

Digitaal Erfgoed Bruikbaar

PROJECTPLAN

Linked data pilotproject Thesaurus Oorlogsbronnen

project B3-3

05-07-2016

NIOD, Edwin Klijn en Lizzy Jongma

Rijksdienst voor het Cultureel Erfgoed, Joop Vanderheiden

Opdrachtgever: NDE, Wilbert Helmus, coördinator WP Bruikbaar

| | | | |
|--------------------|------------------|------------------|---------------------|
| Versie | 2.8 | Auteur(s) | Lizzy Jongma |
| Datum | 04-07-2016 | Status | Definitief |
| Wijzigingen | Publiceer versie | | |

Introductie

Doel van dit document

Het Netwerk Oorlogsbronnen zal binnenkort in OpenSkos een WO2-thesaurus publiceren. Deze thesaurus zal in Netwerk Oorlogsbronnen worden gebruikt als een middel om WO2-collecties inhoudelijk met elkaar te verbinden. Voor het Netwerk Digitaal Erfgoed (NDE) is deze gebruikerswens een ideale *use case* voor de concrete uitwerking van de ambitie om een knowledge graph te realiseren. Landelijk beschikbare terminologiebronnen worden dan in een linked data infrastructuur gebruikt door erfgoedinstellingen om hun lokale collecties te ontsluiten en te verbinden.

Dit document is een beknopte uitwerking van het voorstel voor het uitvoeren van het project “Linked data thesaurus oorlogsbronnen”.

Op basis van dit Projectvoorstel geeft de Stuurgroep Digitaal Erfgoed Bruikbaar goedkeuring voor uitvoering van het project.

Iedere significante wijziging in de inhoud van het Projectvoorstel moet worden afgestemd met de Stuurgroep Digitaal Erfgoed Bruikbaar.

Context

Het project is een onderdeel van het werkprogramma Digitaal Erfgoed Bruikbaar van het NDE. Dit samenwerkingsverband richt zich op de ontwikkeling van een stelsel van landelijke voorzieningen en diensten voor het verbeteren van de zichtbaarheid, bruikbaarheid en houdbaarheid van digitaal erfgoed. Het netwerk is gestart op initiatief van het Ministerie van Onderwijs, Cultuur & Wetenschap (OCW) en past in de nationale strategie digitaal erfgoed.¹ Deelnemers aan het samenwerkingsverband zijn onder meer de Koninklijke Bibliotheek, het Nederlands Instituut voor Beeld en Geluid, de Rijksdienst voor het Cultureel Erfgoed, de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW), het Nationaal Archief en kenniscentrum DEN. Binnen het werkprogramma van de NDE zijn drie werkpakketten overeenkomend met de drie lagen uit de strategie voor digitaal erfgoed, namelijk:

Werkpakket 1: Zichtbaar

Werkpakket 2: Bruikbaar

Werkpakket 3: Houdbaar

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

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

Dit project past daarmee naadloos in de doelstellingen van *Digitaal Erfgoed Bruikbaar* en met name in het onderdeel waarbij het termennetwerk gerealiseerd zal gaan worden.

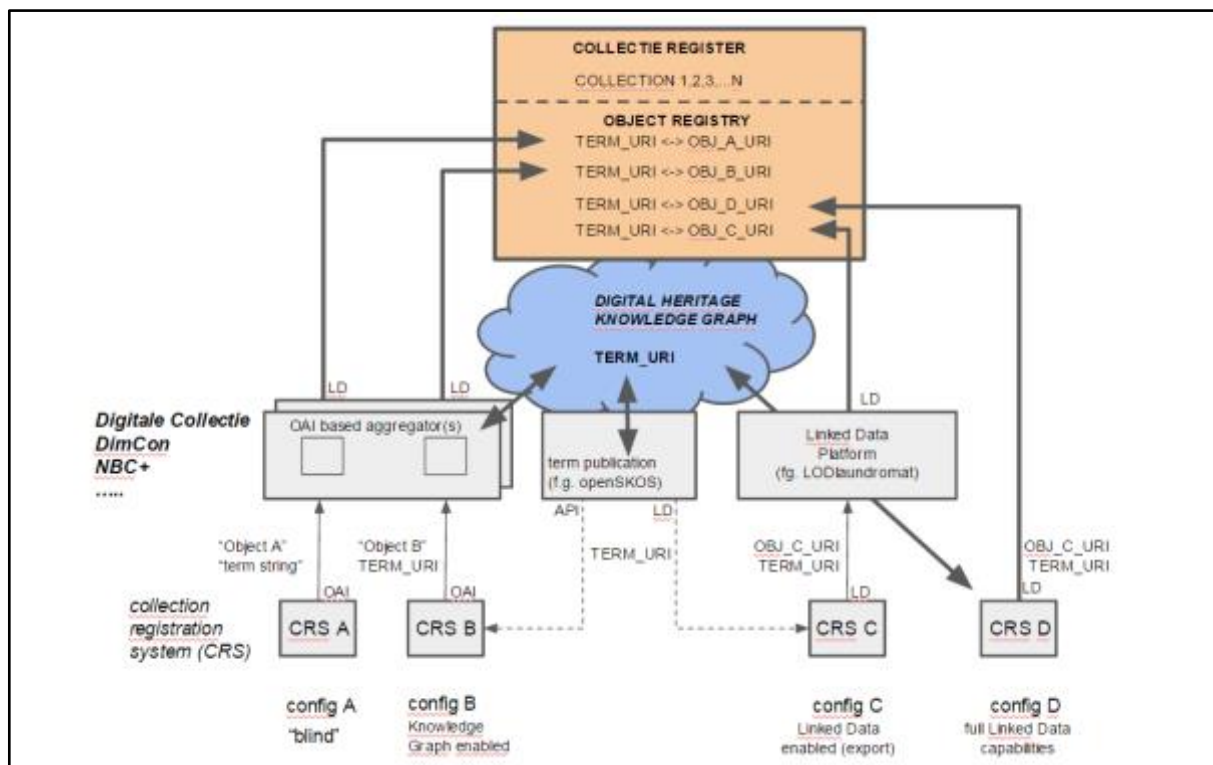
¹ <http://www.rijksoverheid.nl/documenten-en-publicaties/publicaties/2015/03/09/nationale-strategie-digitaal-erfgoed.html>

Uitdaging

Het doel van dit project is om in vier veel gebruikte collectiebeheersystemen (Adlib, De Ree, OCLC, Memorix Maior) de WO2 thesaurus vanuit OpenSkos aan te roepen en zowel retrospectief als nieuwe thesaurustermen toe te kennen aan de WO2 gerelateerde objecten in de collecties. Bij de implementatie hiervan wordt leveranciers gevraagd om een generieke oplossing die ook werkt voor het aanroepen van andere terminologiebronnen in de knowledge graph. Voor De Ree is dit een vervolgproject op het project bij Groninger Archieven.

Doelstellingen zijn:

- dat collecties verrijkt kunnen worden met in het collectiebeheersysteem opgeslagen uri's van thesaurustermen (term_uri uit onderstaand schema)
- dat in genoemde collectiebeheersystemen externe thesauri kunnen worden aangeroepen en bij toekenning als uri lokaal worden opgeslagen (ontwikkeling workflows)
- dat de collecties in het Linked Data Collectieregister bekend gemaakt worden
- dat de WO2 gerelateerde objecten uit die collecties in het Linked Data Object Register bekend gemaakt worden



Figuur 1. Samenhang termennetwerk en het register voor collecties en objecten

Achtergrond Netwerk Oorlogsbronnen

Voor het Netwerk Oorlogsbronnen is het doel van dit project is het ontwikkelen, beheren en onderhouden van een Linked Open Thesaurus voor bronnen WO2. De thesaurus is onderdeel van de eigen basisinfrastructuur digitale Oorlogsbronnen² en dient als koppelstuk tussen verschillende WO2 collecties. De thesaurus bevat begrippen die relevant zijn voor het documenteren van WO2 bronnen en dient als gestructureerde referentie voor Oorlogserfgoed beherende instellingen. Dit wil zeggen dat zij, door gebruik te maken van de begrippen uit de WO2 thesaurus, zich verbinden aan een gedeeld – en voor computers leesbaar – referentiekader.

Binnen dit project zal ook een eerste uitrol van Persistent Identifiers plaatsvinden: in dit project om de thesaurustermen te voorzien van een Persistent URI. In vervolgprojecten zullen de technieken voor persistentie en semantisch verbinden verder uitgebouwd worden.

Op dit moment wordt de WO2 thesaurus beheerd, onderhouden en ontwikkeld door het NIOD. De thesaurus wordt ontwikkeld in de OpenSKOS omgeving van Picturae en Beeld en Geluid.

Projectplan

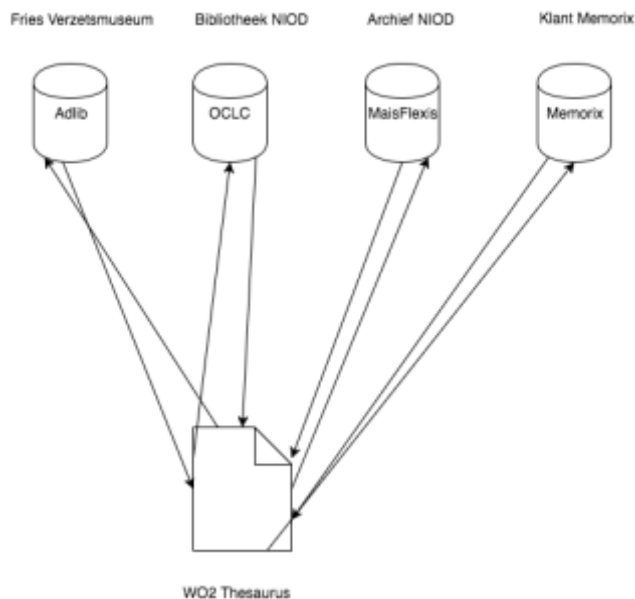


Figure 1 - Algemene indruk ontlenen WO2 termen

² Deze basisinfrastructuur wordt bij het schrijven van dit projectplan nader uitgewerkt en zal in separate projecten ingericht worden.

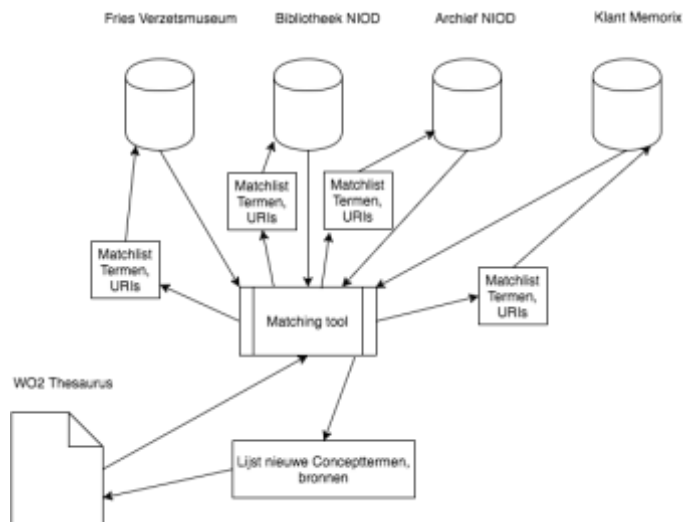


Figure 2 - Algemene indruk proces retrospectivisch matchen termen

Doel van het project is het beschikbaar maken van de WO2 thesaurus voor een viertal doelgroepen:

1. Beheerder(s) van de thesaurus
2. Leveranciers van Erfgoed Registratie Systemen
3. Gebruikers van de thesaurus (collectiebeheerders extern)
4. NDE/OpenSKOS platform

Beschikbaar maken betekent dat de WO2 thesaurus voor alle vier de doelgroepen betekent dat er een basisset aan termen in een hiërarchische structuur in het van tevoren gekozen formaat SKOS open gesteld wordt.

De vier doelgroepen kunnen de termen online bekijken, downloaden en/of via de eigen Registratiesystemen aan beschrijvingen toekennen. Afhankelijk van de rol kunnen betrokkenen (ook) termen voorstellen, aanvullen enzovoorts.

Om de resultaten van dit project te behalen zal samen met drie projectpartners, gekozen omdat de drie partners alle drie over verschillende Registratie Systemen beschikken, en met hun leveranciers vastgesteld worden hoe de WO2 thesaurus geïncorporeerd wordt.

Als de incorporatiefase uitgevoerd is dan zullen de digitale resultaten – lees de semantische verrijkte records van de digitale bronnen van de gebruikers - aangeleverd worden naar het het Objectenregister.

Thesaurustermen van de gebruikers die nog niet in de thesaurus voorkwamen maar in de incorporatiefase zichtbaar zijn geworden zullen als concepten in de WO2 thesaurus opgenomen worden.

Project Thesaurus WO2

Het project kent drie aspecten:

- a. Beheren, onderhouden en ontwikkelen van de thesaurus (NIOD/NOB kant)
- b. Gebruiken van de termen bij registraties (overige doelgroepen).
- c. Retro matchen van termen in bestaande registraties (overige doelgroepen).

De beheerderskant

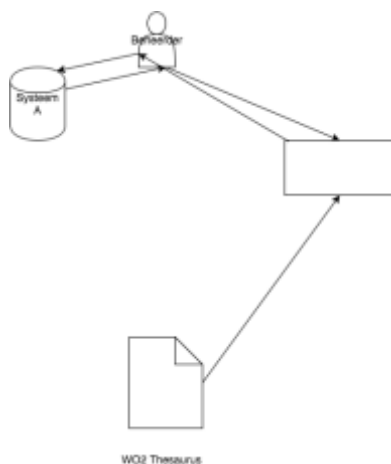
Aan de beheerderskant wordt op dit moment al volop gewerkt aan het ontwikkelen van de WO2 thesaurus. Parallel aan dit project zal een analyse plaatsvinden van de wijze waarop de thesaurus professioneel beheerd, onderhouden en gedeeld kan worden. Uitgangspunt is een open, webbased omgeving waarbinnen het zowel voor de beheerders binnen het NIOD als voor externe partners eenvoudig zal zijn om mee te werken aan de verdere ontwikkeling van de thesaurus. Centraal in dit project staat dat de thesaurus door de mens hanteerbaar en deelbaar en door de machine leesbaar en herbruikbaar dient te zijn: SKOS zal dus het output format blijven en vanuit het Netwerk Oorlogsbronnen zal gekeken worden hoe de beheerderskant geoptimaliseerd kan worden.

Vanuit dit project wordt nauw samengewerkt met de andere NDE projectgroepen, die inmiddels ervaring hebben met het werken met professionele thesaurus software. Het NIOD zal op basis van ervaringen en adviezen vanuit de andere werkgroepen een keuze maken³.

De gebruikerskant: gebruiken van termen bij registraties

Voor de gebruikerskant zijn een drietal scenario's denkbaar voor het werken met de WO2 thesaurus: beginnend bij het meest eenvoudige scenario tot het meest geavanceerde en geautomatiseerde scenario.

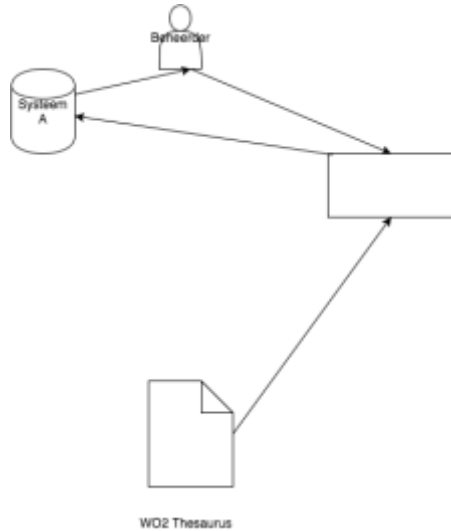
Scenario 1 - light:



Termen en de bij de term behorende URI worden handmatig opgezocht in de door mensen leesbare web interface van de WO2 thesaurus. De termen (wellicht scopenotes) en URI's worden door middel van knippen en plakken in het eigen systeem ingevoerd.

³ Het RCE is op dit moment druk bezig met het testen van Poolparty voor NDE partners. Aansluiten bij dit programma lijkt op dit moment een goede optie, hoewel er ook gekeken zal worden naar xTree, waar het RKD mee werkt in het kader van de AAT.

Scenario 2 - medium:



Termen worden “handmatig” opgezocht in de door mensen leesbare web interface van de WO2 thesaurus. Door middel van een slimme download worden de begrippen in een machine leesbaar formaat binnengehaald en (automatisch) geïmporteerd in het eigen Registratiesysteem.

Scenario 3 – advanced:



Het advanced scenario zal in nauwe samenwerking met leveranciers (via/met de kwartiermaker knooppunt leveranciers) uitgevoerd worden. In dit scenario zullen termen vanuit de WO2 thesaurus in machine leesbaar SKOS-formaat beschikbaar gemaakt worden voor verschillende registratiesystemen. In samenspraak met de leveranciers en de projectpartners zal gekeken worden hoe de thesaurus geheel of gedeeltelijk in de software opgenomen kan worden zodat invoerders bij het registreren van erfgoed automatisch termen uit de WO2 Thesaurus aangeboden krijgen. De semantische verrijking van de trefwoorden wordt ook automatisch opgenomen in de registratiesystemen en kan zodoende mee doorgeleverd worden. Ook voor aanvullingen en aanpassingen worden geautomatiseerde oplossingen bedacht.

De gebruikerskant: retro matchen van termen

Tot slot zit er retro, matching werk vast aan dit project: de projectpartners hebben al veel registraties voltooid en zelf aan deze registraties thesaurustermen/trefwoorden toegekend. Samen met de partners zullen registraties die voorzien zijn van trefwoorden/thesaurustermen geautomatiseerd geanalyseerd worden. De resultaten van dit matchen zullen met de partners besproken worden en matches die voorzien kunnen worden van URI's zullen geschoond, geannoteerd en met de juiste URI en bronvermelding terug geleverd worden aan de partners.

Dit proces zal naar alle waarschijnlijkheid nieuwe termen en inzichten voor de WO2 thesaurus opleveren. De termen en scopenotes zullen (in kandidaat) in de thesaurus opgenomen worden en door de beheerder(s) van de thesaurus geëvalueerd worden.

Projectresultaten

De resultaten van dit project zijn:

1. Gestructureerde werkwijze en beschikbaarstelling van de WO2 thesaurus
2. Basis WO2 thesaurus is online beschikbaar voor mens en machine
3. De termen in de thesaurus zijn voorzien van persistente URI's
4. WO2 thesaurus is beschikbaar onder een <http://opendatacommons.org/licenses/by/1-0/> Licentie: conform andere erfgoed thesauri zoals de AAT en ULAN.
5. Beheer, onderhoud en verdere ontwikkeling van WO2 thesaurus geborgd. Afspraken NOB en NIOD over beheer formaliseren.
6. WO2 thesaurus op verschillende manieren beschikbaar voor erfgoedinstellingen met verschillende Erfgoed Registratie Systemen: scenario 1 t/m 3 zijn getest en met de partners is een keuze voor het voor hen juiste scenario gemaakt
7. Met softwareleveranciers zijn afspraken gemaakt over implementatie en/of uitrol van de WO2 thesaurus in de registratiesystemen waarbij rekening wordt gehouden met verdere ontwikkelingen van de WO2 thesaurus
8. Door de partners aangeleverde data is retro-perspectivisch gematcht en waar mogelijk semantisch verrijkt
9. Objecten en bronnen die semantisch verrijkt zijn, zijn aangeleverd aan het Objectenregister
10. Collectiebeheerders zijn getraind in het werken met de WO2 thesaurus
11. Er is een workshop gehouden voor nieuwe gebruikers van de WO2 thesaurus
12. Het project is geëvalueerd en er is een eindrapportage geschreven die als handleiding kan dienen voor andere thesaurusbeheerders, - ontwikkelaars

Communicatie

Helderheid over de stand van zaken is voor alle betrokkenen in dit project noodzakelijk. Daarom staan er vier project-nieuwsbrieven gepland die gedurende de loop van het project vanuit een centraal communicatiepunt (NOB) naar alle betrokkenen (zie 'projectteam') wordt verstuurd. Om de project-nieuwsbrief zo informatief mogelijk te maken vraagt het centrale communicatiepunt waar nodig om input van projectteamleden.

Daarnaast wordt van alle betrokkenen verwacht zelf ook aandacht te geven aan het project via eigen communicatiekanalen als nieuwsbrieven, websites en sociale media. Daarvoor wordt een toolkit geleverd met o.m. de logo's van alle betrokkenen. Er kan gebruik gemaakt worden van de informatie/teksten uit de project-nieuwsbrief NOB. In overleg kan vanuit het NOB meer content worden aangeleverd. Denk bijvoorbeeld aan reeds opgestelde tweets of facebookberichten, interviews, persberichten, webteksten, etc.

Belangrijk is dat de communicatie naar buiten vanuit projectteamleden plaatsvindt nadat de project-nieuwsbrief is verspreid. Hierdoor wordt voorkomen dat projectteamleden updates uit media i.p.v. intern te lezen krijgen. Uiteraard kan in overleg wel gekozen worden voor gezamenlijk gelijktijdige communicatie over bijvoorbeeld aanvangspunten binnen het project. Het NOB zal initiatief nemen tot het opzetten van een communicatiewerkgroep t.b.v. afstemming.

Verwachte onderwerpen per project-nieuwsbrief zijn:

Eind juli: Goedkeuring projectplan (NOB/NDE), kennismaking betrokkenen (NOB), eerste stappen (NOB), status thesaurus invoer|beheer|onderhoud (NIOD).

Eind september: Status thesaurus invoer | beheer | onderhoud | redactiestructuur (NIOD), verslag analyse Editor OpenSKOS (NIOD/OpenSKOS), status scenario's (NOB), verslag implementatie WO2-thesaurus (NOB/projectpartners/softwareleveranciers), verslag retro matchen ophalen bronnen (NOB).

Begin november: Verslag redactiestructuur thesaurus (NIOD), status thesaurus invoer | beheer | onderhoud (NIOD), verslag ontwikkelplan scenario's (NDE/NOB/softwareleveranciers), status plugins | API's (leveranciers), status test thesaurus (projectpartners), verslag resultaten retro matchen (NOB/projectpartners), status terugleveren data (NOB/projectpartners), start training en overdracht (NOB), status ontwikkeling webinterface (NOB/NIOD/OPENSkos).

Begin januari 2017: Rapportage project: resultaten en evaluatie.

Om de communicatiedoelstellingen – helderheid voor projectteamleden en informatief naar het publiek – te bereiken zijn een aantal meetbare en tijdgebonden Kritische Prestatie Indicatoren opgesteld (KPI's). Deze worden toegepast op de communicatie die vanuit het NOB wordt gevoerd, maar ook op die vanuit projectteamleden. In overleg met projectteamleden worden dan ook indicatoren voor publieksbereik vastgesteld, en gevraagd om de resultaten van communicatie bij te houden. Zie hieronder een opzet:

| Strategie | Middel | Meetbare doelstelling (KPI) |
|--------------------------------------|---------------------------|---|
| Informereren projectteamleden | Project-nieuwsbrief | Aantal ontvangers Ev. aan- en afmeldingen Openingspercentage Bouncepercentage |
| Informereren erfgoedsector | Persbericht | Aantal keer gepubliceerd in vakmedia |
| Informereren erfgoedsector | Nieuwsbrief | Artikel project WO2-thesaurus aantal keer geopend |
| Informereren erfgoedsector / publiek | Website | Artikel project WO2-thesaurus aantal keer geopend |
| Informereren publiek | Persbericht | Aantal keer gepubliceerd in printmedia Aantal keer gepubliceerd op nieuwswebsites / -blogs |
| Dialogiseren publiek | Sociale media: Twitter | Aantal keer getweet |

Projectteam

Management - Opdrachtgever:
Wilbert Helmus (NDE)

Management - Opdrachtnemer:
Edwin Klijn (NOB)

Uitvoering - Projectleider:
Lizzy Jongma (NOB)

Uitvoering - WO2 thesaurus team:
Marjo Bakker (NIOD)

Uitvoering - Matching termen voor partners:
Inhuur (NOB)

Uitvoering - Projectpartners:
Fries Verzetsmuseum
NIOD
Brabant Cloud

Stappen om tot resultaten te komen

1. Ontwikkelen en accepteren definitief projectplan
 - a. Schrijven projectplan
 - b. Goedkeuring projectplan
 - c. Kennismaken projectleider en projectteam
 - d. Initiëren eerste stappen
2. Thesaurus beheer en onderhoud
 - a. Invoeren termen, relaties, scope notes enz.
 - b. Bepalen van mate waarin en wijze waarop externen mee kunnen werken en helpen thesaurus door te ontwikkelen
 - c. Selecteren/vaststellen Persistent Identifiers voor thesaurusterminen (creëren bewustzijn en commitment)
 - d. Procedure voor terugkoppeling nieuwe concepttermen vanuit matching fase
3. Registratie – stap 1: vaststellen
 - a. Vaststellen welke velden in de registraties relevant zijn voor matching
 - b. Vaststellen per softwarepakket of er een uitgebreide thesaurusstructuur aanwezig is met mogelijkheden om URI's en wellicht andere data op te nemen
 - c. Vaststellen welke software geschikt is voor (semi-) geautomatiseerde retro matching
 - d. Vaststellen welke bronnen gematcht worden en omvang van de bronnen vaststellen (analyse)
4. Matchen registraties – stap 2
 - a. Op basis van de analyse in fase 1 zal per partner en softwarepakket een scenario vastgesteld worden om toekomstig registratiewerk middels een plug-in te kunnen verrijken met termen uit de WO2 thesaurus
 - b. Op basis van analyse zal een realisatieplan uitgewerkt worden
5. Retro matchen registraties – stap 3
 - a. Per partner data matchen met WO2 termen
 - b. Per partner bespreken matching resultaten met collectiebeheerder

- c. Op basis van resultaten en gesprek met beheerder terug leveren verrijkte data
 - d. Overall analyse van matches en waarschijnlijke matches: hoe succesvol is het geautomatiseerd retro matchen van de WO2 thesaurus
- 6. Thesaurus online beschikbaar maken in zowel mens- als machine leesbaar formaat en als Linked Open Data bron in het Semantisch Web te plaatsen
 - a. Web interface voor mensen maken
 - b. Downloadformaten ontwikkelen voor half geautomatiseerde gebruikers
 - c. API('s) ontwikkelen voor software leveranciers
 - Koppelen thesaurus
 - i. Optie: om de WO2 thesaurus via een API in de thesaurusstructuur van softwareleveranciers te plaatsen
 - ii. Optie: thesaurus via een API of als periodieke update in de CMS-systemen van erfgoedinstellingen in te lezen
- 7. Trainen en overdracht
 - a. Verzorgen trainingen in gebruik WO2 thesaurus en semantisch verrijken van registraties
 - b. Workshop werken met Thesaurus WO2 voor geïnteresseerde instellingen
 - c. Evaluatie project en projectresultaten
 - d. Rapportage en documentatie project
 - e. Scripts en tools waar mogelijk op Github beschikbaar gemaakt voor derden