

## **Doel**

De pilot Infrastructuur stelt de requirements en criteria op voor de uiteindelijke E&L-infrastructuur en een historische geocoder. De pilot doet onderzoek naar een duurzame basisinfrastructuur voor digitaal cultureel erfgoed en geo-informatie en onderzoekt de mogelijkheden van hergebruik van bestaande technologie en architectuur, en knelpunten van functionaliteit en schaalbaarheid. De pilot onderzoekt eveneens de bouwstenen voor een historische geocoder.

## **Deelnemende partijen**

- Rijksdienst voor het Cultureel Erfgoed - 100 uur
- Waag Society - 100 uur
- Stichting Bibliotheek.nl - 100 uur
- Erfgoed & Locatie - 100 uur
- Bob Coret (Coret Consultancy) - 20 uur

## **Doelgroep**

De pilot test de op te leveren producten voor deze doelgroepen:

- wetenschap, onderzoek, overheid;
- projectpartners, erfgoedinstellingen en creatieve industrie;
- aggregatoren: Europeana, DimCon en Digitale Collectie Nederland.

## **Scope**

De pilot Infrastructuur onderzoekt kleinschalig de basale infrastructurele elementen die nodig zijn voor het goed functioneren van Erfgoed & Locatie als linked (open) geodata-omgeving. Uitgangspunt hierbij is dat de pilot onderzoek doet naar en vergelijking maakt van verschillende bevragsingsmodellen en oplossingsrichtingen, vanuit de vraag in welke koppelvlakken de infrastructuur moet voorzien.

- De pilot buigt zich over de GeoSPARQL standaard als bevragsingsmodel voor de aangesloten collecties en E&L ontologie.
- De Pilot onderzoekt een oplossingsrichting in de vorm van een architectuur met hierin de combinatie van een semantische store (bijvoorbeeld Virtuoso) en een Spatial Data Infrastructure (zoals GeoServer of Mapserver). Een dergelijke oplossingsrichting zal in de Pilot Infrastructuur worden onderzocht door de RCE.
- De pilot maakt een inventarisatie en benoeming van de API's die de functionaliteit van de infrastructuur ontsluiten.
- De pilot realiseert een ontwikkel/testomgeving voor het aansluiten van software afkomstig van andere E&L-pilots, die van de infrastructuur afhankelijk zijn of hiermee communiceren. Op deze ontwikkel/testomgeving combineert het pilootteam reeds ontwikkelde en als open source beschikbare pakketten met aanvullend te ontwikkelen programmacode en API's, waar nodig en haalbaar binnen de pilot. De ontwikkel/testomgeving voldoet op kleine schaal aan de requirements van de beoogde E&L-infrastructuur in het eindstadium zoals voorgesteld in het E&L-architectuurschema. De performance wordt niet getest.
- Het pilootteam brengt verslag uit over knelpunten, gebruikte pakketten en standaarden en doet aanbevelingen over de manier waarop de kleinschalige ontwikkel/testomgeving opgeschaald kan worden naar een hoogwaardige productieomgeving.
- Als onderdeel van de pilot wordt de vorm van de E&L historische geocoder onderzocht die een (uri naar) een geometrie en/of tijdsaanduiding retourneert op basis van een ingegeven en/of

gegenereerde plaats- en tijdsaanduiding. Bij Pilot 2 Geovoculaires wordt de inhoud van de Historische Geocoder onderzocht.

### Randvoorwaarden

- Interoperabel - De ontwikkeling van linked (open) geodata is nog in volle gang. Interoperabiliteit is daarom een belangrijke randvoorwaarde die bereikt kan worden door het toepassen van internationale en Europese standaarden van erkende autoriteit (W3C, OGC).
- Opensource - Vanuit de wens om de in E&L ontwikkelde kennis en producten zo makkelijk mogelijk en tegen zo laag mogelijke kosten ter beschikbaar te stellen aan de erfgoedsector en het risico op een *vendor lock-in* te beperken, wordt gebruik gemaakt van opensource software.
- Op deze pilot zijn de Algemene Pilotvoorwaarden E&L van toepassing.

### Out of scope

De focus ligt op het realiseren van basale functionaliteit en het onderzoek naar schaalbaarheid op het gebied van de geosemantische bevraging. Hieruit volgt dat:

- De uiteindelijke productie-omgeving wordt niet gerealiseerd binnen deze pilot.
- De mapping-functionaliteit voor het aansluiten van externe collecties valt onder pilot 3 Koppeltooling. Indien wenselijk kan dit wel gerealiseerd worden op de ontwikkel/testomgeving van de pilot Infrastructuur.
- De opslaginfrastructuur voor user-generated content kan niet op voorhand ongespecificeerd door de pilot infrastructuur worden toegezegd. Het pilotteam adviseert om in pilot 6 Geotools bij partners in het project te laten inventariseren welke specifieke dienst mogelijk is.
- De semantische bevraging en opslag van thesauri (zowel binnen als buiten E&L) valt onder pilot 2 GeoVoculaires. Indien wenselijk kan dit wel getest worden op de ontwikkel/testomgeving van de pilot Infrastructuur.

### Afhankelijkheden

De pilot levert de infrastructurele basiscomponenten voor:

- de E&L-ontologie van pilot 2 Geovoculaires
- de ontwikkel- en testomgeving voor pilot 3 Koppeltooling
- de interfaces voor de viewers van pilot 5 Toolkits&ProofsofConcepts

Het pilotteam draagt samen met het E&L-projectteam zorg voor onderlinge afstemming tussen de pilotteams wat betreft deze afhankelijkheden, de *lead* hiervoor ligt primair bij de pilotleider.

### Deliverables

Voor alle deliverables geldt dat de oplevering en disseminatie ervan door het pilotteam wordt uitgevoerd in een vorm en op een plaats die wordt aangegeven door het E&L-projectteam. De deliverables zijn aangeduid als D1 t/m D7.

- **D1-PPP.** De algemene pilotdeliverables, waaronder een pilotprojectplan. Lead: Job Spierings
- **D2-Server.** Een gevirtualiseerde serveromgeving met beveiligde toegang voor ontwikkelaars en beheerders van de infrastructuur. De serveromgeving richt in overleg met en op verzoek van andere pilots specifieke serverpakketten in. Lead: Rein van 't Veer

- **D3-Store.** Eén of meer RDF-stores voor verschillende RDF-repositories, met hierin opgenomen een index. Deze index zal gedurende de pilot worden gedefinieerd. De keuze voor semantische modelleersoftware en een triple-store wordt met de andere pilots afgestemd.  
Lead: Rein van 't Veer
- **D4-API's.** Lead: Frans van der Zande
  - **D4A - HG** Specificaties van de E&L Historische Geocoder-API. De Historische Geocoder functioneert als koppelvlak voor het vinden, koppelen en verrijken van geotemporele data en zal in de pilot nader worden gedefinieerd.
  - **D4B - SZ** Specificaties van de E&L semantische zoekfunctionaliteit. De semantische zoek-API handelt zoekvragen semantisch af en retourneert resultaten uit de bij E&L aangesloten collecties. Hiermee is de semantische zoekmachine de primaire interface-toegang voor de ErfGeoViewer (pilot 5 Toolkits & Proofs of Concept). De Semantische zoekmachine zal in de pilot nader worden gedefinieerd.
  - **D4C - TL** - Een mechanisme (API) om de gestandaardiseerde en semantisch verrijkte informatie terug te leveren aan deelnemers.
- **D5-Req.** Use cases (waar mogelijk gebaseerd op door E&L geformuleerde stories), requirements en acceptatiecriteria voor de opgeschaalde serveromgeving in het eindstadium.  
Lead: Enno Meijers
- **D6-Kennis.** Een kennisproduct (wiki, rapport etc.) voor de beschrijving van de E&L-infrastructuur, zo mogelijk Engelstalig. Lead: Job Spierings
- **D7-Business.** Levert informatie voor het vormen van businesscases voor de werkgroep Businessmodellen (E&L-werkpakket 03). Lead: Enno Meijers

### Onderzoeksvragen en aandachtspunten

• *Een index/cache: kan gebruikt worden voor (tijdelijke) opslag en cache van RDF, thumbs en tiles. De centrale vraag is in hoeverre deze voorziening in de E&L-infrastructuur centraal belegd moet worden versus het lokaal cachen bij de collectiehouder. Deelvragen kunnen zijn: is de performance van de web-API's van de grote collectiehouders voldoende om zonder caching bij E&L toe te kunnen? Zijn de kleine- en middelgrote erfgoedinstellingen in staat hiervoor een voorziening te treffen en wat heeft dit voor effect op hun deelname aan E&L? Welke alternatieven zijn hier te bieden? Welke kosten zijn hieraan verbonden en hoe verhouden deze zich tot een verdienmodel?*

In antwoord op deze vraag stelt de pilot Infrastructuur dat Erfgoed & Locatie ernaar zou moeten streven de data zoveel mogelijk bij de bron te laten. Dit laat onverlet dat een antwoord geformuleerd moet worden voor die gevallen (collecties) waar dit niet mogelijk is. De pilot Infrastructuur zal in de loop van de pilot de index/cache nader definiëren.

• *GeoSPARQL is aangewezen als standaard voor de uitwisseling en bevraging van linked (open) geodata door het Open Geospatial Consortium. De vraag is echter in hoeverre de voor de E&L-infrastructuur relevante opensource GIS-producten al GeoSPARQL ondersteunen. De centrale vraag is waar de knelpunten zitten en of E&L duurzaam gebruik kan gaan maken van deze standaard.*

In antwoord op deze vraag stelt de pilot Infrastructuur dat zowel de beschrijvingsmethode als de aanvullende functies van GeoSPARQL wenselijk zijn, maar dat de huidige implementaties onvoldoende stabiel en schaalbaar lijken te zijn. De pilot Infrastructuur zal hierover nader rapporteren in het

eindverslag.

• *Een URI-strategie is een belangrijk onderdeel van elk linked data-initiatief. Welke URI-strategie moet voor E&L worden aangehouden, voor welke objecten en welke gevolgen heeft dit voor de infrastructuur? Dezelfde vragen gelden voor persistent identifiers.*

Bij pilot 2 GeoVoculaires wordt een persistente URI-strategie geformuleerd, dus niet bij de pilot Infrastructuur. Het is wel van belang dat hierover goede afstemming plaatsvindt tussen de pilot Infrastructuur en pilot 2 GeoVoculaires.

• *Linked data is niet de enige vorm van kennisopslag en -overdracht voor E&L. Linked data leent zich minder goed voor de doorgifte van rasterkaartservices (WMS), audiovisuele media, of rasterafbeeldingen. De centrale vraag is welke strategie gevolgd dient te worden om deze media te ontsluiten via de E&L-infrastructuur. Is een centrale WMS-service binnen de E&L-infrastructuur noodzakelijk?*

De infrastructuur-bevraging door toepassingen zal in linked data-vorm antwoord geven. Daarin ligt het niet voor de hand om brondata zoals afbeeldingen, WMS-diensten of audiovisuele media te gaan converteren naar linked data. Deze (resolveable URI's van) mediavormen zullen omschreven moeten worden in RDF, en afgehandeld worden in de front end om bruikbaar te zijn. Hoe deze media in RDF omschreven moeten worden, is aan de pilot Geovoculaires; ook hierbij is goede afstemming tussen de pilots van belang. De pilot Infrastructuur onderzoekt daarbij WMS.

### **Succesfactoren**

- Interoperabiliteit met softwareproducten van andere pilots.
- Duurzame en schaalbare exploitatiemogelijkheden van de pilot-infrastructuur.

### **Pilotleider**

Job Spierings (Waag)

### **Pilotteam**

- Enno Meijers (BNL)
- Bert Spaan (Waag)
- Frans van der Zande (RCE)
- Joop Vanderheiden (RCE)
- Job Spierings (Waag)
- Rein van 't Veer (Erfgoed & Locatie)

### **Planning en aanpak**

De pilot Infrastructuur is in hoofdzaak een onderzoeksproject. Onderzoek en documentatie daarvan vormt daarin de voornaamste post. Verdeling:

*Onderzoek en documentatie: 50% - 250 uur*

- Verzamelen use cases en requirements
- Mappen van use cases naar oplossingsrichtingen

- Toetsen en documenteren van query-resultaten
- Documenteren van prestaties en tekortkomingen
- Onderzoeken en vaststellen API requirements

*Testen en configureren: 10% - 50 uur*

- Installatie en configuratie GeoSPARQL, LD/SDI - 40 uur
- Inladen datasets - 10 uur

*Bijeenkomsten en overleg: 25% - 125 uur*

- 8 bijeenkomsten van +/- 2,5 uur met 6 personen

*Verslaglegging: 15% - 75 uur*

- Schrijven concept-rapportage: 40 uur
- Verwerken en opleveren pre-definitieve rapportage: 20 uur
- Verwerken en opleveren definitieve rapportage: 15 uur

*Totaal: 500 uur, waarvan 100 uur uit te besteden aan externe LOD deskundige/systeembeheerder. De verwachting is dat dit onderhands aan te besteden is aan een zelfstandige of organisatie.*

*Doorlooptijd: 1 oktober 2013 t/m 1 mei 2014.*

### **Meer informatie**

Zie [www.ergoedenlocatie.nl/pilots](http://www.ergoedenlocatie.nl/pilots) voor de laatste versies van de hieronder genoemde E&L-documenten.

- Project Initiatie Document (PID) Erfgoed & Locatie (versie 30; d.d. 08-02-2013)
- Algemene Pilotvoorwaarden E&L (versie 2.1; d.d. 13-02-2013)
- Architectuurschema E&L (versie 3.0; d.d. 03-02-2014)