ontstane integratie van collecties en infrastructuur binnen de erfgoeddomeinen als een positieve ontwikkeling en verwachten dat zij bijdraagt aan de integratie op nationaal niveau. Het is echter wel van belang om te realiseren dat de ontwikkelingen op nationaal niveau ook impact zullen hebben op de werkwijze binnen de erfgoeddomeinen. Een duurzame oplossing vergt de toewijding van de knooppunten, maar even belangrijk is de bereidheid van de instellingen om vanuit een geïntegreerde gedachte te gaan werken.

2.2 Termennetwerk: gedistribueerd beheer, uniforme toegang

Het alternatief voor integratie via één homogeen systeem is een flexibele vorm van integratie waarin voortdurend en waar van toepassing, overeenkomsten en verbanden worden vastgelegd. In dit advies zetten we uiteen hoe deze aanpak kan worden toegepast op de integratie van erfgoedinformatie. Geïnspireerd door initiatieven op het gebied van *Linked Data* noemen we het leggen van deze verbanden ook wel *linken*. Integratie door linken kan worden toegepast op zowel metadata-schema’s als op gecontroleerde vocabulaires zoals termenlijsten en thesauri.

Metadataschema’s linken

In het geval van metadata-schema’s kan er een link worden gemaakt tussen domeinspecifieke schema’s en een overkoepelend schema (Dublin Core, Europeana Data Model, schema.org). Ook is het

mogelijk om direct een link te leggen tussen schema's. Binnen de Digitale Collectie Nederland¹ en binnen Europeana is veel aandacht voor de integratie van metadata-schema's. In Europeana worden bijvoorbeeld de schema's van de individuele collecties gemapped naar het Europeana Data Model². In dit advies gaan we niet verder in op de integratie van metadata-schema's voor collecties maar focussen we ons op het integreren van gecontroleerde vocabulaires. Zoals we later zullen zien speelt schema-integratie echter ook bij de integratie van vocabulaires een rol.

Gecontroleerde vocabulaires linken

In het geval van gecontroleerde vocabulaires gaat het linken om het leggen van verbanden tussen termen uit verschillende bronnen. Door links tussen deze termen ontstaat een termennetwerk dat de bronnen overspant. Het beheer van elk individueel vocabulaire blijft de verantwoordelijkheid van de oorspronkelijke aanbieder. Voor het realiseren en beheren van het termennetwerk is dus sprake van een decentraal, gedistribueerd initiatief. Naar buiten toe moet het termennetwerk juist zorgen voor uniforme toegang tot de collecties. Het linken van de vocabulaires speelt een belangrijke rol om dit technisch te realiseren. Minstens zo belangrijk is dat het termennetwerk naar buiten toe wordt gecommuniceerd als een geheel. Voor de professionals, onderzoekers en het publiek die het termennetwerk gebruiken is het gedistribueerde karakter van het beheer in eerste instantie niet relevant. Zij zoeken veelal naar de wie, wat, waar en wanneer in het netwerk van erfgoedinformatie en zelden naar een specifiek gecontroleerd vocabulaire.

De belangrijkste eisen om een bruikbaar termennetwerk te realiseren en beheren zijn als volgt:

- Erfgoedinstellingen gebruiken gedeelde publieke bronnen voor het beschrijven van hun collectieobjecten. Instellingen moeten zich de vraag stellen, in hoeverre hun interne gecontroleerde vocabulaires relevant zijn voor opname in het publieke domein. Interne vocabulaires verdwijnen als ze geen toegevoegde waarde hebben ten opzichte van de publieke bronnen. Als de relevantie evident is, maakt de instelling van de interne vocabulaire een nieuwe publieke bron, of maakt zij deze onderdeel van een al bestaande publieke bron.
- Iedereen kan bijdragen aan het termennetwerk door een gecontroleerd vocabulaire aan te bieden als een publiek gedeelde bron. De aanbieder is verantwoordelijk voor de kwaliteit van het vocabulaire, de duurzame beschikbaarheid en de integratie met andere bronnen. Bij het Netwerk Digitaal Erfgoed zijn vijf knooppunten betrokken die logischerwijs publieke bronnen kunnen aanbieden. Andere erfgoedinstellingen kunnen eveneens hun gecontroleerde vocabulaires aanbieden binnen het termennetwerk door ze publiek beschikbaar te maken en deze te integreren met de andere, gerelateerde bronnen in het netwerk.
- Verenigingen, stichtingen en particulieren die niet zelf een rol als aanbieder van een gecontroleerd vocabulaire ambiëren, kunnen hun specialistische kennis inbrengen onder het beheer van een bestaande aanbieder, of via een publieke omgeving als Wikidata. De domeinexpert blijft bijdragen aan het inhoudelijk beheer van termen, maar dan onder verantwoordelijkheid van de nieuwe aanbieder.
- De publieke bronnen in het termennetwerk worden onderling met elkaar gelinkt. De beheerder van de bron is verantwoordelijk voor het maken van links naar andere bronnen. Waar mogelijk worden de bronnen in het termennetwerk ook gelinkt met bronnen buiten de erfgoedsector.

¹ <http://digitalecollectie.nl/>

² <http://pro.europeana.eu/page/edm-documentation>

- Het termennetwerk bevat de informatie die nodig is om de wie, wat, waar en wanneer van collectie-objecten te beschrijven. Simple Knowledge Organization System (SKOS) wordt gebruikt voor het representeren van concepten (wat). Andere vocabulaires zijn nodig om de aanvullende eigenschappen van bijvoorbeeld personen (wie), locaties (waar) en gebeurtenissen (wanneer) te representeren.

2.3 Toegang tot het termennetwerk voor annotatie

In de visie van de knooppunten gaan de interne niet-publieke vocabulaires uiteindelijk op in open, publieke vocabulaires. De collectiebeheerder beschrijft dan haar objecten met concepten uit het termennetwerk. De eisen om dit annotatieproces te ondersteunen zijn als volgt:

- Een gebruiker kan zoeken in het termennetwerk over meerdere bronnen heen. Termen die gelinkt zijn worden als één resultaat getoond. De gebruiker heeft wel toegang tot informatie over de bronnen waarin het resultaat voorkomt.
- Een collectiebeheerder kan direct in haar collectiebeheersysteem objecten annoteren met concepten uit het termennetwerk. De bronnen waarin de collectiebeheerder wil zoeken, kunnen verschillen per metadataveld. Ook de manier waarop de resultaten worden getoond en gerangschikt, kunnen verschillend per metadataveld.
- Instellingen zijn niet beperkt tot de vocabulaires die via het knooppunt van hun sector worden aangeboden. Gecontroleerde vocabulaires kunnen informatie bevatten die zowel binnen als buiten het expertisegebied van het knooppunt valt. Een voorbeeld is de geografische tak van veel Nederlandse culturele thesauri. De informatie over locaties is vaak beperkter dan in specialistische geografische bronnen, zoals de Historische Geothesaurus³ of Geonames. Instellingen binnen een erfgoeddomein kunnen ook gebruik maken van deze externe bronnen.
- Een collectiebeheerder moet nieuwe kandidaat-termen kunnen aandragen, bijvoorbeeld als blijkt dat bij het annoteren een term nog niet bestaat. De beheerder van het vocabulaire kan de kandidaat-term vervolgens accepteren, weigeren of aanpassen.

2.4 Toegang tot collecties via het termennetwerk

Een belangrijk speerpunt van de nationale strategie is het verbeteren van de toegang tot collecties. Deze verbeterde toegang begint bij de integratie van de data: op het niveau van de metadataschema's door de collecties op een interoperabele manier aan te bieden en op het niveau van de collectiebeschrijvingen door deze via het termennetwerk te integreren. Als aan deze voorwaarden is voldaan, wordt aggregatie en het gebruik van de geïntegreerde data bereikbaar voor iedereen. Het canonieke voorbeeld dat steeds terugkeert, is de aggregatie van collecties tot een nationale collectie, die vervolgens via een portal integraal toegankelijk kan worden gemaakt.

Verrijken met externe informatie biedt een erfgoedinstelling ook de mogelijkheid om de toegang tot de eigen collectie te verbeteren. In onze ogen is dit minstens zo belangrijk. Een betere eigen online dienstverlening wordt dan mogelijk. Instellingen kunnen toegang tot hun collectie verbeteren door gebruik te maken van rijkere informatie uit het termennetwerk en door informatie uit andere collec-

³ <http://erfgoedenlocatie.nl/>

ties. Zo maken al meerdere erfgoedinstellingen in Nederland en België gebruik van de AAT: Openlucht Museum, Zuiderzeemuseum, Belasting en Douane Museum, Bijzondere Collecties Universiteit Leiden, Antwerpse Musea, Musea Oost-Vlaanderen en Erfgoedplus.be. Omdat de AAT meertalig is, dit in tegenstelling tot de meeste interne bronnen, wordt op eenvoudige wijze meertalig zoeken mogelijk. Een ander voorbeeld is het aanbieden van objecten uit een andere collectie⁴, of het opzetten van een website voor een expositie met objecten uit meerdere collecties. Ook de rijke informatie uit externe bronnen, zoals Wikipedia, kan via het termennetwerk worden gebruikt om een rijkere presentatie te geven; bijvoorbeeld met achtergrondinformatie over een persoon. De integratie van collecties verlaagt daarnaast de drempels voor onderzoekers en ontwikkelaars om analyses over grote culturele datasets uit te voeren, die nu alleen bereikbaar zijn voor grote labs, zoals het Software Studies Initiative van Lev Manovich.⁵

De belangrijkste eisen aan het gebruik van het termennetwerk om collecties toegankelijk te maken zijn als volgt:

Een collectiebeheerder kan (via het collectiebeheersysteem) eenvoudig haar collectie in een interoperabel formaat beschikbaar maken. Hierin worden de referenties naar de termen uit het netwerk meegenomen.

In een systeem waarin meerdere collecties zijn geaggregeerd wordt het termennetwerk gebruikt om over de verschillende collecties te zoeken en te navigeren. Hiervoor moet het systeem om kunnen gaan met zowel de links tussen de metadataschema's en de links tussen de termen.

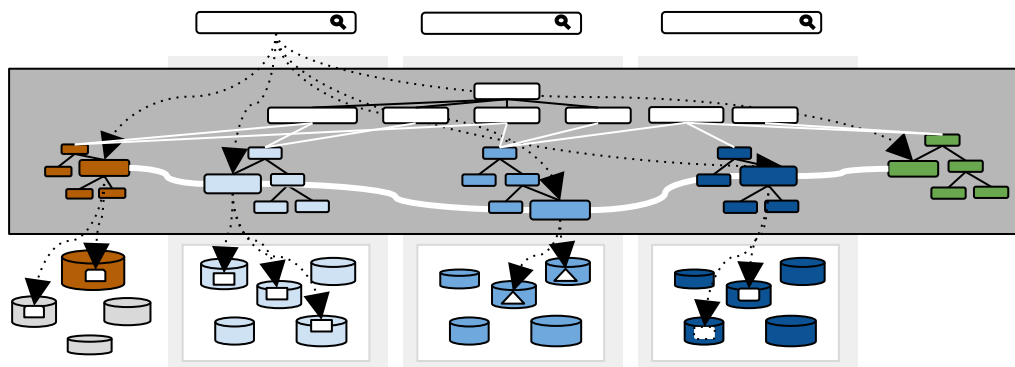
Een instelling kan ook in zijn eigen zoekstelsel gebruik maken van de rijke informatie uit het termennetwerk. Een instelling kan ook andere collecties die verbonden zijn met het netwerk gebruiken voor nieuwe vormen van toegang. In dit geval moet dus ook het lokale zoekstelsel om kunnen gaan met de links tussen de metadataschema's en de links tussen de termen.

⁴ <https://www.comsode.eu/index.php/2015/07/linked-open-images/>

⁵ <http://manovich.net/index.php/projects/cultural-analytics-social-computing>

3 Het termennetwerk als knowledge graph

In deze sectie vertalen we de eisen naar een visie voor een duurzaam en bruikbaar termennetwerk en de concrete rollen en taken, die nodig zijn om deze visie te realiseren.



Figuur 3. Visie van een knowledge graph voor erfgoedinformatie. Gecontroleerde vocabulaires binnen de erfgoeddomeinen worden met elkaar verbonden en met externe bronnen. Collectiebeheerders kunnen gebruik maken van meerdere bronnen voor het beschrijven van hun collectie-objecten.

De visie die wij willen uitdragen is dat het termennetwerk wordt benaderd als *een knowledge graph voor erfgoedinformatie*. Hierin refereert *knowledge graph* naar het initiatief van Google waarin zij de kennis in de wereld op een gestructureerde manier verzamelt⁶. Google gebruikt deze knowledge graph, onder andere, in hun zoekmachine. De knowledge graph is opgebouwd uit informatie afgeleid uit verschillende bronnen, waaronder Wikipedia. Van buiten is de knowledge graph toegankelijk als één geheel, voor mensen via de zoekmachine en voor machines via de API⁷. De eenduidigheid die Google kan realiseren in haar knowledge graph komt voort uit de centrale organisatie van de data en de mogelijkheid die Google heeft om zelf, zonder inmenging van anderen, één geïntegreerd geheel te construeren uit de verschillende onderliggende bronnen. Op schaniveau wordt de informatie gestructureerd aan de hand van schema.org⁸.

Op een vergelijkbare manier kan een knowledge graph voor erfgoedinformatie worden gerealiseerd. In de erfgoedsector ontbreekt echter de centrale instantie, die alle kennis uit de sector kan integreren. Wij stellen dan ook voor om de knowledge graph voor erfgoedinformatie op een gedistribueerde, decentrale manier te realiseren en beheren. Figuur 3 toont een schematische weergave van deze visie, in relatie tot de bestaande erfgoeddomeinen. Hierin spelen de gecontroleerde vocabulaires uit de erfgoedsector zelf een belangrijke rol en waar relevant ook externe bronnen als Wikidata (en/of Wikipedia). De bronnen worden geïntegreerd tot één geheel door de termen (horizontaal) met elkaar te linken. De verschillende type concepten worden geïntegreerd door de schema's van de vocabulaires te integreren (verticaal).

⁶ <https://www.google.com/intl/bn/insideseach/features/search/knowledge.html>

⁷ <https://developers.google.com/knowledge-graph/?hl=en>

⁸ <http://schema.org/docs/full.html>

Wij adviseren het gebruik van de schema's zo veel mogelijk open te houden. Wel is het raadzaam om te adviseren over de schema's die reeds beschikbaar zijn en hoe deze het best kunnen worden gebruikt. Daarnaast kunnen bestaande links tussen schema's beschikbaar worden gemaakt.

Het gevaar van een gedistribueerde aanpak is dat deze kan leiden tot wildgroei, waardoor het termen-netwerk niet meer als één geheel te benaderen is. Enkele eenvoudige voorwaarden kunnen deze wildgroei beperken, zodat een bruikbaar geïntegreerd netwerk kan ontstaan zonder centrale autoriteit.

De voorwaarden voor het decentraal organiseren van de knowledge graph voor erfgoed-informatie zien wij als volgt:

- Iedereen kan een knooppunt worden en een gecontroleerd vocabulaire registreren. Als startpunt kan het netwerk beginnen met de gecontroleerde vocabulaires van de knooppunten uit de nationale strategie.
- Knooppunten garanderen dat een gecontroleerd vocabulaire openbaar beschikbaar is en blijft, en dat de bron interoperabel is met andere bronnen door de schema's te integreren.
- Beheerders van een gecontroleerd vocabulaire zijn verantwoordelijk voor de integratie met andere vocabulaires in het netwerk door de termen te linken.
- Collectiebeheerders beschrijven hun collecties met termen uit het netwerk. Hiermee zijn ze niet meer beperkt tot de bronnen die binnen hun erfgoed-domein beschikbaar zijn, maar kunnen ze informatie gebruiken uit het netwerk als geheel.

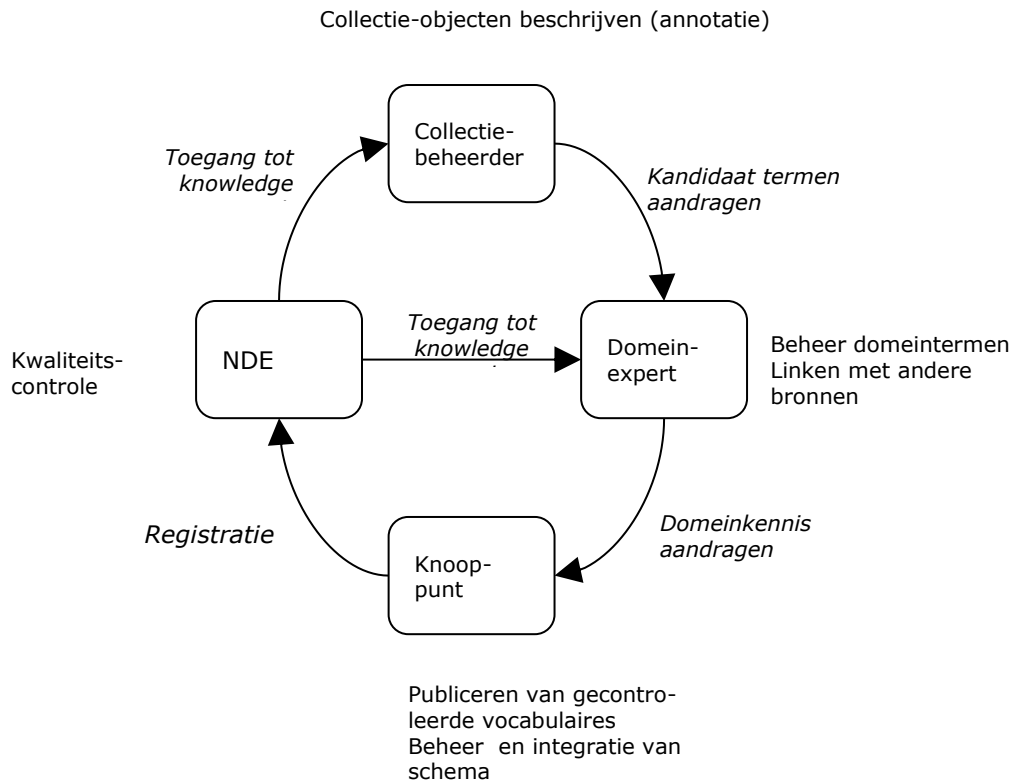
Omdat iedereen gebruik kan maken van en kan linken naar de bronnen die zij het meest geschikt achten, verwachten wij dat, op termijn, het belang van bronnen (of delen van een bron) zal uitkristalliseren. Als gevolg hiervan zouden beheerders zich moeten focussen op de informatie binnen hun expertisegebied en andere aspecten overlaten aan de andere experts binnen het netwerk.

Naast de reeds bestaande knooppunten binnen het Netwerk Digitaal Erfgoed zullen verschillende erfgoedinstellingen een rol als knooppunt op zich moeten nemen. Het advies van Vincent de Keizer geeft een overzicht van de gecontroleerde vocabulaires in de erfgoedsector en daarmee de kandidaten voor knooppunten. Het ligt voor de hand dat grote instituten een belangrijke rol zullen spelen. Het RKD beheert bijvoorbeeld meerdere bronnen in het kunstdomein, zoals de *RKDartists&*, *RKDimages*, de Nederlandse AAT en Iconclass. Naturalis beheert het nationale soortenregister. Kleinere instituten, met bijzondere domeinkennis, kunnen ervoor kiezen zelf een knooppunt te worden of hun kennis onder te brengen bij een bestaand knooppunt. Daarnaast kunnen thematische initiatieven, zoals het project visuele thesaurus voor online mode-erfgoed en het netwerk oorlogsbronnen⁹, nieuwe domeinkennis inbrengen bij een bestaand knooppunt of leiden tot nieuwe knooppunten. Voor tijdelijke projecten is het essentieel dat de duurzaamheid wordt gegarandeerd en moet de eindverantwoordelijkheid al binnen het project geregeld worden.

Rollen en taken

Om een knowledge graph in de erfgoedsector op een duurzame en bruikbare manier te realiseren identificeren we vier rollen. Figuur 4 toont de vier rollen, hun taken en de onderlinge relaties.

⁹ <http://www.oorlogsbronnen.nl/>



Figuur 4 Rollen in het onderhouden van een duurzame knowledge graph voor erfgoedinformatie. NDE maakt de knowledge graph toegankelijk voor collectiebeheerders en domeinexperts. Collectiebeheerders gebruiken de knowledge graph om hun collectie-objecten mee te beschrijven en dragen nieuwe kandidaat termen als deze nog niet beschikbaar zijn. Domeinexperts beheren termen en linken deze met de andere termen in de knowledge graph. Het knooppunt publiceert de gecontroleerde vocabulaires en registreert zichzelf bij het netwerk.

Collectiebeheerder

Collectiebeheerders integreren hun collectie op landelijk niveau door hun collecties te beschrijven met concepten uit de knowledge graph (annotatie). Als tijdens het annoteren blijkt dat er concepten nodig zijn die nog niet beschikbaar zijn in de knowledge graph, legt de collectiebeheerder deze voor aan een domeinexpert.

Domeinexpert

Domeinexperts zijn erfgoedinstellingen, verenigingen, stichtingen of individuen die specialistische kennis hebben van een domein. Deze kennis kan gebruikt worden om een bestaande bron van een knooppunt uit te breiden. De verantwoordelijkheid voor de publicatie van de bron ligt bij het knooppunt. De domeinexpert en het knooppunt coördineren samen het gedistribueerde beheer. De domeinexpert integreert zijn kennis met andere bronnen door concepten te linken. Het is ook mogelijk dat een domeinexpert zijn kennis bijdraagt via een open platform zoals Wikidata. Als een domeinexpert liever het beheer van hun kennis in eigen hand wil houden, wordt zij zelf een knooppunt, met de daarbij behorende taken.

Knooppunt

Een knooppunt is verantwoordelijk voor het publiceren van gecontroleerde vocabulaires. Het knooppunt beheert het schema van een gecontroleerd vocabulaire en zorgt voor het linken hiervan met de andere schema's. Voor de inhoudelijke aspecten heeft het knooppunt een coördinerende rol met de domeinspecialisten. Het knooppunt registreert zichzelf en zijn bronnen bij het netwerk.

Netwerk Digitaal Erfgoed

Het Netwerk Digitaal Erfgoed maakt het mogelijk dat de knooppunten zich kunnen registreren in de knowledge graph. Knooppunten zijn zelf verantwoordelijk voor het beschikbaar stellen van de gecontroleerde vocabulaires en de integratie. Het is de verantwoordelijkheid van het netwerk om het geheel te monitoren. Dit betekent dat het netwerk moet nagaan of de bronnen, die een knooppunt beschikbaar stelt, voldoen aan de kwaliteitseisen. Een onderdeel van de kwaliteitscontrole is het maken van een overzicht van de beschikbare bronnen en statistieken over de content, kwaliteit en gebruik. Het netwerk heeft ook een adviserende rol voor de knooppunten. Hoe maak je een gecontroleerd vocabulaire op een interoperabele manier beschikbaar? Welke schema's zijn er beschikbaar en welke zijn relevant? Welke andere bronnen zijn er om mee te linken en hoe doe ik dat?

Het bepalen van rollen

In het netwerk van digitaal erfgoed kan een collectiebeheerder tevens de rol van domeinexpert én van knooppunt vervullen. Hier is sprake van zodra een erfgoedinstelling besluit om een interne gecontroleerde vocabulaire publiek beschikbaar te maken in het termennetwerk. Om te bepalen welke van de bovenstaande rollen van toepassing is, maakt de erfgoedinstelling een inventarisatie van het interne vocabulaire ten opzichte van de knowledge graph. De inventarisatie kan leiden tot drie gevallen:

1. De concepten in het interne vocabulaire zijn al beschikbaar in de knowledge graph en ze voegen geen meerwaarde toe. Het interne vocabulaire wordt opgegeven en vervangen door links naar de knowledge graph.
2. De interne bron definieert een subdomein van een bestaande bron, dat nog niet of beperkt is uitgewerkt. De erfgoedinstelling wordt dan domeinexpert binnen een bestaand knooppunt. Erfgoedinstelling en knooppunt bepalen samen hoe de bronnen worden geïntegreerd.
3. Er is geen overlap met de knowledge graph, of er is overlap met de knowledge graph maar de bron geeft een ander (aanvullend) beeld op de knowledge graph. De collectiebeheerder maakt de interne bron publiek beschikbaar, integreert deze met het termennetwerk en neemt het beheer op zich. In dit geval vervult de erfgoedinstelling dus ook de rollen van domeinexpert en knooppunt.

Erfgoedinstellingen kunnen de inventarisatie van interne vocabulaires ook collectief uitvoeren. Een mooi voorbeeld hiervan vormt het project visuele thesaurus voor online mode-erfgoed. In dit project worden de interne vocabulaires eerst met elkaar vergeleken, om vervolgens de gezamenlijke overlap met de AAT te bepalen.

Werken en realisaties

Een interessant aspect van het opzetten van een knowledge graph voor erfgoedinformatie is dat het a-priori niet zonder meer duidelijk is waar de grenzen liggen van wat relevante erfgoedinformatie is. We kunnen ons bijvoorbeeld afvragen of collectie-objecten ook onderdeel moeten zijn van de knowledge

graph? Dit zal afhangen van het soort informatie dat nut heeft voor de erfgoedsector. Voor publicaties kan het bijvoorbeeld nuttig zijn om onderscheid te maken tussen “het werk” en de “realisatie” van een werk. Als een gebruiker van een bibliotheek op zoek is naar een boek, dan is zij meestal niet geïnteresseerd in een specifieke uitgave van dat boek. In eerste instantie wil de gebruiker zoeken naar “het werk”. Als zij vervolgens het werk wil lenen, is de specifieke realisatie van het werk wel van belang. Is er een realisatie beschikbaar op deze locatie? Een gebruiker moet dus kunnen zoeken op “het werk” en deze moet gelinkt zijn aan de realisaties. Om dit mogelijk te maken is het nuttig om een bron met alle werken in de knowledge graph beschikbaar te maken. In andere domeinen zien we vergelijkbare voorbeelden, zoals de “programma reeks” voor het jaarnaal, en “het werk” de denker van Rodin. Wikipedia bevat bijvoorbeeld precies deze interpretatie https://nl.wikipedia.org/wiki/NOS_Jaarnaal en https://nl.wikipedia.org/wiki/De_Denker.

4 Dienstenconcept

In de vorige sectie hebben we de taken geïdentificeerd die nodig zijn voor het beheer en gebruik van de knowledge graph. In deze sectie beschrijven we de diensten die nodig zijn om deze taken uit te voeren. We behandelen alleen de taken met een technische component en niet de advies- en coördinatietaken.

4.1 Collectie-objecten beschrijven (annoteren)

In collectiebeheersystemen zijn annotatievelden normaliter gekoppeld aan een interne bron. Om gebruik te maken van een externe knowledge graph met erfgoed informatie kan een collectiebeheersysteem op twee manieren te werk gaan:

- (i) maak gebruik van een externe API om te zoeken naar termen, of
- (ii) aggregeer de relevante bronnen en maak deze binnen het systeem beschikbaar.

Binnen de erfgoedsector zien we bij verschillende systemen ontwikkelingen voor beide aanpakken.

Naast de technische integratie van een knowledge graph in een collectiebeheersysteem, zullen er ook nieuwe aspecten aan het licht komen bij het zoeken naar termen. Zo zal het per erfgoedinstelling verschillen welke bronnen relevant zijn. Ook het rangschikken van resultaten en de presentatie van concepten kunnen specifieke eisen hebben, die verschillen per erfgoedinstelling en annotatieveld. Het onderzoek van Hildebrand en Ossenbruggen naar annotatie met meerdere thesauri bij het Rijksmuseum¹⁰, geeft inzicht in de eisen en mogelijke oplossingen. Om een efficiënte set-up te realiseren voor erfgoedinstellingen, zullen de collectiebeheersystemen het zoeken naar en de presentatie van resultaten beter moeten kunnen aanpassen aan de wensen van verschillende erfgoedinstellingen. Bovendien moet het mogelijk zijn voor collectiebeheerders om het zoekstelsel voor het termen-netwerk te configureren naar hun wensen.

Het opbouwen en in gebruik nemen van de knowledge graph door de instellingen zal inspanning en tijd vergen. Veranderingen in de workflow en het gebruik van het collectiebeheersysteem zullen nodig zijn. Hierdoor is het niet direct mogelijk om de knowledge graph, vanaf het eerste moment van beschikbaarheid, gelijk volledig te benutten. Het is dan ook aannemelijk, dat het vervangen van interne vocabulaires geleidelijk aan zal plaatsvinden. Tot die tijd kan een instelling de integratie van haar collectie wel al bevorderen, door de interne vocabulaires te linken naar de knowledge graph.

4.2 Aandragen kandidaat-termen

Als tijdens het annoteren blijkt dat een concept nog niet beschikbaar is in de knowledge graph, dan moet een collectiebeheerder deze als kandidaat-term kunnen aandragen. Voor het beheersysteem van een vocabulaire betekent dit dat externe gebruikers een kandidaat-term moeten kunnen aanmaken. De beheerder van het vocabulaire moet deze kandidaat vervolgens kunnen beoordelen. Het vocabulaire beheersysteem moet de aandrager van de kandidaat-term informeren over de beslissing van de beheerder, bijvoorbeeld door een notificatie te versturen. Voor collectiebeheersystemen

¹⁰ <http://oai.cwi.nl/oai/asset/13989/13989D.pdf>

betekent dit, dat ze moeten worden uitgebreid met functies om (i) de kandidaat termen bij een extern systeem te kunnen opgeven, en (ii) notificaties kunnen ontvangen wanneer de kandidaat-termen in het externe systeem zijn verwerkt.

De collectiebeheerder moet zelf bepalen aan welke domeinexpert een kandidaat-term wordt voorgesteld. Op termijn zou het mogelijk kunnen zijn om een dienst te ontwikkelen die suggereert (of zelfs bepaalt) waar de kandidaat-term naar toegaat.

Domeinexpert

Beheer domeintermen

Er zijn reeds verschillende systemen beschikbaar voor het beheer van gecontroleerde vocabulaires. Het beheer van domeintermen wordt ook vaak gedaan in collectiebeheersysteem. Deze systemen zijn meestal alleen intern bij een erfgoedinstelling te benaderen, zodat een exportfunctie nodig is om de termenlijst publiekelijk beschikbaar te maken. Voor het gezamenlijk en gedistribueerd beheren van domeinkennis kan ook gekozen worden voor bestaande systemen zoals Wikipedia en Wikidata. Onder beheer valt ook het opschonen van een gecontroleerd vocabulaire, bijvoorbeeld om dubbele termen te verwijderen of het formaat van namen gelijk te trekken.

Linken van gecontroleerde vocabulaires

Domeinexperts moeten hun gecontroleerde vocabulaires verbinden met andere bronnen in het termennetwerk. Voor een bruikbare oplossing moet de gebruiker controle hebben over de kwaliteit van de resultaten. Dit is de reden dat volledig automatische technieken meestal niet voldoende zijn. Deze technieken moeten gecombineerd worden met een manier om de resultaten te kunnen analyseren. Voor een duurzame oplossing is tevens belangrijk dat domeinexperts niet alleen éénmalig links kunnen leggen, maar dat termen ook gelinkt blijven als er updates plaats vinden.

Het linken naar een externe bron kan ook gebruikt worden om een analyse te maken van een interne bron. De links geven bijvoorbeeld inzicht in de overlap tussen bronnen. Andersom geldt dat hetgeen dat niet gelinkt is, waarschijnlijk niet beschikbaar is in de externe bron. Op deze manier kan het linken gebruikt worden voor bijvoorbeeld het analyseren van de sterke punten van interne vocabulaires.

Het analyseren van meerdere bronnen gezamenlijk, wordt nog niet ondersteund door bestaande tools. Het gebrek aan tools voor deze analyse wordt duidelijk in het project Visuele thesaurus voor online mode-erfgoed, waar veel handwerk nodig is om de interne bronnen van verschillende erfgoedinstellingen te vergelijken.

Knooppunt

Publiceren gecontroleerde vocabulaires

Het is essentieel dat de gecontroleerde vocabulaires op een interoperabele manier worden gedeeld. Een voor de hand liggende optie is om deze data te delen met behulp van de Semantic Web standaarden van het World Wide Web consortium¹¹, zoals RDF. Details van het gebruik van deze representatietalen vallen buiten de scope van dit advies.

¹¹ https://www.w3.org/2001/sw/wiki/Main_Page

Naast het formaat waarin de vocabulaires worden gedeeld, is ook de wijze van communicatie van de data van belang. Een voor de hand liggende manier is om de data simpelweg als files beschikbaar te maken. Daarnaast zijn er verschillende initiatieven die meer controle geven in de toegang tot de gepubliceerde data. Vanuit de Semantic Web community zijn er bijvoorbeeld het publiceren via de principes van Linked Open Data¹², het beschikbaar maken via een SPARQL endpoint¹³ en Linked Data Fragments¹⁴. Het “Open Archives Initiative Protocol for Metadata Harvesting” (OAI-PMH)¹⁵ biedt een oplossing voor het periodiek ophalen van updates. Via OAI-PMH kan niet alleen alle data worden opgevraagd, maar ook bijvoorbeeld de updates vanaf een bepaalde datum.

Beheer en integratie van schema

De knooppunten beheren de schema's waarbinnen domeinexperts informatie kunnen representeren. Voor personen en plaatsen is deze informatie anders dan voor concepten of gebeurtenissen. Er bestaan reeds verschillende schema's voor types, zoals personen en plaatsen. Het Netwerk Digitaal Erfgoed zou samen met de bestaande knooppunten een voortrekkersrol moeten nemen om te bepalen welke schema's relevant zijn voor de erfgoedsector.

Registratie knooppunten

Knooppunten registreren zich bij het Netwerk Digitaal Erfgoed. Het knooppunt geeft aan waar welke bronnen te vinden zijn. Vanuit de Open Data community zijn er verschillende platformen ontstaan voor de registratie van datasets, een voorbeeld is CKAN¹⁶.

Netwerk Digitaal Erfgoed

Kwaliteitscontrole

Het Netwerk Digitaal Erfgoed controleert de kwaliteit van de bronnen die worden aangeboden door de knooppunten. Hiervoor moet ze de content periodiek aggregeren. De exacte kwaliteitsaspecten moeten nader worden bepaald. Voorbeelden zijn de beschikbaarheid van de bron, syntactische correctheid, gebruik van datatypes etc.

NDE brengt ook in kaart welke bronnen beschikbaar zijn en hoe deze met elkaar gelinkt zijn. De resulterende data wordt publiek gedeeld en visueel beschikbaar gemaakt als een data cloud. In een uitgebreidere versie, zou de gebruiker door de schema's van de bronnen kunnen navigeren. Welke soort concepten zijn beschikbaar en wat zijn hun attributen en onderlinge relaties?

Geavanceerdere diensten zouden in kaart kunnen brengen waar nieuwe mogelijkheden liggen: Welke bronnen overlappen maar zijn nog niet geïntegreerd? Welke (sub)-domeinen zijn nog niet beschikbaar?

Toegang tot de knowledge graph

Om een collectie te annoteren met het termennetwerk, moet een collectiebeheerder kunnen zoeken naar concepten uit het hele netwerk. Een dienst, vergelijkbaar met de API van de Google Knowledge Graph¹⁷, is nodig voor het termennetwerk. Bovendien moet het mogelijk zijn voor instellingen om de configuratie van het zoekstelsel aan te passen naar wat nodig is voor hun collecties en specifieke annotatievelden.

¹² <http://linkeddata.org/guides-and-tutorials>

¹³ <https://en.wikipedia.org/wiki/SPARQL>

¹⁴ <http://linkeddatafragments.org/>

¹⁵ <https://www.openarchives.org/OAI/openarchivesprotocol.html>

¹⁶ <http://ckan.org/>

¹⁷ <https://developers.google.com/knowledge-graph/>

In bestaande vocabulaire beheersystemen is het mogelijk om te zoeken naar de termen binnen dat systeem. Op dit moment is er echter geen overkoepelende dienst, waarmee een knowledge graph van erfgoed informatie als geheel doorzocht kan worden, en waarmee de configuratie van het zoek-systeem aangepast kan worden aan de eisen van de collectie en de annotatievelden.

5 Tot slot

5.1 Realisatie dienstenconcept

In dit advies hebben we de contouren aangegeven voor een dienstenconcept rondom een knowledge graph van erfgoed informatie. We raden het Netwerk Digitaal Erfgoed aan om voor de verdere uitwerking aan te sluiten bij bestaande initiatieven in de erfgoedsector. Hierbij is het raadzaam om zowel de doelen van het werkpakket bruikbaar als het werkpakket zichtbaar mee te nemen. Het uiteindelijke doel van verbindingslaag is immers om de zichtbaarheid te verbeteren. Voorbeelden van bestaande initiatieven waarin de rol van de verbindingslaag concreet kan worden toegepast zijn de visuele thesaurus voor online mode-erfgoed, het netwerk oorlogsbronnen, de afstemming van de AAT en de Erfgoed Thesaurus.

5.2 Toegang tot collecties

Integratie van erfgoedcollecties via een gedeeld termennetwerk is een noodzakelijke voorwaarde om de toegang tot deze collecties te verbeteren. Daarnaast zijn er tools en/of services nodig die gebruik maken van het termennetwerk om toegang te geven tot de collecties. De exacte eisen zullen afhangen van de specifieke toepassing. Algemeen toepasbare operaties die nodig zullen zijn omvatten in elk geval:

- Zoeken naar concepten uit de knowledge graph
- Links tussen concepten gebruiken om verschillende interpretaties in de knowledge graph te communiceren naar de gebruiker (denk aan: disambiguatie pagina's in Wikipedia)
- Zoeken en navigeren naar (collectie)objecten (mogelijk uit verschillende collecties) via concepten uit de knowledge graph
- Links tussen concepten gebruiken om via andere terminologie of taal te zoeken
- Overzicht verkrijgen van de collecties die objecten hebben over een concept
- Bij een gegeven collectie-object (of set van objecten) via concepten uit de knowledge graph suggesties krijgen van objecten uit andere collecties

Naast de mogelijkheden om uniforme toegang te realiseren over de collecties heen, zien wij ook een belangrijke meerwaarde voor de erfgoedinstellingen zelf. Van oudsher hebben musea, bibliotheken en archieven de rol om mensen de weg te wijzen in de fysieke gebouwen die ons erfgoed huisvesten. Peter Edsen¹⁸ schetst in een interessant betoog hoe musea ook op het Web deze sociale rol moeten opeisen. Om dit te realiseren moet het ontwikkelen van nieuwe applicaties drastisch eenvoudiger worden. De instituten, maar ook de knooppunten en samenwerkingsverbanden, moeten makkelijk kunnen experimenteren en snel nieuwe vormen van gebruikerservaringen kunnen ontwikkelen.

Grote nieuwsplatformen, zoals van de Guardian, spelen al langere tijd met haar rol op het Web. Alhoewel hun doelen een sterkere commerciële motivatie zullen hebben, kunnen culturele instellingen hier wel wat van leren. David Weinberger legt uit, hoe de ontwikkeling van platformen binnen de nieuwsorganisaties hier een belangrijke rol in spelen.¹⁹ Die platformen waren, in eerste instantie, niet anders bedoeld dan het faciliteren van het hergebruik van de eigen content buiten de muren van de eigen organisatie: hergebruik in de apps, hackathons en visualisaties die de wereld zouden gaan ver-

¹⁸ <https://medium.com/tedx-experience/dark-matter-a6c7430d84d1>

¹⁹ <http://shorensteincenter.org/open-news-platforms-david-weinberger/>

anderen. Terugkijkend blijken diezelfde platformen een uitermate grote meerwaarde te bieden voor de nieuwsorganisaties zelf. Ze kunnen nu veel makkelijker *zelf* nieuwe applicaties maken dan voorheen, experimenten met nieuwe vormen van uitgeven en daarnaast natuurlijk samenwerken met interessante nieuwe partijen.

Het is uitermate belangrijk dat erfgoedinstellingen hun ruwe data delen en integreren. Maar daarnaast moeten instellingen zelf ook een gebruiker worden van deze data; misschien blijken ook zij straks wel de belangrijkste gebruikers van de knowledge graph voor erfgoed.