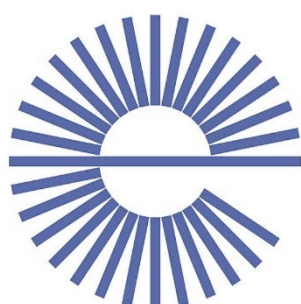


Beschrijving van de praktijkervaringen
van de implementatie van
Persistent Identifiers (ARK)
bij de
Gouda Tijdmachine (Omeka S)

Opgesteld in het kader van het Netwerk Digitaal Erfgoed



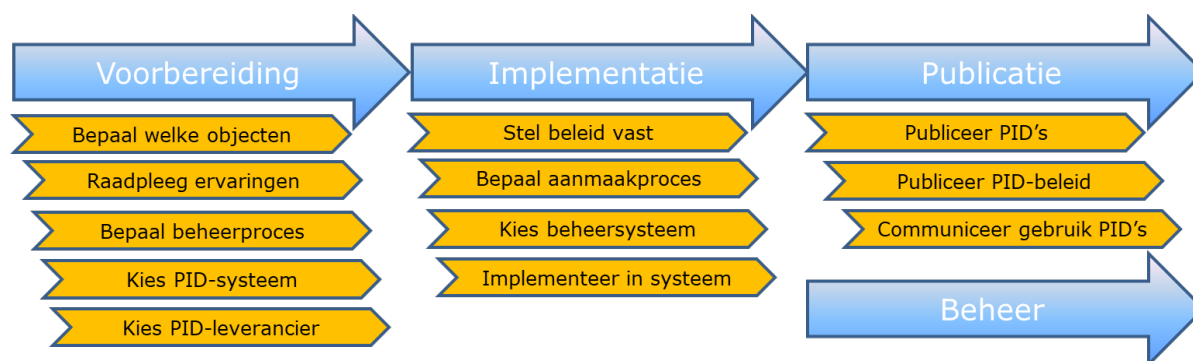
**netwerk
digitaal
erfgoed**

1 Inleiding

Als erfgoedinstelling heeft u ervoor gezorgd (of zorgt u er momenteel voor) dat uw digitale erfgoed vindbaar blijft. U gebruikt hiervoor Persistent Identifiers (PID's). Een PID is een unieke identificatiecode van een digitaal object, die op een afgesproken plaats wordt geregistreerd. Hij blijft gegarandeerd werken, ook al verandert het webadres van een organisatie. Zo zorgt u ervoor dat de verwijzing naar een digitaal object (een scan, audiovisueel bestand, metadata record, etc.) altijd blijft werken. Ook als de bewaarlocatie of het webadres van het digitale object verandert.

Via deze good-practice beschrijving stimuleert het project Persistent Identifiers van het Netwerk Digitaal Erfgoed het delen van praktijkervaring m.b.t. het implementeren van PID's. Qua vorm en inhoud sluit de bestpracticebeschrijving zoveel mogelijk aan op het "Stappenplan voor implementatie van Persistent Identifiers", opgesteld in het kader van hetzelfde project. Deze uniformiteit zorgt ervoor dat het stappenplan en de bestpracticebeschrijvingen onderlinge samenhang vertonen en vergelijkbaar zijn.

De volgende hoofdstappen en deelaspecten worden in het stappenplan beschreven (zie het stappenplan voor de precieze beschrijvingen) en daarom ook in de bestpracticebeschrijving gebruikt:



2 Organisatie en contactpersoon

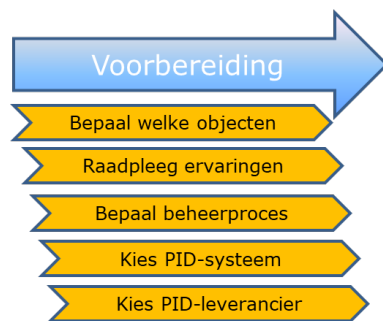
Het project Gouda Tijdmachine wil uit de enorme hoeveelheid informatie de ontwikkeling van de stad volledig reconstrueren. Alle informatie over gebouwen, wegen, water, bedrijven en bewoning moet in kaart worden gebracht.

De Gouda Tijdmachine is een werkgroep van de Historische Vereniging Die Goude in samenwerking met het Streekarchief Midden-Holland. Die Goude is organisatorisch bij het project betrokken. Streekarchief Midden-Holland stelt haar bronnen en indexen

beschikbaar voor het project en stuurt de vrijwilligers aan die op haar locatie in de Chocoladefabriek werken met de originele bronnen.

Binnen de Gouda Tijdmaschine houdt het IT team zich bezig met de infrastructuur, waaronder ook de PID's. Aanspreekpunt voor dit onderdeel is Bob Coret.

3 Voorbereiding



3.1 Welke objecten?

De Gouda Tijdmaschine maakt gebruik van het open-source semantische collectie beheer systeem Omeka S¹. Via Omeka S wordt er informatie vastgelegd over “items” als straten, wijken, panden, percelen, adressen, gebouwen, bruggen, waterlopen, sluizen, gemalen, personen en persoonsvermeldingen, organisaties, plaatsen, archieftoegangen, datasets (en distributies) boeken, artikels en afbeeldingen (kaarten, foto's, scans).

Een object bestaat uit meta data en nul, één of meer media bestanden. Deze media bestanden krijgen geen eigen PID.

3.2 PID-beheerproces

Om als organisatie eigen PID's te maken op basis van ARK dient er eerst een “Name Assigning Authority Number (NAAN)” aangevraagd te worden. Op basis van informatie over de organisatie en de belofte om zorg te dragen voor de duurzaamheid, krijgt de organisatie een nummer toegewezen. Dit nummer maakt onderdeel uit van de ARK die als structuur heeft `ark:{NAAN}/{assigned name}`.

De gratis registratie van de NAAN zorgt ervoor dat ARK's via de Name 2 Things (N2T) revolver² worden verwezen naar de in de registratie opgegeven website. Het verder

¹ <https://omeka.org/s/>

² <https://n2t.net/>

verwijzen naar het juiste object (op basis van de {assigned name}) is de eigen verantwoordelijkheid. Voor de Gouda Tijdmachine zorgt een lokale resolver dat het juiste object (in het gevraagde) formaat wordt geserveerd.

Omeka S maakt automatisch voor elke item en item-set een ARK identifier aan. Geconfigureerd kan worden in welk property deze waarde wordt opgeslagen. De Gouda Tijdmachine heeft gekozen voor de property `det erms : i dent i fi er`.

3.3 PID-systeem

De keuze voor een PID-systeem werd deels beïnvloed door de beschikbare ondersteuning (via separate open-source modules) van PID-systemen.

De Ark & Noid module³ is een module voor het maken en beheren van ARK identifiers. De Persistent Identifiers module⁴ biedt soortgelijke functionaliteit en zijn gericht op de PID services EZID (ARK) en DataCite (DOI). Er is gekozen voor de Ark & Noid module om niet afhankelijk te zijn van PID-services maar het geheel in de Gouda Tijdmachine infrastructuur op te lossen (waartoe ook de benodigde technische kennis voor aanwezig was).

3.4 PID-leverancier

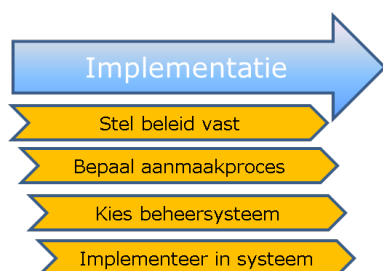
De “Name Assigning Authority Number” wordt uitgegeven door de ARK Alliance. De ARK Alliance is ontstaan uit een samenwerking tussen de California Digital Library en DuraSpace/LYRASIS. Het werd gelanceerd in 2018 en kreeg de steun van 43 instellingen. In 2021 werd de ARK Alliance een trots lid van de NDSA, een internationaal consortium dat zich inzet voor het langdurig bewaren van digitale informatie.

Het {assigned name} deel van de ARK wordt lokaal aangemaakt en opgeslagen bij de metadata van de objecten. Er is

³ <https://github.com/Daniel-KM/Omeka-S-module-Ark>

⁴ <https://omeka.org/s/modules/PersistentIdentifiers/>

4 Implementatie



4.1 PID-beleid

Bij het aanvragen van een NAAN bij de ARK Alliance dient er er commitment afgegeven te worden voor data persistentie:

“Does the memory organization commit to data persistence? In assigning ARKs to data, the memory organization, or a service provider acting on its behalf, agrees to make a best effort to store and manage that data and any core metadata such that persistent access to it can be provided. (The NAAN holder is encouraged but not required to make all data and metadata publicly available. Sample policies are linked from the ARKs FAQ.)”

De Gouda Tijdmachine, dus het Streekarchief Midden-Holland en historische vereniging Die Goude, hebben dit commitment gegeven en geborgd in de organisatie.

4.2 PID's aanmaken

De Gouda Tijdmachine kent (geautomateerd) identifiers toe binnen het ARK-domein onder de NAAN 60537 volgens de volgende principes:

- Er wordt geen ARK opnieuw toegewezen; dat wil zeggen, zodra een ARK-object-associatie openbaar is gemaakt, zal die associatie voor onbepaalde tijd als uniek worden beschouwd.
- Het {assigned name} gedeelte van de door de Gouda Tijdmachine toegewezen ARK's bevat geen algemeen herkenbare semantische informatie.
- De door de Gouda Tijdmachine gegeneerde ARK's bevatten een “terminal-control” karakter welke herkenning van enkelvoudige karakterfouten en omzettingfouten mogelijk maakt.

4.3 PID-beheersysteem

Het {assigned name} deel van de ARK wordt automatische, lokaal toegekend via de Omeka S module ARK & NOID (de “minter”).

4.4 Implementatie

Het gebruik van ARK is vrij laagdrempelig doordat zowel de PID's (ARK's) als PID-systeem (Omeka S + Ark & Noid module) gratis zijn te gebruiken. Bij de Gouda Tijdmachine was technische kennis voor handen om dit in te richten (er is geen externe leverancier gebruikt, behalve dan de Omeka S en ARK communities).

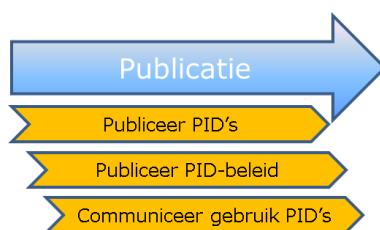
Enige technische kennis was nodig om de lokale ARK resolver te realiseren die werkt wanneer er gebruik wordt gemaakt van meerdere Omeka S subsites en content-negotiation.

Wanneer er nu in de browser via de N2T resolver het object met PID `ark:/60537/bumH54` wordt opgevraagd, dus <https://n2t.net/ark:/60537/bumH54> dan zal dit uitkomen op de pagina <https://www.goudatijdmachine.nl/data/document/ark:/60537/bumH54>

Wanneer er een machine-leesbare RDF representatie van hetzelfde object gewenst is kan dit als volgt worden opgevraagd:

```
curl -L -H "Accept: text/turtle" https://n2t.net/ark:/60537/bumH54
```

5 Publicatie



5.1 Publiceer PID's

De Gouda Tijdmachine heeft via Omeka S ook een publicatiemiddel om alle objecten online toegankelijk te maken. In de web-interface wordt op elke item en collectie pagina PID getoond. Dit is tevens een link naar het object via de N2T resolver.

Omeka S is een semantisch collectie beheersysteem dat via de API de meta data in JSON-LD toont, inclusief de PID via de `dcterms:identificer` property.

Alle data wordt ook als linked open data gepubliceerd via een triplestore (GraphDB). Waar elke resource een identifier heeft (bijv.

<https://www.goudatijdmachine.nl/data/api/items/17782>) wordt via een `owl:sameAs`

relatie de link met de ARK-via-resolver gelegd (in dit voorbeeld: <https://n2t.net/ark:/60537/bumH54>).

Wanneer de data via SPARQL wordt bevraagd worden de ARK-via-resolver als identifiers gebruikt. Een toepassing die de triplestore van de Gouda Tijdmachine gebruikt is het NDE Termennetwerk⁵. Wanneer er via het Termennetwerk wordt gezocht naar het “Achter de Vismarkt” in de “Straten in Gouda” bron dan is het resultaat de URI <https://n2t.net/ark:/60537/b74Nrs>.

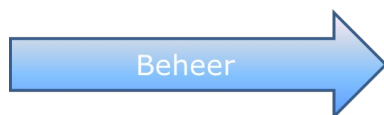
5.2 Publiceer PID-beleid

Het PID-beleid wordt op de voor ARK standaard wijze gepubliceerd, maar is nog minimaal. Daarnaast is het gebruik beschreven voor de gebruikers van het collectiebeheersysteem. Waar het Streekarchief Midden-Holland al wel bekend zijn met PIDs (maar nog geen eigen PID-systeem hebben geïmplementeerd) is dit voor de historische vereniging een lastiger onderwerp.

5.3 Communiceer PID-gebruik

Het PID-gebruik, de introductie van ARK, is niet los gecommuniceerd. Het was onderdeel van de introductie van Omeka S als semantisch collectiebeheersysteem.

6 Beheer



Er kunnen wel items verwijderd worden uit het collectiebeheersysteem. Een ‘tombstone’ pagina voor dergelijke objecten om aan te geven dat het object bestaan heeft maar is verwijderd, is op dit moment nog niet beschikbaar via Omeka S.

⁵ <https://termennetwerk.netwerkdigitaalergoed.nl/>

7 Wie heb je nodig?

De organisatie dient bewust te zijn van de noodzaak van persistente identifiers en hiervoor de organisatie en de techniek in te richten.

Aan de aanvraag van een NAAN bij de ARK Alliance zijn geen kosten gemoeid, wel dient (er door management) een belofte gedaan te worden om zorg te dragen voor duurzame toegankelijkheid.

Wanneer er gebruik wordt gemaakt van Omeka S als collectiebeheersysteem is het de applicatiebeheerder die simpel wel de ARK module installeert en configureert.

Het aanmaken van PID's voor elk item gebeurt nu automatisch. Gebruikers van het collectiebeheersysteem en ook eindgebruikers dienen voorgelicht te worden over nut en noodzaak van deze PID's.

Met name voor eindgebruikers helpt het als in de template van de website een ARK als een link wordt gepresenteerd via de N2T resolver.

Voor de LOD-gebruikers kan het helpen om de PID (met resolver) te gebruiken als identificatie voor alle resources, bijvoorbeeld via een owl:sameAs relatie.