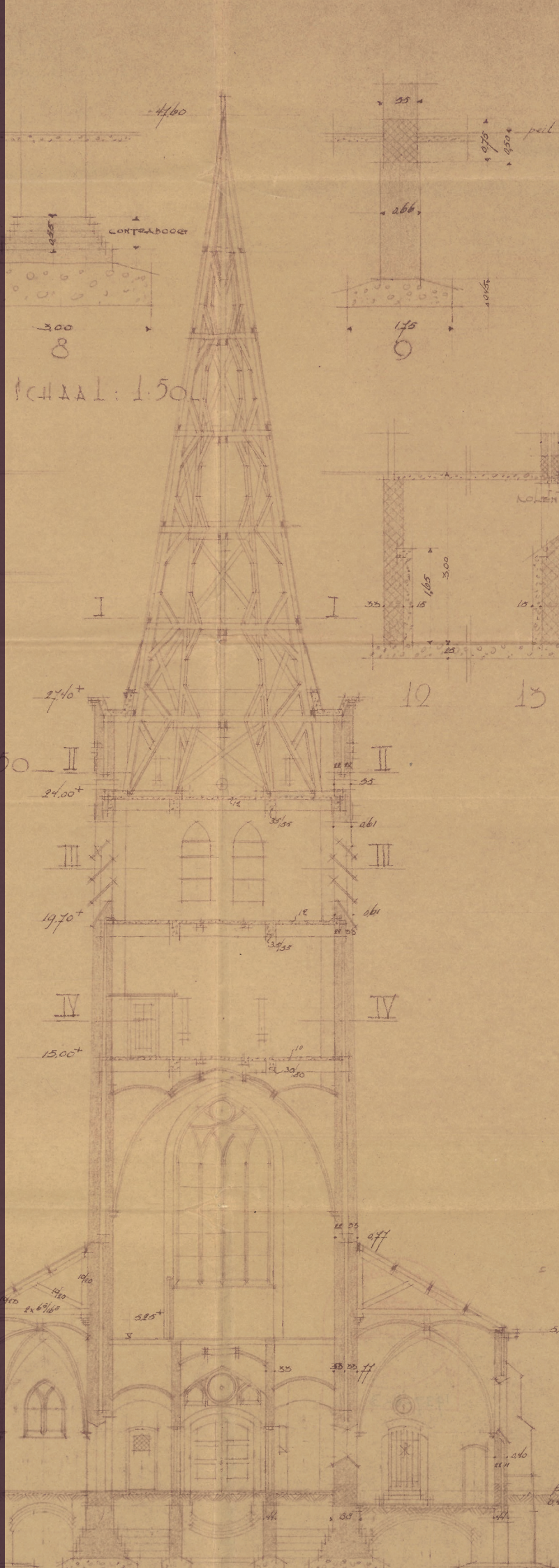


# Papier digitaal

Eindrapport Fase 2

Pilot e-depot Zuid-Utrecht

1 januari 2021





## Inhoud

3	<b>Vowoord</b>
3	<b>Managementsamenvatting</b>
4	<b>Inleiding</b>
5	<b>De doelen van fase 2: 'gedigitaliseerd beschikbaar maken'</b>
5	Opname en beheer van gedigitaliseerde stukken
7	Privacy en auteursrechtenbescherming
8	Verrijking van metadata
11	<b>Conclusie</b>

### Eindrapport Fase 1 pilot E-depot

1 januari 2020

#### Auteurs

Bram Klapwijk (RAZU)

Wietse Bakker (RAZU)

#### Bijdragen van

Corinna Willigenburg

Astrid van Zanten

#### Afbeelding voorzijde:

Bouwtekening veranderen\vernieuwen  
R.K.Kerk aan de Klooster Leuterstraat 35,  
WbD, 1937  
RAZU - 1937-8 - Toegang 404,  
Bouwvergunningen Wijk bij Duurstede  
1935 - 1995.

#### Bezoekadres

Karel de Grotestraat 30  
3962 CL in Wijk bij Duurstede

#### Postadres

Regionaal Archief Zuid-Utrecht  
Postbus 64 3960 BB Wijk bij Duurstede

Telefoon: 088 5300 170

E-mail: info@razu.nl



## Voorwoord

De interesse in gemeentelijke bouwdoSSIERS, als onderdeel van onze overheidsarchieven, groeit snel. Zo zijn er in maart 2020, vlak voor de eerste corona lockdown, circa 80 vragen en verzoeken over bouwvergunningen geregistreerd op de studiezaal van het RAZU. Nog lang niet alle bouwvergunningen zijn overgebracht naar het RAZU, een nog kleiner deel is gedigitaliseerd en slechts een fractie daarvan is ook al geanonimiseerd. Hierdoor is inzage in bouwdoSSIERS een grotendeels handmatige, erg arbeidsintensieve en in groeiende mate digitale taak voor het RAZU en onze gemeenten. De grotendeels vastgelopen huizenmarkt en naderende energietransitie zullen naar verwachting voor nog meer verbouwingen zorgen waardoor de vraag naar bouwdoSSIERS verder zal blijven stijgen. Alle reden dus om in de pilot te kijken naar mogelijkheden voor het e-depot om een deel van deze werklust over te nemen. Met fase 2, een tussenfase aan het einde van 2020, hebben we naar dit vraagstuk gekeken.

## Managementsamenvatting

In januari 2020 is het RAZU samen met de gemeenten Wijk bij Duurstede en Vijfheerenlanden een pilot voor het opzetten van een e-depot gestart. Het e-depot moet in de toekomst het digitale archief van de gemeenten op kunnen nemen. Omdat dit een nieuwe taak voor het RAZU vormt met veel onbekende factoren, was een pilot met leerdoelen noodzakelijk. Hierbij worden zowel de technische eisen onderzocht als ook het geheel aan benodigde mensen en middelen. Zowel het RAZU als de gemeenten moeten aan het einde van de pilot weten waar ze aan toe zijn en welke werkzaamheden ze moeten oppakken. Om te komen tot deze inzichten heeft de projectgroep, bestaande uit het RAZU en medewerkers van de twee deelnemende gemeenten, een traject van twee jaar uitgezet. De pilot is opgedeeld in drie fasen, waarvan de eerste fase is afgerond in september 2020. Het doel van de eerste fase was het kennismaken met het 'opnemen, beheren en tonen' van zaakdoSSIERS in het e-depot. Dit rapport beschrijft de resultaten van de tweede fase, die zich heeft gericht op het opnemen van gedigitaliseerde bouwvergunningen. De tweede fase liep van oktober tot en met december 2020.

De tweede fase had tot doel om gedigitaliseerde bouwdoSSIERS in het e-depot op te nemen en ervaring op te doen met het gedeeltelijk

beperken van de toegang tot deze doSSIERS op grond van de AVG en de Auteurswet. Gedurende de fase is hier een derde doel aan toegevoegd: het verrijken van metadata met geografische data. Voor deze fase is gebruik gemaakt van twee bestanden. Het eerste kleine bestand bestond uit reeds overgebrachte en door het RAZU geanonimiseerde bouwvergunningen. Het tweede en grotere bestand bestond uit recente(re) bouwdoSSIERS die door de gemeente Wijk bij Duurstede zijn geanonimiseerd, maar nog niet zijn overgebracht. Met beide bestanden is het proces van voorbereiding en opname doorlopen. Helaas bleek de laatste stap van het online publiceren van de doSSIERS, door technische gebreken en slechte handleidingen, net niet te lukken. In fase 3 zal verder gegaan worden met deze laatste stap, omdat het oplossen van deze gebreken ook de reguliere dienstverlening van het RAZU zal ondersteunen. Het gesprek met de leverancier is hierover aangegaan.

De fase heeft een aantal belangrijke inzichten opgeleverd met betrekking tot de voorbereiding, bewerking en opname van doSSIERS. Zo is het (voorlopig theoretisch) mogelijk om persoonsgegevens te verbergen in metadata, kunnen watermerken in grote aantallen PDFs tegelijk worden toegevoegd en kunnen metadata op grote schaal en tegen zeer lage kosten worden verrijkt met extra metadata. Ondanks de ontbrekende publicatiestap beschouwt de projectgroep de fase daarmee toch succesvol.



## Inleiding

De tweede fase van de e-depot-pilot richtte zich enkel op papieren bouwdoSSIERS die zijn gedigitaliseerd. Veel gemeenten ondernemen de actie tot digitaliseren van de omvangrijke bouwdoSSIERarchieven om de dienstverlening te vereenvoudigen, of, om door middel van vervanging de papieren doSSIERS te mogen vernietigen.

Uit de eerste fase bleek dat de moderne digitale zaakdoSSIERS complex in elkaar zaten en dat er nog verder onderzocht moest worden welke eigenschappen allemaal bijdragen aan de authenticiteit van dergelijke doSSIERS. Gedigitaliseerde bouwdoSSIERS zijn een stuk minder complex. Ze bestaan uit een aanvraag, tekeningen, een vergunning en eventuele aanvullende stukken. Waar er op papier nog gekozen kon worden voor verschillende vormen van papier, zijn deze unieke eigenschappen in de gedigitaliseerde vorm niet meer aanwezig: met de digitalisering worden alle documenten

omgezet in PDF. Om deze reden werd er bij de planning van de tweede fase van oktober tot en met december 2020 ingeschat dat de korte tijdsduur voldoende moest zijn om te experimenteren met deze vorm van archief.

Dit rapport zal kort ingaan op de gestelde doelen in het projectinitiatiedocument (PID). Daarnaast gaat het ook in op een aanvullend punt, de verrijking van de doSSIERS met geografische metadata. Tot slot geeft het rapport een vooruitblik op de derde fase van de pilot e-depot.

*Afbeelding 1- De bouw van de archiefbewaarplaats van het RAZU, één van de gebouwen waarvan de gemeente Wijk bij Duurstede de bouwvergunning van heeft gedigitaliseerd. Foto: W. Leeman, 2010.*





## De doelen van fase 2: 'gedigitaliseerd beschikbaar maken'

Het PID vormt de basis van de pilot e-depot. In dit PID staan de randvoorwaarden, kaders, planning, het budget en de te behalen doelen centraal. Elk fase-eindrapport start daarom met een terugblik op de in het PID benoemde doelen. Zijn alle resultaten behaald? Is er op punten afgeweken van de doelen of de planning?

De tweede fase had als doel kennis op te doen van het opnemen van gedigitaliseerde archiefstukken, in dit geval specifiek gedigitaliseerde bouwdoSSIERS. Er is gebruik gemaakt van bouwdoSSIERS met bijbehorende metadata van de gemeente Wijk bij Duurstede, die al haar bouwvergunningen al gedigitaliseerd heeft. Het PID stelde twee doelen:

- Opname en beheer van gedigitaliseerd archief;
- Ervaring opdoen met het beheer van auteursrechten en persoonsgegevens in gedigitaliseerd archief.

Voor beide doelen zal worden toegelicht hoe te werk is gegaan en in hoeverre ze behaald zijn. Voorafgaand aan de toelichting moet opgemerkt worden dat, even als in de eerste fase, er technische beperkingen waren, zij het in mindere mate. Daarbij geldt dat de vraagstukken omtrent auteursrecht en persoonsgegevens voornamelijk vanuit een technische invalshoek zijn onderzocht. Er is deels gekeken in hoeverre het wettelijk is toegestaan auteursrechtelijk materiaal of persoonsgegevens te publiceren.

Tot slot hebben we tijdens de tweede fase een aanvullend doel gesteld, dat aansluitend

besproken zal worden:

- Ervaring opdoen met het verrijken van metagegevens.

### Opname en beheer van gedigitaliseerd archief

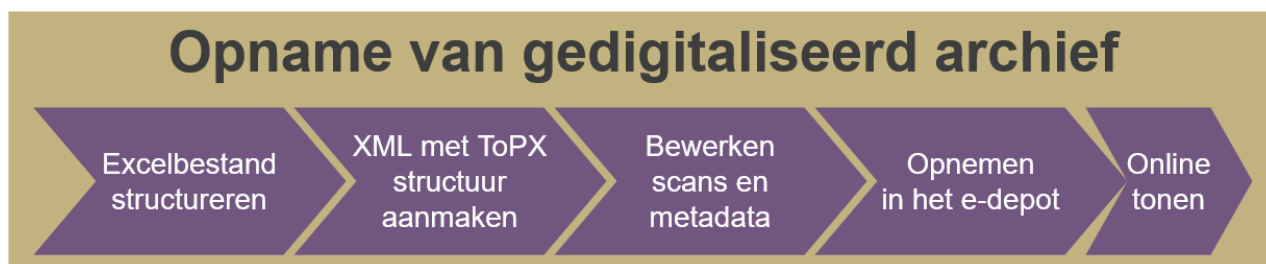
Het belangrijkste doel van deze fase was het technisch werkend krijgen van de opname en het beheer van gedigitaliseerde bouwdoSSIERS. Dit doel is grotendeels behaald, waarbij er in elke stap van het proces een aantal leerpunten naar voren zijn gekomen. Het proces is hieronder in figuur 1 uitgebeeld. Elk van de stappen wordt kort besproken. Voor dit doel hebben we gebruik gemaakt van twee soorten bestanden:

- 10 bouwvergunningen van Wijk bij Duurstede van vóór 1995 die het afgelopen jaar ad-hoc zijn geanonimiseerd door de studiezaalmedewerker van het RAZU;
- Circa 3200 geanonimiseerde bouwvergunningen van Wijk bij Duurstede van na 1995.

Deze laatste, grote, batch hebben we gebruikt om met het verwerken van grote aantallen vergunningen te testen. Aangezien deze doSSIERS nog niet formeel zijn overgebracht zijn ze niet gepubliceerd en worden ze na de pilot weer uit ons systeem verwijderd.

#### *Excelbestanden structureren*

Gemeenten beheren de metadata over bouwdoSSIERS vaak in een Exceldocument, of in een systeem dat deze metadata als Excelbestand kan exporteren. Denk hierbij aan het jaar waarin de vergunning is verstrekt, het doSSIERNUMMER en de inhoudelijke omschrijving van de vergunning. Deze metadata staan in



Figuur 1 - Opname proces voor gedigitaliseerd archief

de Excelvorm los van de scans. Echter, in het e-depot moeten de metadata en de scans aan elkaar gekoppeld worden, zodat de dossiers met bijbehorende scans gevonden kunnen worden. MAIS-Flexis, ons e-depot, neemt in principe CSV- en XML-bestanden aan. In onze pilot is gekozen om gebruik te maken van XML-bestanden die volgens het e-depot format (ToPX) van het Nationaal Archief zijn gestructureerd. Hiermee worden de bouwvergunningen zodanig voorbereid dat ze in principe in elk (in Nederland gebruikt) e-depot kunnen worden opgenomen. Voordat het Excelbestand omgezet kan worden naar een XML-bestand, moet het eerst goed gestructureerd zijn. Dit was een relatief kleine inspanning, maar bevatte wel handmatig werk. Belangrijk voor de toekomst is het vaststellen van een format waaraan Excelbestanden moeten voldoen. Bijvoorbeeld, hoe om te gaan met de meervoudige bouwvergunningen. In het aangeleverde bestand kreeg elk adres een aparte regel. Dit zou betekenen dat een vergunning voor 20 adressen dan ook 20 keer in het e-depot te vinden zou zijn. Dit vinden we onwenselijk. Als 1 dossier geldt voor 20 adressen dan moet het eenmalig geregistreerd worden maar wel vindbaar zijn op al die adressen. Op basis van documentatie vermoeden we dat ons e-depot deze mogelijkheid zou moeten bieden, maar dit is door meerdere technische redenen niet gelukt. We gaan hier zo verder op in.

Daarnaast werden in het aangeleverde Excelbestand alleen de dossiernamen, nummers en adressen geregistreerd, terwijl voor het aanmaken van de dossiers ook de bestandsnamen in het Excel moeten zijn benoemd. Ook hiervoor moesten een aantal handmatige handelingen uitgevoerd worden in het project, die wellicht in de toekomst bij aanlevering volgens een format al gestandaardiseerd kunnen zijn.

#### *XML met ToPX structuur aanmaken*

Voor het aanmaken van een XML-bestand in ToPX-formaat werd de RMTTool van VHIC gebruikt. In principe hoeven de metadata voor MAIS-Flexis niet te voldoen aan de ToPX-structuur, maar het gebruik van de landelijke standaard is duurzaam met het oog op eventuele toekomstige migraties. Het gebruik van de RMTTool is redelijk eenvoudig, aangezien kolommen in het Excelbestand handmatig gekoppeld kunnen worden aan de verschillende elementen van de ToPX-structuur. Vervolgens kunnen alle dossiers en bijbehorende documenten door de tool in één keer voorzien worden van metadata in de vorm van XML-bestanden. Eventuele fouten kunnen door de tool hersteld of overschreven worden. Deze stap

verliep verder zonder problemen.

#### *Bewerking scans*

Voordat de scans worden geïmporteerd in MAIS-Flexis, kunnen deze nog bewerkt worden. Deze bewerkingen worden gedaan op de zogenaamde raadpleegkopieën. De originelen blijven onaangetast. Hierbij kan gedacht worden aan het anonimiseren van persoonsgegevens of het toevoegen van watermerken. De pilot had als doel hiermee te experimenteren en de resultaten hiervan zullen bij het tweede doel worden besproken. Samenvattend blijkt dat het goed mogelijk is scans te bewerken, wat bij kan dragen aan het toekomstig beschikbaar stellen van bouwdoSSIERS zonder in strijd te zijn met de AVG of de Auteurswet. Een vraag die naar voren komt is wanneer de bewerking precies gedaan moet worden en of dit een handeling aan de kant van de gemeente of de archiefdienst moet zijn.

#### *Opname in het e-depot*

Nadat alle voorbereidingen getroffen waren, konden de dossiers worden opgenomen in het e-depot. In MAIS-Flexis betekende dit dat er een zogenaamd verwerkingsmodel opgesteld moest worden. Veel van de stappen van het bestuurlijke proces uit fase 1 konden in dit nieuwe verwerkingsmodel overgenomen worden. Dit maakte dat er snel een nieuw verwerkingsmodel stond en dat de bouwdoSSIERS in het e-depot opgenomen konden worden. We liepen tegen twee technische beperkingen aan op dit vlak. Enerzijds geeft MAIS-Flexis geen of onduidelijke foutmeldingen. Het ontbreken van een (duidelijke) foutmelding maakt het importeren van bestanden, opstellen en aanpassen van verwerkingsmodellen en opnemen van verwerkte bestanden in het e-depot onnodig lastig. Hierdoor was er weinig ruimte voor ideeën voor het verbeteren of uitbreiden van het verwerkingsmodel. Eventueel kan dit probleem worden opgelost met een aanvullende cursus in het maken van verwerkingsmodellen, al zal het deels ook aan het systeem liggen. Een andere, nog onopgeloste beperking, is dat MAIS-Flexis op dit moment niet om weet te gaan met identieke velden. TMLO en ToPX staan voor bepaalde velden meerdere waarden toe, de zogeheten kardinaliteit is hoger dan 1. Zo mogen er meerdere locatiegegevens aanwezig zijn, bijvoorbeeld de plaatsnaam en de straatnaam. Wanneer een metadatabestand met dergelijke meervoudige velden in MAIS-Flexis wordt geïmporteerd dan verdwijnt in het opnameproces één van de locatievelden. Op dit moment gaan we er vanuit dat de fout aan

Weikom / Beschrijven / OTV\_74\_17\_VWK - Gedigitaliseerde bouwvergun... / 404-1969-35\_1-AANVRAAG-EN-VERGUNNING\_Zwartgemaakt.pdf

**Toegang kiezen**

**OTV\_74\_17\_VWK - Gedigitaliseerde bouwvergun...**

- Bouw 44 woningen, Badorfstraat 61, Wijk bij Duurstede
- 404-1969-35\_1-AANVRAAG-EN-VERGUNNING\_Zwartgemaakt.pdf
- 404-1969-35\_2-TEKENINGEN.pdf
- 404-1969-35\_3-OVERIGE-DOCUMENTEN.pdf

**404-1969-35\_1-AANVRAAG-EN-VERGUNNING\_Zwartgemaakt.pdf** [Download](#) [Details...](#) [Galerie](#)

GEMEENTE WIJK BIJ DUURSTEDEN  
BOUWVERGUNNING

Burgemeester en wethouders van Wijk bij Duurstede;  
gelezen het verzoek d.d. 14 maart 1969 van de N.V. Stichting Bouwbedrijf  
te Wijk bij Duurstede om vergunning ingevolge de woningwet tot het bouwen van 44  
woningen op de percelen kadastraal bekend gemeente Wijk bij Duurstede sectie C  
nummers 373, 477, 800 en 442, plaatselijk bekend Frankeweg;  
overwegende, dat het bouwplan in strijd is met het vigerende bestemmings-  
plan (uitbreidingsplan de kadastrale kadastraal bekend gemeente Wijk bij Duurstede sectie C  
nummers 373, 477, 800 en 442, plaatselijk bekend Frankeweg);

**404-1969-35\_1-AANVRAAG-EN-VERGUNNING\_Zwartgemaakt.pdf** [Opslaan](#)

Nummering	Beschrijven	Trefwoorden	Openbaarheid	Instellingen
<b>Nummering</b>				
Voorlopig nummer	0			
Numer				
Code				
<b>Beschrijven</b>				
Naam document	404-1969-35_1-AANVRAAG-EN-VERGUNNING_Zwartgemaakt.pdf			
Datering		Beginjaar		Eindjaar
Bestandsformaat	Portable Document Format 1.4			
Documenttype	AANVRAAG-EN-VERGUNNING			
Status document				

Afbeelding 2 - Een bouwdoossier na succesvolle opname in het e-depot.

de kant van het verwerkingsmodel ligt en niet bij MAIS-Flexis. Door het eerder genoemde probleem van onduidelijke foutmeldingen is het lastig een oplossing te vinden. Dit zal in de toekomst nader worden besproken met De REE, de leverancier van MAIS-Flexis

### Online tonen

Na opname in het e-depot waren dossiers nog niet direct zichtbaar. Het zichtbaar maken, of online tonen, is de laatste stap in het proces. Dit werd gedaan door de laatste stap in het verwerkingsmodel: 'toewijzen aan een toegang'. In de praktijk betekent dit dat er een online toegang aangemaakt moet worden. Deze toegang is een profiel dat bepaalt welke onderdelen van het dossier ook daadwerkelijk weergegeven mogen worden. Bij het inrichten van de toegang kunnen een aantal keuzes gemaakt worden, zoals het afschermen van velden met persoonsgegevens. Ook kan er gekozen worden om informatie enkel te tonen voor bepaalde gebruikersprofielen, zodat deze informatie enkel toegankelijk is onder strenge voorwaarden.

Deze laatste stap hebben we niet succesvol kunnen doorlopen vanwege een mogelijk foutieve inrichting van onze MAIS-Flexis

omgeving. We schrijven "mogelijk", omdat door het gebrek aan foutmeldingen en de zeer versnipperde en onduidelijke documentatie het voor ons lastig is om in te schatten waar de fout ligt. We hebben een afspraak gemaakt met de REE om hier nader op in te gaan, aangezien dit hele proces wel werkend (en voor het RAZU begrijpelijk) moet worden gemaakt voor toekomstige gebruik.

### Ervaring opdoen met het beheer van auteursrechten en persoonsgegevens in gedigitaliseerd archief

De Auteurswet en de Algemene Verordening Gegevensbescherming (AVG) hebben beide gevolgen voor de informatiehuishouding van zowel de gemeente als de archiefdienst. De twee wetten beschermen de rechten van personen en organisaties van enerzijds hun creaties (Auteurswet) en anderzijds hun persoonsgegevens (AVG). In het geval van bouwdoossiers zitten doorgaans in elk dossier tekeningen waarvan vastgesteld moet worden of deze auteursrechtelijk beschermd zijn. Anderzijds zitten er persoonsgegevens van de aanvragers, architecten en eventuele andere betrokkenen in het bouwdoossier. De tweede fase heeft zich vooral gericht op de technische kant van deze vraagstukken, ofwel welke technische

mogelijkheden er zijn voor het waarborgen van de rechten. De ondernomen acties zullen per wet kort worden toegelicht.

#### *Auteurswet*

Wat betreft het auteursrecht heeft de pilot één mogelijke oplossing onderzocht, namelijk het gebruik van een watermerk. Het bleek dat dit watermerk eenvoudig voorafgaand aan de opname in het e-depot kon worden toegevoegd aan de tekening. Op het watermerk wordt weergegeven dat hergebruik van de tekening niet is toegestaan zonder toestemming van de auteur. Hiervoor werd gebruik gemaakt van een gratis proefversie van Adobe Acrobat Pro, waarmee geautomatiseerde verwerkingsstappen konden worden ingericht. Met deze verwerkingsstappen konden de vergunningen, vanwege technische redenen in batches van 200, volledig automatisch worden voorzien van watermerken. Het vraagstuk of het toevoegen van een watermerk voldoende is om de tekening te mogen publiceren, staat nog open. Hiervoor is contact opgenomen met de branchevereniging van archiefinstellingen (KVAN/BRAIN). Ze hadden nog geen antwoord op dit vraagstuk, maar hebben wel een onderzoek naar de verschillende werkwijzen rondom auteursrechtelijk materiaal op de planning voor 2021. Omdat de tekeningen in de bouwvergunningen van onze gemeenten vaak gemaakt zijn door lokale architecten wordt er ook nagedacht over het contact opnemen met architecten(-bureaus) uit de omgeving om het vraagstuk te bespreken.

#### *AVG*

Op het gebied van de AVG zijn verschillende acties ondernomen. Enerzijds kan het voorkomen dat persoonsgegevens worden aangetroffen in de documenten, anderzijds kunnen persoonsgegevens in de metadata van de dossiers en documenten staan. Het vraagstuk omtrent privacy is al opgeworpen in de eerste fase van de pilot en zal hoofdzakelijk worden uitgewerkt in de derde fase van de pilot. Er is daarom op dit moment nog geen eenduidig antwoord op de vraag welke persoonsgegevens verstrekt of gepubliceerd mogen worden. Desondanks konden de technische mogelijkheden wel getest worden.

Een eerste technische oplossing is het anonimiseren van documenten. Het gaat in dit geval om het aflakken van de persoonsgegevens in (kopieën van) de documenten die verstrekt of gepubliceerd worden. Op dit moment is dit een handmatige actie die veel tijd vergt. Mogelijke oplossingen zijn het uit laten voeren van anonimiseren door gespecialiseerde bedrijven of het gebruik van anonimiseringssoftware.

Ten tweede is er onderzocht of metadatavelden van de documenten en dossiers afgeschermd konden worden. Op het moment van schrijven is het nog niet volledig duidelijk hoe dit technisch werkt in MAIS-Flexis, maar dat er technische mogelijkheden bestaan in het systeem is duidelijk.

Tot slot biedt MAIS-Flexis de mogelijkheid om openbaarheidsbeperkingen toe te voegen. In dit geval worden de metadata wel weergegeven, maar de documenten niet. Deze openbaarheidsbeperkingen worden door middel van een datum in het systeem gezet, waardoor het duidelijk is op welk moment het dossier volledig openbaar wordt.

#### **Ervaring opdoen met het verrijken van metadata**

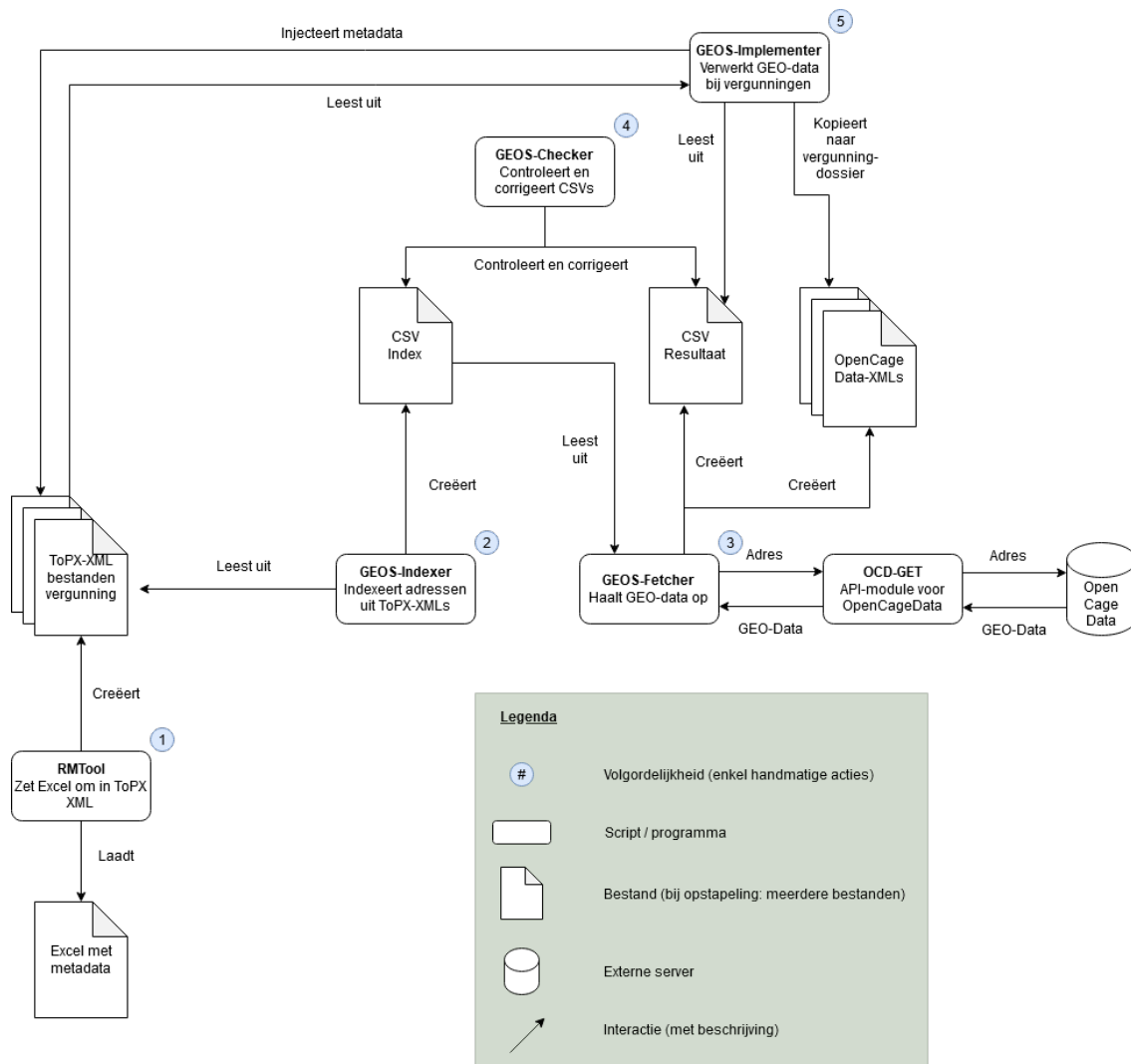
Tijdens de wekelijkse overleggen van de projectgroep werd het idee geopperd om te kijken of we de metadata van bouwvergunningen ook zouden kunnen verrijken. Onder verrijken verstaan we het toevoegen of aanpassen van metadata zodat een breder gebruik van de metadata of vergunningen mogelijk wordt. Op nationaal niveau zijn er bijvoorbeeld al experimenten met het automatisch herkennen van persoonsnamen in gescande documenten. Dit kan van pas komen voor het beschermen van persoonsgegevens (AVG) of in het identificeren van architecten op bouwvergunningen zodat hun auteursrechten beter kunnen worden beschermd. Deze technologie is echter nog té experimenteel om in deze fase te testen. Wat beter binnen bereik leek te liggen was het automatisch toevoegen van geografische metadata, de lengte- en breedtegraden. Met geografische metadata kan een vergunning gekoppeld worden aan een locatie op een kaart. In een eerste verkenning bleek dit op twee manieren te kunnen: een uitdraai uit de BAG of het Kadaster, bijvoorbeeld door RHC Overijssel toegepast; of op basis van het geocoderen van een adres. Het eerste leek, wederom vanwege de omvang, niet passend voor deze fase. Geocoderen, dat is het opvragen van geografische metadata bij een database via een API, was de meest kansrijke optie. Omdat het verrijken van metadata niet in de planning of doelen van fase 2 was opgenomen is dit onderdeel grotendeels in vrije tijd opgepakt. Toen dit bleek te werken is er ook binnen het project aandacht aan besteed. Dit heeft tot twee resultaten geleid: het GEOS-scriptpakket en kennis met het toepassen van geo-data in MAIS-Flexis.



## GEOS-Scriptpakket

Door het toepassen van de RMTTool van VHIC werd er bij elke vergunning een XML-bestand toegevoegd waarin de basis-metadata, waaronder het adres, was opgenomen. In fase 1 is ervaring opgedaan met het openen en bewerken van XML-bestanden via Powershell, een computerscripttaal van Microsoft. Een automatische extractie van adressen via Powershell lag voor de hand. Voordat dit in gang werd gezet is gezocht naar een API waarmee de geografische data konden worden opgehaald. We vonden hier twee opties voor, Google Maps en OpenCageData. Google kon óf via MAIS-Flexis, óf via een eigen script worden opgeroepen. Google is echter redelijk prijzig en bood maar beperkte metadata, waardoor ook gekeken is naar een alternatief. OpenCageData, een Duitse start-up, bood een goed alternatief dankzij goede documentatie, veel rijkere aanlevering van metadata en lagere kosten. Binnen de pilot konden we gebruikmaken van een gratis ontwikkelingslicentie, waardoor in fase 2 geen kosten zijn gemaakt voor het geocoderen.

Met een idee voor een script en een goede API is gestart met de ontwikkeling van een aantal scripts, samen GEOS genoemd: Geografische Scripts. De werking van de scripts is hieronder schematisch weergegeven. In het kort komt het er op neer dat het eerste script (Indexer) de metadatabestanden van bouwvergunningen zoekt, uitleest en de adressen uit de XML-bestanden als lijst in een CSV-bestand zet. Het tweede script (Fetcher) leest dit CSV-bestand uit en communiceert via een tussenscript (OCD-Get) met OpenCageData. Het zet de belangrijkste metadata van OpenCageData in een nieuw CSV en slaat het ruwe resultaat van elk adres op als XML in een centrale map. Het derde script (Implementer) kijkt vervolgens weer naar de vergunningen. Daarbij leest het de metadata-XML's uit en zoekt de verrijkte metadata bij elkaar op basis van een adres binnen de CSV-bestanden en OpenCageData-XML-bestanden. Dit zet het script vervolgens bij de vergunning neer en injecteert het voor een deel ook in de bestaande metadata van de vergunning. Omdat de originele metadata niet altijd even 'schoon' bleken te zijn is er een vierde script (Checker) geschreven dat fouten van de eerste twee



Figuur 2 - Overzicht van het GEOS-Scriptpakket

scripts kan terugdraaien. Denk bijvoorbeeld aan een adres als 'Karel de Grotestraat 30 (nabij)'. Dit adres wordt door de eerste twee scripts wel verwerkt, maar omdat OpenCageData niet om kan gaan met een toevoeging als '(nabij)' zal dit geen bruikbare metadata opleveren. Het vierde script markeert dit adres voor menselijke interventie en haalt het uit de CSV-bestanden.

Door deze scripts op een dergelijke wijze uit elkaar te houden kan de geocoding op een later moment gemakkelijker worden hergebruikt. We zien een grote meerwaarde in het toepassen van GEOS voor de collectie van het RAZU, maar kunnen ons ook een inzet van de scripts binnen de gemeente voorstellen. GEOS zal net als SPIRA na afronding als open source op GitHub worden gepubliceerd, zodat het publiekelijk kan worden hergebruikt. Hierbij tekenen we wel aan dat het RAZU geen ondersteuning zal bieden aan gebruikers (buiten onze deelnemers) en doorontwikkeling enkel zal plaatsvinden op basis van wensen van het RAZU en zijn deelnemers.

### *Geo-data in MAIS-Flexis*

MAIS-Flexis biedt de mogelijkheid om geografische metadata, en dan specifiek de lengte- en breedtegraad, in speciale velden op te nemen. Door de metadata hierin op te nemen kan MAIS-Flexis een kaart weergeven met daarop markeringen voor vergunningen. Dit kan op het niveau van individuele bouwvergunningen, maar ook door middel van speciale adres-registraties. Op die wijze kan er gewerkt worden met enkelvoudige registratie van adressen, die vervolgens in meerdere vergunningen gerefereerd kunnen worden. Ook wordt het meervoudige adres bij een vergunning, bijvoorbeeld voor een blok rijtjeshuizen, hierdoor gemakkelijker en overzichtelijker.

Dankzij de GEOS-scripts was het voor ons mogelijk om de geografische data op grote schaal toe te passen en klaar te zetten voor het importeren in MAIS-Flexis. Hier liepen we echter tegen een aantal beperkingen aan, die ook al eerder in dit rapport zijn beschreven. Zo kon MAIS-Flexis niet omgaan met de volgens de ToPX-gestructureerde XML-bestanden waarin meer dan een enkel adres was opgenomen. Hierdoor hebben we GEOS-Implementer moeten aanpassen zodat de XML-bestanden nu afwijken van de ToPX, maar wel te verwerken zijn door MAIS-Flexis. Voor duurzaam hergebruik is dit onwenselijk, maar op dit moment onvermijdelijk. Vervolgens liepen we er tegen aan dat het model voor de speciale adres-registraties in onze MAIS-Flexis-omgeving niet goed is ingericht. Dit bleek binnen fase 2 niet op

te lossen te zijn. Uiteindelijk is, zoals beschreven in Online tonen, de volledige doorvoer van onze geo-data nog niet voltooid, waarschijnlijk vanwege een (deels) foutieve inrichting van onze MAIS-Flexis omgeving.



## Conclusie

Op basis van de bevindingen concluderen we dat beide doelen uit het PID grotendeels zijn behaald. Het eerste doel beoogde het totale proces in kaart te brengen voor het opnemen en beheren van de bouwdoSSIERS. Het tweede doel beoogde kennis op te doen van de technische mogelijkheden om de auteurs- en persoonsrechten te waarborgen. Ten derde is er tijdens de pilot het aanvullende doel gesteld om ervaring op te doen met het verrijken van metadata, wat praktisch neer kwam op het toevoegen van geografische metadata aan de doSSIERS. Alle doelen zijn grotendeels bereikt. Het bleek in enkele gevallen nog niet mogelijk het goed in te richten aan de kant van MAIS-Flexis. Hiervoor is contact opgenomen met de leverancier. De verwachting is dat het meeste op te lossen is.

De derde en laatste fase van de pilot neemt het hele jaar 2021 in beslag. In het rapport van fase één werd er in de vooruitblik al aangegeven dat niet alle op voorhand gestelde doelen behaald kunnen worden in de derde fase. De technische beperkingen uit fase één en twee leiden er toe dat er minder aandacht gegeven zal worden aan het technische aspect van het overbrengen, zoals het opstellen van handleidingen en protocollen. Er zal daarentegen veel nadruk

gelegd worden op de afstemming met alle gemeenten. In de eerste maanden wordt er met de gemeenten onderzocht wat zij verstaan onder het 'digitale object'. In andere woorden: als er digitale doSSIERS overgebracht moeten worden, wat zijn de minimale kenmerken van dat dossier die mee moeten komen naar het e-depot om het dossier 'authentiek' te kunnen noemen? Dit vraagstuk is aangekaart in het rapport van fase één. Daarnaast zal het wetgevingsvraagstuk een belangrijk thema vormen in fase 3. Er zullen visies en werkwijzen omtrent de AVG en Auteurswet opgesteld moeten worden, ook dit wordt opgepakt met de gemeenten. Ten derde zal er een visie worden opgesteld voor het toekomstige e-depot, waarbij aandacht is voor bewaren bij de bron en common ground. Hier hoort ook een lijst van eisen en wensen bij voor een e-depot dat wél aan onze verwachtingen voldoet. Of dit door de markt geleverd kan worden zal moeten worden bekeken. In het eindrapport van fase 3 wordt daarmee niet alleen ingegaan op de behaalde doelen, maar zal tevens een advies voor de toekomst worden gegeven. Op basis hiervan moeten de gemeenten en het RAZU bepalen hoe er verder gegaan wordt in de toekomst.

*Afbeelding 3 - Een bericht van het heden aan de toekomst: Het begraven van een tijdsapsule op de bouwplaats van de archiefbewaarplaats van het RAZU in 2010. Foto: W. Leeman, 2010.*

