

DE BASIS – Vindbaarheid

Basisnormen ICT voor Nederlandse
cultureel erfgoedinstellingen



DEN-rapport
28 januari 2008

INHOUDSOPGAVE

Inleiding

Deel 1 Het project

- Doel
- Achtergronden
- Methodiek
- Toepassingsgebieden

Deel 2 DE BASIS

- DE BASIS vindbaarheid: overzicht
- DE BASIS vindbaarheid
- Standaarden en begrippen
 - Interoperabiliteit
 - Webdiensten
 - HTTP
 - XML
 - UTF-8
 - Toegang
 - Harvesting
 - OAI-PMH
 - Gedistribueerd zoeken
 - SRU
 - Beschrijving
 - Dublin Core
 - Identificatie
 - URI
 - Persistentie
 - Resolving
- Scenario

Bijlage 1: De totstandkoming van DE BASIS

- Inventarisatieronde experts
- Brede inspraakronde door middel van RFC procedure
- Wiki.den.nl
- Analyseperiode toetsing DE BASIS in het veld
- Vaststelling en publicatie

Bijlage 2: Samenvatting van de analyseperiode: DE BASIS en de huidige praktijk

- Informatiebeleid
- Vindbaarheid
- Standaardisatie en normering
- Beschrijving en Dublin Core
- Identificatie
- Toegang en interoperabiliteit
- Mogelijke aanvullingen
- De wiki
- Tot slot

INLEIDING

In 2007 is DEN het project DE BASIS (Digitaal Erfgoed: Bouwen aan een Succesvolle ICT Strategie) gestart, om te komen tot een set basiseisen die gesteld kunnen worden aan alle Nederlandse cultureel erfgoedinstellingen. Het betreft Erfgoedinstellingen die ICT willen inzetten voor hun dienstverlening of deel willen nemen aan een ICT-samenwerkingsproject.

Erfgoedinstellingen die voldoen aan de basiseisen zijn ervan verzekerd dat de wijze waarop zij ICT inzetten effectief en toekomstgericht is. Bovendien is het een garantie dat de digitale diensten van deze ICT-projecten gebaseerd zijn op internationale standaarden en dat ze in staat zijn te communiceren (interoperabiliteit) met andere diensten.

Dit document beschrijft het eerste resultaat, DE BASIS Vindbaarheid. In 2007 is Prioriteit gegeven aan dit thema. In deze eindrapportage is het proces van totstandkoming, een uitgebreide beschrijving van de standaarden die geselecteerd zijn voor DE BASIS en een uitwerking in een scenario opgenomen. Deze eindrapportage is bedoeld als verantwoording van het proces en brengt alle informatie over het project bijeen.

In 2008 wordt DE BASIS voor de thema's Vervaardiging en Duurzaamheid verder uitgewerkt.

Digitaal Erfgoed Nederland (DEN)

Annelies van Nispen

Marco Streefkerk

DEEL 1: HET PROJECT

Doel

Het doel van dit project is het vaststellen en publiceren van een set ICT-kwaliteitsinstrumenten die als basisnorm geldt bij digitalisering (in brede zin) van cultureel erfgoed. In het proces van normering staat het **belang van het publiek** voorop. Om die reden heeft in de eerste fase het thema **Vindbaarheid** prioriteit gekregen.

De vaststelling vindt plaats op basis van een inventarisatie onder experts, gevolgd door een brede inspraak- en testprocedure, waarbij alle erfgoedsectoren actief betrokken worden. De genormeerde instrumenten worden als zodanig gemarkeerd in het ICT-register voor het Cultureel Erfgoed

Achtergronden

"Cultureel erfgoed is ons maatschappelijk kapitaal, ons collectieve geheugen. Het is een onschatbare bron van kennis, creativiteit en identiteit. Digitalisering maakt deze bron toegankelijk voor iedereen en maakt nieuwe toepassingen mogelijk. Daarom bouwen wij aan een Digitale Collectie Nederland.", aldus het ministerie van OCW¹. Om de totstandkoming van die Digitale Collectie Nederland te ondersteunen, heeft OCW de opdracht aan DEN gegeven een kwaliteitszorgsysteem op te zetten. Dit kwaliteitszorgsysteem moet alle erfgoedinstellingen in Nederland in staat stellen op een goede wijze ICT in te zetten om de collecties te digitaliseren. Op deze wijze moeten de erfgoedinstellingen in staat zijn te participeren in de Digitale Collectie Nederland.

Het kwaliteitszorgsysteem is tevens een platform om kennis over ICT te verwerven en te delen. Onderdeel van de opdracht van het ministerie aan DEN is om de normering van de kwaliteitsinstrumenten op basis van zelfregulering tot stand te brengen. Het vaststellen van DE BASIS, zoals uiteengezet in dit document, is de eerste stap op deze weg. In plaats van individuele instrumenten in willekeurige volgorde te onderwerpen aan een normeringstraject wordt allereerst gekeken welke instrumenten in de ogen van de erfgoedsector zelf door iedere instelling, klein of groot, zouden moeten worden toegepast.

Door snelle veranderingen in de wereld van de ICT kan deze normering van ICT-instrumenten geen eeuwigheidswaarde hebben. DE BASIS moet voeling houden met de ontwikkelingen: als een instrument verouderd raakt of als een ander instrument met vergelijkbare functionaliteit in de praktijk de overhand neemt, wordt de set hierop bijgesteld.

Methodiek

De totstandkoming van DE BASIS is een zorgvuldig proces geweest, waarbij DEN heeft samengewerkt met experts en instellingen uit de erfgoedsector.

Het proces bestond uit:

- een inventarisatieronde bij experts;
- een brede inspraakronde gebaseerd op de RFC-procedure 2026;
- de toetsing in het veld;
- definitieve vaststelling en publicatie.

In de bijlage is een uitgebreide beschrijving opgenomen van de totstandkoming van DE BASIS Vindbaarheid.

¹ <http://www.digitaliseringerfgoed.info/>

Toepassingsgebieden

Het besluit DE BASIS vast te stellen komt niet voort uit een behoefte om de instellingen een keurslijf op te leggen. DE BASIS moet veeleer gezien worden als een veilig vertrekpunt voor verdere ontwikkeling, als een veilig fundament onder de digitaliseringsactiviteiten van cultureel erfgoedinstellingen. DE BASIS kan in de volgende situaties ingezet worden:

- **Beleidsvorming**

Instellingen die nog geen specifiek informatiebeleid geformuleerd hebben kunnen DE BASIS gebruiken als hulpmiddel bij het opstellen van beleidsdocumenten, zoals een informatiebeleidsplan of een digitaliseringsplan. Door DE BASIS in de discussies over het toekomstige beleid te betrekken, wordt er een fundament gelegd voor kwaliteitszorg voor ICT.

- **Systeemontwikkeling**

DE BASIS moet worden opgenomen in de functionele en technische ontwerpen voor ICT-systemen, die in opdracht van de erfgoedinstellingen worden ontwikkeld of aangepast (bijv. collectieregistratiesystemen of websites). DE BASIS is een goed ijkpunt om de diensten van commerciële dienstverleners in de erfgoed- en informatiesector met elkaar te vergelijken en om te onderhandelen met deze partijen.

- **Samenwerkingsverbanden**

DE BASIS garandeert dat de (meta)data van de instelling aan internationale standaarden voldoen en eenvoudig uitwisselbaar zijn met de data van andere instellingen ('interoperabiliteit'). DE BASIS is ook een meetlat voor de erfgoedportals waar de instelling aan zou kunnen deelnemen. DE BASIS helpt de instellingen beter in te schatten hoe groot de inspanningen zullen zijn om deel te nemen aan landelijke en/of internationale erfgoedportals. Het resultaat is een grotere zichtbaarheid van de eigen collectie.

- **Subsidieaanvragen**

In de meeste subsidieregelingen waarin ICT een rol speelt, wordt de nadruk gelegd op innovatie en interoperabiliteit. Welke criteria daarbij gehanteerd worden, is vaak niet duidelijk. DE BASIS helpt om de criteria voor projectaanvragen te verduidelijken. Duidelijkheid over welke normen er vanuit het veld bepaald worden, helpt de opsteller van een aanvraag met het beschrijven van de uitgangspunten. De subsidiegever heeft met DE BASIS een meetlat in handen om te kijken in hoeverre de aanvragen aan de basis goed zitten.

- **Evaluatie van dienstverlening**

Met het toenemend belang van digitale diensten (bijvoorbeeld de website en collectiedatabases) van de erfgoedinstellingen, is het nuttig die diensten periodiek door te lichten. Welk publiek wordt ermee bereikt? Hoe is de gebruiksvriendelijkheid? Hoe toekomstvast is de dienst? DE BASIS helpt om op enkele elementaire gebieden de digitale diensten door te lichten en met andere te vergelijken.

- **Effectieve en duurzame digitalisering**

Het gebruik van standaarden uit DE BASIS zorgt voor een efficiënte en duurzame digitalisering van de dienstverlening, doordat State-of-the-Art en (inter)nationaal toegepaste standaarden worden gebruikt die samenwerking bevorderen.

Het grotere verband

Het kunnen beschikken over een set normen die op basis van zelfregulering tot stand is gekomen is een belangrijke stap voorwaarts naar een kwalitatief betere inzet van ICT door Nederlandse erfgoedinstellingen.

Dit is een goede opstap naar de totstandkoming van de Digitale Collectie Nederland. Maar: DE BASIS is slechts één van de bouwstenen hiervoor.

Professionalisering van de inzet van ICT door erfgoedinstellingen vergt meer dan een set afspraken over instrumenten die hun nut kunnen bewijzen bij het uitbouwen van digitale dienstverlening. Het vergt ook afspraken op hoger niveau, beleid op het gebied van de kennisinfrastructuur, technische infrastructuur en vooral ook architectuur. Het is de missie van Digitaal Erfgoed Nederland met haar kennisplatform een bijdrage te leveren aan kwaliteitsverbetering van de digitalisering in het cultureel erfgoed.

DE BASIS Vindbaarheid

Vindbaarheid wordt gegarandeerd door tijdens het digitaliseerproces te voldoen aan minimale eisen op het gebied van:

Interoperabiliteit: de informatie is beschikbaar via webdiensten

- over **HTTP**,
- verpakt in **XML**,
- gecodeerd in **UTF-8**.

Toegang: de informatie is beschikbaar voor

- harvesting door ondersteuning van **OAI-PMH** en/of
- gedistribueerd zoeken door ondersteuning van **SRU**.

Beschrijving: de informatie is beschikbaar als

- metadata in **Dublin Core** als structuur met
- identificatie van de bron in **URI**.

DEEL 2: DE BASIS VINDBAARHEID

DEN heeft er, in samenspraak met de Redactie van het ICT-register, voor gekozen om het thema Vindbaarheid als eerste uit te werken. Het normeren van alle fasen van de digitale levenscyclus was gezien tijd en menskracht niet haalbaar in 2007. Het thema Vindbaarheid heeft prioriteit gekregen, omdat:

- De gebruiker centraal staat;
- Zichtbaarheid en gebruik van digitaal erfgoed het uiteindelijke doel is van digitalisering van collecties.

Vindbaarheid

Vindbaarheid is het digitaal beschikbaar stellen van informatie over erfgoed op basis van gangbare technologie, zodanig dat deze zonder kennis vooraf kan worden hergebruikt door zowel mens als machine.

DE BASIS concentreert zich op vindbaarheid door machines. Dit is belangrijk voor de huidige dienstverlening/eindgebruikerstoepassingen, maar ook voor vindbaarheid in zoekmachines als bijvoorbeeld Google en voor deelname aan samenwerkingsprojecten als erfgoedportals.

De toegankelijkheid van informatie over erfgoedobjecten zal hergebruik stimuleren. Dat hergebruik zal weer resulteren in grotere toegankelijkheid van het (digitale) erfgoed. Blijvende vindbaarheid en toegankelijkheid van digitaal erfgoed rechtvaardigt de gedane investeringen. Erfgoed dat toegankelijk en herbruikbaar is, vormt nu en in de toekomst een blijvende basis voor (publieks)diensten van de beherende instelling en haar samenwerkingspartners.

Het gebruik van de genormeerde standaarden is een garantie dat de informatie interoperabel is en dat deze informatie (her)gebruikt kan worden voor diensten van derden. Bijvoorbeeld groepen samenwerkende instellingen, centrale portals of (commerciële) marktpartijen, zoals Google kunnen de informatie automatisch ophalen en toepassen in hun eigen diensten. Dergelijke diensten geven het publiek meer zicht op de Digitale Collectie Nederland en op de samenhang daarbinnen. Voor de erfgoedbeheerders biedt elke dienst een potentieel publiekscontact met hun collectie.

In de huidige praktijk realiseren erfgoedinstellingen vindbaarheid van gedigitaliseerd erfgoed veelal via specifieke (eindgebruikers)diensten op een verder gesloten website. DE BASIS ziet blijvende vindbaarheid als een zelfstandig doel, los van de actuele doelgerichte dienst waarvoor op dat moment gedigitaliseerd wordt. De informatie over erfgoedobjecten dient blijvend vindbaar te worden gemaakt door de erfgoedinstelling zelf of via sectoroverstijgende portals. De metadata bevatten minimaal een doorverwijzing naar het (digitale) erfgoed bij de beherende instelling.

Erfgoedinstellingen die voldoen aan deze normen hebben de zekerheid een goede, blijvende contentleverancier van erfgoed te zijn; voor hun eigen dienstverlening, voor samenwerkingsprojecten en voor de Digitale Collectie Nederland.

DE STANDAARDEN en BEGRIPPEN

Interoperabiliteit: de informatie is beschikbaar via **webdiensten**

- over **HTTP**;
- verpakt in **XML**;
- gecodeerd in **UTF-8**.

Interoperabiliteit

Interoperabiliteit is het uitwisselen van (meta)data tussen systemen en software via een gemeenschappelijke set procedures/internetprotocollen.

In de context van informatiediensten van erfgoedinstellingen geldt, dat de wijze van beschikbaar stellen hergebruik en uitwisseling van data moet ondersteunen.

Webdiensten

Dit is de verzamelnaam voor technologie waarmee aangesloten systemen onderling communiceren. De communicatie verloopt via het standaardprotocol van het World Wide Web (WWW) **HTTP**; de 'mark-up taal' die gebruikt wordt is **XML**.

Webdiensten maken het technisch mogelijk om digitale informatie over het erfgoed, zonder menselijke tussenkomst, op te vragen ten behoeve van hergebruik. Omdat de informatie verpakt is in XML, is geautomatiseerde verwerking mogelijk. Hergebruik is uiteindelijk alleen mogelijk als de informatie zelf interpreteerbaar is. Dat stelt aanvullende eisen wat betreft de beschrijving van de informatie. Praktische voorbeelden: interfaces (gedocumenteerde API's)

HTTP

HTTP (HyperText Transfer Protocol) is de wereldwijde standaard voor informatie-uitwisseling op het WWW. Vrijwel alle erfgoedinstellingen hebben een presentatie op het web, waarbij de informatie in HTML-opmaak wordt gepresenteerd aan eindgebruikers. Vrijwel alle Content Management Systemen hebben tegenwoordig een mogelijkheid om data via HTTP beschikbaar te stellen.

Het beschikbaar stellen van digitale informatie over erfgoed volgens de definitie van vindbaarheid binnen DE BASIS is primair gericht op hergebruik door systemen. Dat vereist dat toegang kan worden verkregen tot de informatie zonder context en opmaak. Daarin wordt voorzien door een gestandaardiseerd formaat (XML) en structuur (Dublin Core).

Overigens kan informatie verpakt in XML door middel van XSL Transformaties en Cascading Style Sheets (CSS) worden gepresenteerd als HTML. Beschikbaarstelling van de eigen content in XML via HTTP kan dus zowel de externe vindbaarheid als de presentatie in de eigen website voeden.

XML

Bij het digitaliseren van informatie is structuur een belangrijk hulpmiddel om betekenis te geven aan de digitale data. XML (Extensible Markup Language) vormt een gestandaardiseerde manier om onafhankelijk van de vorm, de inhoud, het platform en het gebruik, digitale data te structureren. XML maakt gebruik van platte tekst en is leesbaar voor mensen en systemen en daarom duurzaam.

Bij XML wordt alle informatie geplaatst tussen een begin en eindlabel (tags). Op die manier wordt aan de informatie een nadere betekenis gegeven.

UTF-8

UTF-8 (8-bit Unicode Transformation Format) is in staat om elk teken uit de Unicode-tekenset te coderen en op te slaan. Dit betekent dat internationale tekensets, bijv. Chinese karakters of Cyrillisch ook opgeslagen kunnen worden. Ook worden de 128 primaire Latijnse tekens op dezelfde wijze gecodeerd als binnen de ASCII tekenset wat betekent dat ook oude systemen en data zonder problemen geïmporteerd en geconverteerd kunnen worden (achterwaartse compatibiliteit). UTF-8 ontwikkelt zich richting de universele tekencodering op internet en vormt een onmisbaar onderdeel binnen DE BASIS voor vindbaarheid.

Digitaal erfgoed, zelfs beperkt tot de Collectie Nederland, is internationaal. Als gevolg komen in beschrijvingen van erfgoed en in objecten die tekst bevatten, tekens voor uit verschillende gebieden en talen. Voor de vindbaarheid van Nederlands digitaal erfgoed is het belangrijk dat data die toegankelijk worden gemaakt voor hergebruik, in dezelfde tekenset zijn gecodeerd en dat deze tekenset ruimte biedt om willekeurig welk teken in digitale vorm te bewaren. Alleen op die manier kunnen tekstuele data uit verschillende bronnen op uniforme wijze worden gevonden, gecombineerd en gepresenteerd.

Toegang: de informatie is beschikbaar voor

- **harvesting** door ondersteuning van **OAI-PMH**
- en/of
- **gedistribueerd zoeken** door ondersteuning van **SRU**

Toegang

Er kan geen twijfel over bestaan dat toegang is een eis voor elke dataleverancier. De technische voorwaarden voor toegang worden beschreven bij interoperabiliteit. De minimale eisen die gesteld moeten worden aan toegankelijke informatie zijn dat zonder kennis vooraf:

- de locatie waar de informatie toegankelijk is kan worden achterhaald en
- deze locatie kan worden gebruikt om (een deel van) deze informatie op te vragen in een herbruikbare vorm.

Dit is mogelijk door gebruik te maken van standaardprotocollen. In een protocol liggen afspraken vast over de manier waarop contact kan worden gelegd met de dataleverancier en hoe vervolgens het opvragen van informatie kan worden gerealiseerd. Het opvragen gebeurt 'just in case' (via harvesting) dan wel 'just in time' (via gedistribueerd zoeken). Effectief hergebruik van opgevraagde informatie is uiteraard afhankelijk van de soort beschrijving die is gehanteerd voor het digitale erfgoed.

Harvesting

Bij webharvesting worden data van een instelling automatisch opgehaald door een derde partij. De harvester is bekend met een startadres van een dataset op het web en haalt vanaf dat adres alle data binnen de set op. Doel van dit proces is om de opgehaalde data elders eventueel in combinatie met andere geïmporteerde datasets opnieuw te gebruiken. Het is een voorbeeld van een 'just in case' aanpak: data worden opgehaald om beschikbaar te hebben voor mogelijk hergebruik. (Kwaliteit van) hergebruik is mede afhankelijk van het formaat waarin de beschikbare metadata zijn opgeslagen. Ook zal de harvester regelmatig controleren of er wijzigingen of updates zijn.

Het principe van harvesting is sinds het midden van de jaren negentig uitgewerkt als model voor het uitwisselen van gestructureerde data: het Open Archives Initiative (OAI). OAI wordt succesvol toegepast in een netwerk van Institutional Repositories binnen de universitaire wereld. Het is een voorbeeld van succesvolle interoperabiliteit. Het OAI-model en het uitwisselingsprotocol OAI-PMH is echter generiek en breed toepasbaar voor (erfgoed)data.

OAI-PMH

OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting) is een protocol voor uitwisseling van metadata. OAI-PMH vraagt een set metadata records op, de dataleverancier is verplicht om Dublin Core te ondersteunen. Maar daarnaast is het mogelijk andere formaten uit te wisselen, zoals CDWA lite (Categories for the Description of Works of Art), een museaal beschrijfschema voor kunstobjecten.

Een metadata record wordt voor OAI-PMH voorzien van een OAI header. De header bevat o.a. een datum die aangeeft wanneer het betreffende record is geharvest van de dataleverancier (dataprovier) en een identificatie die zorgt voor de link naar het overeenkomstige item bij de dataleverancier. De datum wordt gebruikt als controlemiddel door de serviceprovider (degene die de data ophaalt) om te controleren op wijzigingen en updates. Identificatie maakt het mogelijk om aanvullende gegevens over het item op te halen.

Het OAI model biedt erfgoedinstellingen de mogelijkheid om gedigitaliseerd erfgoed (de data, d.w.z. metadata en objecten) herbruikbaar te maken voor de eigen (wisselende) eindgebruikerspresentaties (de diensten) en tegelijkertijd voor portals en Google.

Momenteel lopen er diverse (onderzoeks)projecten om het model van OAI uit te breiden. Enerzijds door niet alleen metadata maar ook complexe, samengestelde objecten uitwisselbaar te maken. Anderzijds door het protocol ook toe te passen voor uitwisseling van data van serviceprovider naar dataprovier, bijvoorbeeld om informatie waarmee de data bij de serviceprovider is verrijkt blijvend aan het bron-item bij de dataprovier toe te voegen.

Gedistribueerd zoeken

Bij gedistribueerd zoeken worden data die een dataleverancier vindbaar heeft gemaakt opgehaald via het web door middel van een zoekactie. Het zoekstelsel en de doorzochte dataset staan los van elkaar en worden veelal door verschillende partijen beheerd. In het model van gedistribueerd zoeken kan één zoekstelsel bij meerdere dataleveranciers tegelijkertijd zoeken. Omgekeerd kan ook één dataset door verschillende zoeksystemen worden bevroegd.

Het zoekstelsel heeft het webadres waarop een dataset vindbaar is gemaakt nodig en gebruikt dit adres om de data behorende bij een zoekvraag op te halen. De opgehaalde data kunnen, eventueel in combinatie met andere opgehaalde data, worden gebruikt in een webpagina. Voor dit hergebruik kan het nodig zijn de data overall in een bepaald formaat op te vragen, bijvoorbeeld in Dublin Core.

De opgehaalde data zijn daarna los komen te staan van de bron. Bij elk volgend gebruik wordt de zoekactie in principe opnieuw uitgevoerd om zeker te zijn van de actualiteit van de opgehaalde data. Het is een voorbeeld van een 'just in time' aanpak: data worden opgehaald om direct beschikbaar te zijn voor hergebruik.

Het principe van gedistribueerd zoeken is vanaf de jaren zeventig gebruikt binnen bibliotheken om wereldwijd catalogi met wetenschappelijke literatuurverwijzingen te bevragen. Voor dit doel is een protocol ontwikkeld waarin afspraken zijn gemaakt over de manier waarop zoekstelsel en dataset met elkaar communiceren bij het afhandelen van de zoekacties: Z39.50. De zoekportalen die vooral academische bibliotheken nu op hun website aanbieden maken veel gebruik van Z39.50. Op die manier is het voor eindgebruikers mogelijk om in één zoekactie data van verschillende wereldwijd verspreide datasets op een bepaald vakgebied als één resultaatset te verkrijgen. Met de opkomst van het web ontstond de behoefte de mogelijkheden van Z39.50 ook daar te benutten. Dit heeft geresulteerd in de SRU standaard.

SRU

SRU (Search/Retrieval via URL) gebruikt HTTP om een zoekvraag af te handelen en wisselt de data uit in XML-formaat. De vraag is verpakt in URL/XML.

SRU is een standaard om realtime informatie van één of meerdere dataleveranciers met één zoekacties op te halen voor hergebruik. Erfgoedinstellingen die gedigitaliseerd erfgoed (de data, metadata en objecten) via SRU vindbaar maken gebruiken dit om eigen wisselende eindgebruikerspresentaties (de diensten) op te zetten. Bovendien is deze informatie ook toegankelijk voor hergebruik in bijvoorbeeld sectorportalen, de European Digital Library en in Web 2.0 toepassingen als AJAX.

SRU voorziet in de behoefte om hergebruik van erfgoed informatie mogelijk te maken zonder kennis van de data (structuur, inhoud) en de techniek (indexering, zoektaal, sessiebeheer) bij dataleveranciers. De standaard stelt het gebruik van de Common Query Language (CQL) verplicht. Verplicht vastgelegd is het beschrijvingsformaat van de dataleverancier, in welke indexen gezocht kan worden en in welke metadata structuren zoekresultaten kunnen worden opgehaald.

Na uitlezing van de beschrijving is een zoekstelsel in principe zelfstandig in staat om een correcte zoekactie uit te voeren en de resultaten ervan op de juiste manier te verwerken.

Beschrijving: de informatie is beschikbaar als

- metadata in **Dublin Core** als structuur met
- **identificatie** van de bron in **URI**.

Beschrijving

Om het duurzaam delen en uitwisselen van data mogelijk te maken, is het essentieel dat niet alleen uitwisselbare bestandsformaten worden gebruikt, maar ook standaardschema's voor beschrijving, inhoudelijke ontsluiting en identificatie van collecties en objecten. Bij het gebruik van sector- en domeinspecifieke beschrijvingsmodellen moet rekening gehouden worden met de mogelijkheid tot hergebruik over de sector- of domeingrens heen.

Metadata schema's (en indien nodig inhoudelijke ontsluiting) moeten door de partij die de (meta)data bevraagt geïnterpreteerd kunnen worden. Alleen als aan die voorwaarde wordt voldaan zijn diensten mogelijk die content van meerdere leveranciers samenbrengen. Daarnaast is identificatie noodzakelijk bij hergebruik van metadata. Deze houdt de connectie in stand met de oorspronkelijke bron waar de metadata worden beheerd. Identificatie maakt het verkrijgen van de metadata van het (digitale) erfgoedobject mogelijk.

Dublin Core

Dublin Core Metadata element set is een NISO en ISO gecertificeerde standaard voor online uitwisselbare metadata. Dublin Core is internationaal het meest gebruikte datamodel binnen het erfgoed, vanwege de sectoroverstijgende toepassingsmogelijkheden.

Het Dublin Core datamodel is generiek, eenvoudig en flexibel. Er zijn 15 basisvelden (elements) waarmee elk erfgoedobject kan worden beschreven.

In Nederlandse vertaling zijn dat: titel (title), maker (creator), onderwerp (subject), omschrijving (description), uitgever (publisher), andere medewerkers (contributor), datum (date), type (type), formaat (format), identificatie (identificatie), bron (source), taal (language), relatie (relation), dekking (coverage), rechten (rights).

Alle velden kunnen 0, 1 of meer keer voorkomen.

Hoewel de generieke opzet de uitwisselbaarheid groot maakt, is het voor sommige vormen van hergebruik nuttig of zelf noodzakelijk om velden specifiek te gebruiken en/of om de inhoud ervan meer te controleren.

Het datamodel voorziet in deze wens door de mogelijkheid om aan velden extra kwalificaties geven: Dublin Core Qualified (DCQ). Om het onderscheid duidelijk te maken wordt Dublin Core zonder qualificiers wel aangeduid met: Dublin Core Simple. De kwalificaties bij DCQ kunnen bestaan uit, ofwel een toevoeging aan het veld, ofwel een beperking van het gebruik ervan. Een voorbeeld van de eerste kwalificaties is bijvoorbeeld de toevoeging van "alternatief" aan het veld "titel". Een voorbeeld van een beperking is het voorschrift om de datum te noteren volgens de internationale standaard ISO8601.

Als Dublin Core op de juiste manier is geformatteerd volgens het officiële XML- of RDF-schema kunnen de beschrijvingen automatisch verwerkt worden. Op die manier kunnen de qualificiers naar behoefte worden genegeerd en komt de brede uitwisselbaarheid van Dublin Core (Simple) niet in gevaar.

Identificatie

Identificatie is noodzakelijk om bronnen blijvend te kunnen traceren. Dit vereist het maken van afspraken over de organisatie en de techniek. Bronnen dienen uniek te worden benoemd. Voor identificatie zijn persistente identificatie en permanente adressering noodzakelijk.

URI

De URI (Uniform Resource Identifier) is het adres van de data of metadata op het WWW, zichtbaar bovenin de browser. Binnen URI worden twee subclasses onderscheiden. URL's (Uniform Resource Locators) identificeren objecten aan de hand van een netwerklocatie die uniek, maar tijdelijk is. De URN's (Uniform Resource Name) zijn niet direct gekoppeld aan een netwerkadres, maar aan een namespace.

Persistentie

Persistentie betreft de verkrijgbaarheid op lange termijn. Webnamen/adressen kunnen veranderen en daardoor kunnen objecten moeilijker of niet meer te vinden zijn. Persistentie zoekt naar een oplossing.

Binnen de context van de Digitale Collectie Nederland lijkt één uniforme benaming (namespace) niet realistisch. Maar de **URN** standaard biedt de mogelijkheid om diverse verschillende naamruimtes door elkaar te gebruiken.

De URN is een combinatie tussen een identifier (een tekenstring) en een naamruimte-aanduiding. Binnen de naamruimte zijn de identifiers uniek en in een gespecificeerd formaat. De autoriteit die verantwoordelijk is voor de naamruimte ziet daar op toe. Voor URN is het een vereiste dat elke naamruimte-aanduiding wereldwijd uniek is, waarmee ook de URN zelf uniek is.

Voor sommige objecttypen binnen de Digitale Collectie Nederland bestaan al (internationale) naamruimtes. Bekende voorbeelden zijn het Internationaal Standaard Boeknummer (ISBN) bv: urn:ISBN:90-9017753-1 en het National Bibliographic Number (NBN) bv urn:nbn:de:bsz:93-opus-59. Voor andere objecttypen kunnen nieuwe naamruimtes geregistreerd worden bijvoorbeeld door koepelorganisaties. Toekenning en gebruik van URN's voor persistente identificatie binnen de Digitale Collectie Nederland is dus afhankelijk van een dekkende structuur van organisaties die als registratieautoriteit voor (een deel van) het erfgoed willen optreden.

Resolving

Resolving betekent dat de gebruiker achter de schermen, bijvoorbeeld door doorverwijzingen, geleid wordt naar het actuele adres van het gezochte object of zoekvraag.

URN's zijn voorzien van functionaliteit die de gebruiker (mens of machine) zo dicht mogelijk bij het object brengt dat door de URN geïdentificeerd wordt. Idealiter resulteert het activeren van de URN direct tot het verkrijgen van het object.

In de praktijk zijn er enkele obstakels. Zo zijn lang niet alle objecten digitaal. In dat geval brengt een URN de gebruiker zo dicht mogelijk bij het fysieke object, dus bijvoorbeeld de (digitale) beschrijving van een schilderij, archiefstuk of een aanvraagmogelijkheid van een bibliotheekboek.

Een ander probleem is dat de digitale objecten uiteindelijk verkregen worden op een netwerklocatie (URL), veelal bij de beherende organisatie. De URL is veelal tijdelijk, want afhankelijk van de organisatie, het gebruikte systeem en het mogelijke formaat. Om dit probleem op te lossen wordt de URN bij activeren vertaald in de actuele netwerklocatie d.m.v. resolving.

Bij resolving wordt de gebruiker ofwel achter de schermen enkele keren doorgestuurd om uiteindelijk te belanden bij de lokale resolver die kennis heeft over de actuele netwerkadres van het gevraagde object, ofwel er is één resolver die kennis van alle actuele netwerkadressen heeft. Vanuit beheersogpunt heeft de eerste methode de voorkeur. De verkrijgbaarheid van (digitale) erfgoedobjecten binnen de Collectie Nederland is dus afhankelijk van de opzet en beheer van een hiërarchisch netwerk van resolvers.

Scenario DE BASIS

Dit scenario beschrijft hoe een erfgoedinstelling DE BASIS kan toepassen in de praktijk. Uitgangssituatie is dat een erfgoedinstelling een collectie digitaliseert als basis voor gebruik op de eigen website, een regionale portal, de Digitale Collectie Nederland en Google

Aanpak

De erfgoedinstelling registreert haar collecties en objecten in een Collectie Beheer Systeem (CBS). Bepaalde velden worden bij registratie uitgedrukt in de originele taal met de originele karakters van die taal, waaronder in (modern) Chinees. Dit is mogelijk omdat DE BASIS standaard voor tekstcodering UTF-8 de karakters kan coderen. Bovendien garandeert UTF-8 dat op oude apparaten of apparaten met beperkte mogelijkheden tenminste de belangrijkste karakters van de westerse tekenset (ASCII) correct getoond worden.

De instelling kent aan ieder object een identifier toe. Er wordt voor gezorgd dat de identifier uniek is binnen de instelling. De fysieke objecten dragen een dergelijk kenmerk bij zich. In een register wordt op basis van dit nummer de actuele standplaats van het object bijgehouden. De objecten kunnen altijd gevonden worden.

Bij het digitaliseren wordt van ieder object één of meerdere digitale reproductie(s) gemaakt. De afbeeldingen worden opgeslagen. Een register houdt op basis van de identificatie de actuele 'standplaats' van de digitale reproducties bij: de server, de map en de bestandsnaam. Omdat de server voorzien is van een webserver kunnen plaatjes via DE BASIS voor digitale communicatie: HTTP naar een browser of ander websysteem worden gestuurd.

De instelling heeft een OAI-module in beheer. Via deze module is elke collectie-item uit het Collectie Beheer Systeem beschikbaar als Dublin Core record. De Dublin Core records worden gegenereerd uit de collectie-items waarbij de instelling heeft aangegeven welke velden uit de interne collectieregistratie de 15 mogelijke elementen van Dublin Core voeden. Naast een Dublin Core record beheert de OAI-module voor elke item ook de datum/tijd van de laatste wijzigingen en een unieke OAI-sleutel.

De OAI-module is verbonden aan internet via HTTP. Elke browser of andere server met een internetverbinding kan contact zoeken met de OAI-module door diens unieke hostname en/of IP-nummer. De OAI-module 'reageert' op verzoeken die voldoen aan DE BASIS standaard voor het ophalen van digitale data: OAI-PMH. OAI levert gestandaardiseerde webdiensten voor de data uit de database en/of aanvullende meta informatie (zoals bijvoorbeeld de naam van de database). Voorbeelden van ondersteunde verzoeken zijn: geef het Dublin Core record op basis van een gevraagde sleutel of geef alle Dublin Core records met een tijdsaanduiding later dan gisteren. De reactie van de OAI-module is verpakt in DE BASIS standaard voor gestructureerde data: XML en opnieuw verstuurd over HTTP.

Naast een OAI-module heeft de instelling ook een SRU-module als toegangspoort tussen het web en de collectie-informatie in het CBS. Via de SRU-module is elk collectie-item uit het CBS beschikbaar opnieuw als Dublin Core verpakt in XML. Via een willekeurige browser of andere server met een internetverbinding kan contact gezocht worden met de SRU-module. SRU is DE BASIS standaard voor zoeken op afstand. Onderdeel van het SRU-protocol is de mogelijkheid om bij elke SRU-server de functionele informatie (explain-informatie) op te vragen. De bevrager krijgt in dit scenario dan in XML (dus machine-leesbaar) als antwoord

dat collectie-items kunnen worden opgevraagd op basis van woorden uit de titel, naam vervaardiger, trefwoord en datering en dat Dublin Core als metadata formaat wordt ondersteund.

Resultaat

De instelling heeft zelf XSLT (Extensible Stylesheet Language Transformations) geprogrammeerd die de XML van de SRU Explain informatie vertaalt naar een zoekinterface met het actuele functionele aanbod van het CBS via SRU. Het SRU zoekresultaat in XML bestaat uit collectie-items in het eigen niet-gestandaardiseerde metadata formaat omdat daarin de meeste informatie kan worden opgenomen.

Het XML-zoekresultaat wordt opnieuw met XSLT vertaald naar HTML voor presentatie en functionaliteit op de website van de instelling. Het veld met de identifier wordt gepresenteerd als interactief element. Bij het laden van de webpagina wordt op basis van dit kenmerk het actuele webadres van bijbehorende digitale reproducties opgevraagd en in de pagina getoond. Is er geen digitale reproductie dan krijgt de gebruiker de mogelijkheid het object of een reproductie ervan aan te vragen.

Voor bepaalde pagina's in de website is het volledig zoekformulier vervangen door een directe SRU-opdracht met één tot alle parameters vast ingebed in de paginacode. Bij aanroep van dergelijke pagina's door bezoekers van de website wordt automatisch een actuele set van collectie-item(s) uit het CBS al dan niet met afbeeldingen gepresenteerd. In de presentatielaag van de website combineert de instelling de SRU-functionaliteit van verschillende systemen. Zo worden zoekresultaten en overzichten uit het archief gepresenteerd met afbeeldingen uit de beeldbank en elders plaatjes met van objecten met contextinformatie uit digitale tentoonstellingscatalogus samengevoegd.

Vernieuwing

Op deze manier zijn de gebruikersdiensten van de instelling dynamisch, flexibel en makkelijk te onderhouden. Nieuwe datasets die met de voortschrijdende digitalisering beschikbaar komen zijn bij oplevering direct zichtbaar in de website. Recentelijk is de informatie uit het nieuw geïmplementeerde bezoekersinformatiesysteem vrij gemakkelijk in de bestaande website geïntegreerd. De op verzoek van de instelling toegevoegde zoekindex op techniek is ook direct door bezoekers te raadplegen en de nieuwe toeschrijving van de serie tekeningen op basis van recent onderzoek wordt ook direct op de website zichtbaar.

De in ontwikkeling zijnde extra gebruikersdienst van de instelling in dit scenario om een screensaver te kopen op basis van een collectie-item kan snel worden opgezet omdat er gebruik wordt gemaakt van dezelfde inhoud en beheersystemen als de website. Een nieuwe huisstijl kan in het tijdsbestek van één onderhoudsvenster via de stylesheets over alle webdiensten worden toegepast en voor volgend jaar heeft de instelling een hele website gepland compleet met web 2.0 functionaliteit. Daar wordt aan gebouwd door de XSLT-, en HTML-ontwikkelaars op basis van de documentatie van de huidige onderliggende infrastructuur.

Deelnemer erfgoedportal

De instelling gebruikt dezelfde opzet en geniet dus dezelfde voordelen om haar collectie-informatie ook te presenteren in de regionale erfgoedportal. Daar profiteert de instelling van het feit dat een gecombineerde presentatie van informatie van meerdere instellingen meer gebruikers aanspreekt. Bovendien genereert de organisatie rond de erfgoedportal meer publiciteit.

Een verschil met de presentatie op de eigen website is wel dat de instelling voor participatie in de erfgoedportal met de andere instellingen een gemeenschappelijk metadataformaat heeft moeten kiezen. Het formaat moest generiek genoeg zijn zodat alle instellingen er hun interne formaten op een aanvaardbare manier op konden mappen. Maar ook was een zo rijk mogelijk formaat gewenst om de bezoekers betekenisvolle informatie en functionaliteit te kunnen bieden.

Ook was het nodig om de zoekfunctionaliteit van SRU onderling af te stemmen ten einde het bereik van een zoeken over collecties heen zo groot mogelijk te maken.

Geogst

De instelling heeft haar collectie-informatie ook aangemeld in een landelijke erfgoedbrede portal. Deze dienst maakt gebruik van OAI en de beschrijvingsstandaard van DE BASIS: Dublin Core. In dat geval wordt de informatie over de collectie-items niet 'just-on-time' verzameld (zoals bij SRU) maar 'just-in-case'. De collectie-informatie wordt door de erfgoedportal opgevraagd en vervolgens opnieuw opgeslagen. In de zo ontstane centrale opslag (repository) wordt item-informatie van verschillende collecties en instellingen tot één collectie samengevoegd. Bovendien worden door de erfgoedportaldienst data van verschillende bronnen in één record samengevoegd. Zo worden de Dublin Core records met geografische data zoveel mogelijk verrijkt met een plaatsnaam afkomstig uit een GEO-collectie en worden beschrijvingen toegevoegd voor de speciale webdossiers die het nationale erfgoedportal regelmatig inricht op het centrale repository.

Google-n

Een aansprekende meerwaarde voor de instelling is het hergebruik van de collectie-informatie door Google op basis van OAI. De OAI-toegang is door de instelling aangemeld bij Google en de Google spider gebruikt de OAI-functionaliteit om alle item-informatie door te lopen en de metadata vindbaar te maken in Google Search. Indexering in Google is voor elke erfgoedinstelling een belangrijk instrument om nieuwe doelgroepen in aanraking te brengen met haar collectie.

Long tail

Bovenstaand hergebruik van collectie-informatie buiten de eigen instelling met behulp van OAI en/of SRU is voor de instelling meervoudig toepasbaar nu en in de toekomst. In principe kan dat zelfs zonder bemoeienis van de instelling zelf. Uiteraard houdt de instelling altijd de regie over waar, wanneer en waarvoor haar collectie-informatie wordt gebruikt. Dat doen ze door de vindbaarheid via OAI en SRU van collectie-informatie (velden of deelverzamelingen) voor externe partijen deels af te sluiten. Bovendien neemt de instelling soms voorwaarden voor hergebruik op in individuele iteminformatie of voor deelverzamelingen.

Belangrijk voor het overwinnen van aarzeling bij de directie van de instelling om collectie-informatie vindbaar te maken is het inzicht dat alleen de meta-informatie buiten de instelling wordt hergebruikt. Het erfgoed zelf (of digitale reproducties ervan) blijft binnen de instellingen zelf. De identificatie, een onderdeel van DE BASIS voor vindbaarheid, vormt de koppeling tussen de meta-informatie en het collectie-item zelf. Om die koppeling werkend te krijgen publiceert de instelling haar unieke kenmerken voor collecties en items volgens DE BASIS voor identificatie: URI. In URI is het kenmerk uitgebreid met een identificatie voor de beherende instelling en een identificatie die aangeeft dat deze URI onderdeel uitmaakt van het schema voor identificatie van Nederlands erfgoed. Dit schema is internationaal geregistreerd en wordt ondersteund door een netwerk van servers (resolvers) die in staat zijn URI in een aantal stappen te 'vertalen' naar een geldige URL waar op dat moment het item behorende bij de URI te vinden is.

De URI biedt gebruikers van de meta-informatie waar dan ook een directe verbinding met het erfgoedobject of de collectie op de site van de instelling zelf. Het vindbaar maken van collectie-informatie draagt dus direct bij het verkeer naar de eigen website en het gebruik van het beheerde erfgoed.

Digitale Collectie Nederland

Feit is dat de instelling door (delen van) haar collecties vindbaar te maken voor hergebruik (onder voorwaarden) zich manifesteert als contentleverancier van (digitaal) erfgoed. Als lid van een breed netwerk van dergelijke contentleveranciers waarop een dynamische aanbod van commerciële en niet-commerciële diensten is gebouwd levert onze voorbeeldinstelling een concrete en publieksgerichte bijdrage aan de Digitale Collectie Nederland.

BIJLAGE 1: DE TOTSTANDKOMING VAN DE BASIS

De totstandkoming van DE BASIS is een zorgvuldig proces geweest, waarbij DEN heeft samengewerkt met experts en instellingen uit de erfgoedsector.

Het proces bestond uit:

- inventarisatieronde bij experts,
- brede inspraakronde gebaseerd op de RFC-procedure 2026
- toetsing in het veld.

Inventarisatieronde experts

De eerste fase in het project was een inventarisatie onder experts uit het erfgoedveld. Hiertoe werden zowel de redactieleden van het ICT-register van DEN gerekend, als enkele specialisten uit het erfgoedveld die betrokken wilden worden bij de inventarisatie. De redactieleden van het ICT-register vertegenwoordigen alle sectoren binnen het erfgoedveld en vormen zo een representatieve afspiegeling. Het redactiestatuut en meer informatie over de redactie zijn te raadplegen op de DEN-website (www.den.nl/register/redactie).

De inventarisatie bestond uit twee onderdelen.

Eerst werd door iedere expert afzonderlijk in vrije tekst omschreven:

- op welke terreinen van ICT in zijn of haar ogen behoefte bestaat aan standaardisatie en normering;
- voor welke deelonderwerpen binnen de terreinen dat met name geldt;
- hoe standaardisatie en normering gerealiseerd kunnen worden.

De indeling in terreinen is overgenomen uit het ICT-register. De basis hiervoor ligt bij het Europese project ePSINet. De indeling beschrijft de levenscyclus van digitale data: Vervaardiging – Beschrijving - Opslag – Toegang – Presentatie – Beheer – Gebruik – Interoperabiliteit - Duurzaamheid.

De resultaten van deze inventarisatie zijn meerdere malen plenair besproken door de redactie en na de vaststelling van de terreinen en deelonderwerpen voor standaardisatie, werd de tweede stap in de eerste fase gezet: de vertaalslag naar concrete ICT-standaarden. Hierbij is besproken welke standaarden beschikbaar zijn, en in welke mate ze beschouwd kunnen worden als norm voor Nederlandse erfgoedinstellingen. Hierna zijn de standaarden voorlopig vastgesteld door de redactie.

Brede inspraakronde door middel van RFC-procedure

Nadat de redactie de standaarden voorlopig had vastgesteld, zijn deze als voorstel voor DE BASIS bekendgemaakt. Om de vaststelling van DE BASIS goed te laten verlopen, is de RFC-procedure 2026 (Request for Comment) gevolgd.

Deze RFC beschrijft het proces voor standaardisatie van internettechnologieën. Hierin zijn drie versies van de standaard of norm te onderscheiden: het voorstel, het concept en de definitieve versie.

In het kort verloopt de procedure als volgt:

1. Beschikbaarstelling van het voorstel ten behoeve van commentaar van iedereen die maar wil. Het document heeft nog geen formele status, en er kunnen nog steeds wijzigingen in aangebracht worden.
2. Beschikbaarstelling van de conceptversie. In deze fase wordt gekeken naar de praktische implicaties van DE BASIS. Bij een representatieve groep erfgoedinstellingen wordt geanalyseerd in hoeverre DE BASIS al in praktijk wordt gebracht, en wat de consequenties zijn als dat nog niet het geval is. Wijzigingen zijn alleen nog maar toegestaan om onduidelijkheden weg te nemen.
3. Vaststelling en validatie door de redactie van het ICT-register. De precieze criteria voor validatie zullen door de redactie op basis van de reacties uit het veld worden opgesteld en bekend gemaakt. Na de definitieve validatie zal DE BASIS worden gepubliceerd.

Er is een communicatieplan opgesteld met een inventarisatie van alle mogelijke en geschikte communicatiemiddelen. Het voorstel voor DE BASIS is verspreid via de in het plan opgenomen communicatiemiddelen, o.a. de communicatiekanalen van DEN, aanverwante organisaties en daarnaast via enkele goedgelezen discussielijsten als bijv. Nedbib-L.

In samenspraak met de redactie is voor een wiki gekozen als middel tot interactie met het erfgoedveld. De wiki die DEN heeft opgezet, wiki.den.nl, beschrijft alle voorgestelde standaarden, het proces en de voornaamste reden waarom voor een wiki is gekozen is de mogelijkheid tot directe interactie.

Wiki.den.nl

Na openstelling van de Wiki is iedereen in de gelegenheid gesteld het voorstel te becommentariëren. Elke commentaar in de vorm van een aanpassingen aan de voorgestelde set basisnormen, verbeterde artikelen over de geselecteerde onderdelen ervan, discussiepunten of vragen binnen deze periode kon gegarandeerd worden meegenomen in de commentaarperiode.

Aan de hand van de statistieken kan worden geconcludeerd dat de vraag welke ICT-instrumenten en richtlijnen in gebruik zijn bij de digitalisering van erfgoed in Nederland velen interesseert. De circa 4500 bezoeken hebben echter niet tot wezenlijke aanpassingen of aanvullingen geleid; niet wat betreft de voorgestelde set en ook niet wat betreft de inhoud van de artikelen over de geselecteerde ICT-instrumenten. Wel zijn er een aantal tekstuele aanpassingen voorgesteld die vrijwel allemaal zijn overgenomen. Met het aflopen van de commentaarperiode is voorstel nu vastgesteld als concept.

Analyseperiode toetsing DE BASIS in het veld

Dit concept is na de analyseperiode door DEN getoetst aan de praktijk bij representatieve erfgoedinstellingen. Doel was te analyseren of de normen toepasbaar zijn in de praktijk van erfgoedinstellingen en om inzichtelijk te maken wat het resultaat van toepassing is voor de vindbaarheid van collecties. DEN heeft in overleg met de redactie van het ICT register een aantal representatieve instellingen geselecteerd.

De volgende instellingen zijn bereid gevonden deel te nemen:

- Geheugen van Nederland
- Internationaal Instituut en Archief voor de Vrouwenbeweging (IIAV)
- Brabantse Museum Stichting
- Museum Jan Cunen
- Naturalis
- Universiteitsbibliotheek UvA
- Historisch Centrum Overijssel
- Regionaal Archief Nijmegen
- Nationaal Archief

Aanpak

DEN heeft een vragenlijst opgesteld waarvan de beantwoording inzicht moet geven in de praktische implicaties van de Basisnormen ICT voor vindbaarheid. Deze vragenlijst is ook gepubliceerd in de wiki (<http://wiki.den.nl/Analyseperiode>).

Medewerkers van DEN hebben de instelling bezocht. Tijdens dit bezoek is aan de hand van de vragenlijst diepgaand ingegaan op:

- de praktijk van vindbaarheid;
- welke wijze de basisnormen uit het voorstel al worden toegepast;
- hoe wenselijk en haalbaar toepassing van de complete set is;
- wat de resultaten van toepassing van de basisnormen zijn of zouden kunnen zijn.

De discussie tijdens het bezoek en de beantwoording van de vragenlijst zijn door DEN uitgewerkt in een rapportages per instelling. Alle rapportages kunnen geraadpleegd worden op wiki.den.nl (<http://wiki.den.nl/Analyseperiode>).

Vaststelling en publicatie

De redactie draagt de eindverantwoordelijkheid bij de vaststelling van DE BASIS. DEN is verantwoordelijk voor het beheer en beschikbaarstelling van de vastgestelde versie en de bijbehorende documentatie.

DE BASIS zal in enkele versies beschikbaar komen. Allereerst als een overzichtelijk en leesbaar document, dat zowel online (bijv. PDF) als in gedrukte vorm eenvoudig verspreid kan worden. Daarnaast is DE BASIS als een dynamische lijst op te vragen uit het ICT-register op de website van DEN. Alle geselecteerde instrumenten worden zodanig gemarkeerd, dat de set eenvoudig online is op te vragen. Op de website is ook de aanvullende documentatie over de afzonderlijke ICT-standaarden te vinden. Door de context van het register kan ook snel gezien worden welke instrumenten ook als best practices worden gezien, maar toch niet de drempel van DE BASIS hebben gehaald.

Tenslotte zal DE BASIS als geheel (of onderdelen eruit) gemarkeerd worden in de andere kwaliteitszorgsystemen die er zijn voor de erfgoedsector, zoals de kennisbank van de Taskforce Archieven.

BIJLAGE 2: SAMENVATTING VAN DE ANALYSEPERIODE: DE BASIS EN DE HUDIGE PRAKTIJK

De volgende samenvatting is geschreven op basis van werkbezoeken aan de negen genoemde erfgoedinstellingen.

Informatiebeleid

Veel instellingen zijn nog bezig met formuleren van beleid en plannen voor digitalisering. Een aantal doet dat met behulp van een subsidie uit de regeling Digitaliseren met Beleid. De noodzaak van een planmatige aanpak voor digitalisering wordt echter breed gevoeld. De huidige praktijk rond digitalisering en de gevolgen die activiteiten op dit gebied in het verleden hebben gehad op het dagelijks beheer worden als problematisch ervaren. Zo is een integrale zoekfunctie op de website voor de diverse datasets die erfgoedinstellingen inmiddels voor het publiek beschikbaar hebben op dit moment voor iedereen lastig te realiseren.

Bij veel instellingen zijn digitalisering en de rol van open standaarden daarin een expertiseterrein van enkelen. Deze experts ervaren vaak gebrek aan medewerking en draagvlak zowel bij de collega's (collectiebeheerders, registrators) als het management en ze zijn vaak overbelast. Informatieplanning zou daarin verandering kunnen brengen. Investeren in ICT-kennis bij inhoudelijke experts zou hier ook sterk verbetering kunnen brengen.

Veel van de bezochte instellingen zijn zich bewust van de *ad hoc* aanpak rond digitalisering de afgelopen jaren en erkennen de inefficiëntie en de risico's daarvan. Dit is echter mede het gevolg van het feit dat digitalisering voornamelijk gefinancierd wordt via projectgelden en subsidies. Daarbij zijn de korte termijn projectdoelen meestal belangrijker dan de beheersbaarheid en herbruikbaarheid van data op de lange termijn.

Op één uitzondering na, moeten overal de meeste keuzes over de informatie- en systeemarchitectuur nog worden gemaakt. De verwachting is veelal wel dat de standaarden van DE BASIS voor vindbaarheid daarin belangrijke peilers zullen zijn. Zo streven diverse instellingen naar toepassing van XML als formaat voor het beheer en uitwisseling van content.

Veel instellingen achten grote veranderingen in infrastructuur en organisatie noodzakelijk waarbij het echter zeer onzeker is of de huidige organisatie die kan realiseren.

Vooraf in de archiefsector zien instellingen hun leverancier als een obstakel voor het werken met open standaarden. In de museale sector worden leveranciers meer gezien als bondgenoot. Eén instelling laat de kennisontwikkeling m.b.t. standaarden en de toepassing ervan zelfs liever over aan de leverancier die op dat gebied echt als een strategisch partner wordt beschouwd.

Vindbaarheid

De definitie die wordt gehanteerd voor vindbaarheid als thema binnen DE BASIS komt niet overeen met de associaties die de meeste gesprekspartners bij de instellingen met dit begrip hebben. Daarin is vindbaarheid voornamelijk gerelateerd aan (doelgroepgerichte) diensten via websites. Vindbaarheid staat primair in dienst van eindgebruikers. Prioriteit geven aan vindbaarheid buiten de eigen website of websites van samenwerkingsverbanden komt weinig voor. De gangbare manier om de eigen collectie op

het web te presenteren is via internetmodules en webinterfaces op de eigen (beheer)systemen, gekoppeld aan de website. Enige uitzondering is wellicht de vindbaarheid in Google, maar daarbij gaat het niet direct om object- en/of collectiebeschrijvingen, maar meer om websites en webpagina's.

Incidenteel spreken gesprekspartners hun twijfel uit over nut en noodzaak van het vindbaar maken van de resultaten van basisdigitalisering voor hergebruik elders: Wie zit daar op te wachten? Gebruikers zien liever verhalen dan objecten en de collectiebeheerders zijn als geen ander in staat zinvolle presentaties te maken rondom objecten.

Functioneel hergebruik van digitale objecten door partijen anders dan de collectiebeheerder is lastig. Aan de ene kant is de basale informatie te oppervlakkig voor experts, aan de andere kant is de informatie zonder context niet te begrijpen voor het grote publiek. Eén instelling had vanuit die gedachte de ambitie om digitaal erfgoed vindbaar te maken in combinatie met een gebruikersschil. De schil levert experts meer functionaliteit en de leek meer context.

In het algemeen is hergebruik van dezelfde digitale content voor verschillende presentatiedoelen, zelfs intern, zeldzaam. Veelvuldig wordt een zelfde object op meerdere plekken vindbaar gemaakt zonder onderling verband. Maar één instelling heeft zijn architectuur zo ontworpen dat (her)gebruik van erfgoed in presentaties (diensten) wordt gefaciliteerd op basis van generieke vindbaarheidstelling van de data, zij het vooralsnog uitsluitend intern.

De andere instellingen zien DE BASIS vooral als hulpmiddel bij samenwerkingsprojecten. De motivatie om deel te nemen aan samenwerkingsprojecten ligt meestal in de aanvullende financiering. Hoewel presentatie van de informatie over digitaal erfgoed in samenwerkende erfgoedportals vaak als doel heeft het verkeer naar de eigen site te stimuleren, sluiten beide sites meestal niet mooi aan voor gebruikers. Ook is erfgoed dat is gedigitaliseerd voor een samenwerkingsproject, zelden in dezelfde vorm beschikbaar op de eigen website. Dat komt mede doordat de content meestal gecentraliseerd en op maat gemaakt wordt voor de presentatie van de erfgoedportal. Dit geldt voor metadata en digitale objecten.

De kennis van de conversie wordt centraal opgedaan en is moeilijk herbruikbaar door de leverancier van de content voor andere doeleinden. Conversie en/of verrijking van data is ook vrijwel altijd noodzakelijk omdat veel beschrijvingen van erfgoed gemaakt zijn vanuit het perspectief van de collectiebeheerder ten behoeve van het collectiebeheer. Door de centralisatie en specifieke conversies aldaar is het actualiseren binnen samenwerkingsverbanden lastig. Als oplossing wordt veelal overwogen een beheertoegang te maken op de centrale content, waardoor instellingen te maken krijgen met dubbele administratie.

Standaardisatie en normering

Over het nut en de noodzaak van standaardisatie en normering wordt verschillend gedacht. Met name bij instellingen waarbij samenwerkingsverbanden gemeengoed is, is de noodzaak duidelijk. Ook de planmatige aanpak die de digitalisering van de grote meerderheid van instellingen steeds meer kenmerkt leidt ertoe dat standaardisatie als belangrijk stuurmiddel wordt gezien.

Daarentegen worden er ook bezwaren aangegeven. Bijvoorbeeld dat het erg tijdrovend is om met een grote en diverse groep instellingen overeenstemming te bereiken over de keuze voor een standaard en de manier van implementatie.

Een ander argument is de snelheid waarmee het ICT-domein verandert. Standaarden van vandaag zijn morgen alweer verouderd.

Beschrijving en Dublin Core

Erfgoedinstellingen hebben meestal geen beschrijvende metadata die zondermeer geschikt is voor publiekspresentatie voor het grootste gedeelte van hun collectie. Met name in musea is hier nog een flinke investering nodig. Maar ook voor andere sectoren geldt dat collecties in het verleden zijn beschreven vanuit de beheersfunctie. Deze beschrijvingen zijn nauwelijks geschikt voor publieksdoeleinden.

Om hier toch in te voorzien wordt meestal ter gelegenheid van een concreet presentatiedoel beschrijvende metadata toegevoegd die dan soms niet geschikt zijn voor andere doeleinden.

Er zijn twijfels over de toepasbaarheid van Dublin Core. Niemand acht Dublin Core geschikt als intern datamodel. Als erfgoedbrede uitwisselstructuur is er echter geen alternatief voor Dublin Core. Elke instelling erkent dan ook de noodzaak om metadata over het erfgoed in beheer beschikbaar te kunnen stellen in Dublin Core, liefst 'on the fly'. De 20 velden van de basisregistratie voor musea zijn bijvoorbeeld goed te projecteren op Dublin Core. In andere domeinen, bijvoorbeeld bij geografische informatie, is het niet altijd eenduidig of en, zo ja, hoe bestaande informatie in Dublin Core beschikbaar kan worden gesteld, bijvoorbeeld omdat er moeilijk een objectenniveau is aan te duiden of omdat het gaat om afgeleide informatie, bijvoorbeeld tot stand gekomen op basis van onderzoekswerk van vrijwilligers.

De ervaring uit meerdere projecten leert dat meestal toch conversie en/of verrijking nodig is om Dublin Core records met erfgoed informatie van meerdere instellingen voor één gebruiksdoel samen te brengen. Ook wordt het veld subject vaak een vergaarbak.

Daarnaast zijn de informatiediepte en daarmee de toepassingsmogelijkheden bij uitwisseling in Dublin Core beperkt. Informatieverlies wordt gezien als een probleem. Het hergebruik van bijvoorbeeld samengestelde objecten of archiefcollecties is wel erg beperkt op basis van Dublin Core.

Dublin Core is echter te beschouwen als de ondergrens. Er is geen alternatief voor Dublin Core. Naar behoefte en mogelijkheden kunnen instellingen aanvullingen op Dublin Core ondersteunen. Daarbij zijn er verschillende oplossingsrichtingen: een aanvulling op Dublin Core (DCX) of ondersteuning van aanvullende metadata schema's (bv EAD, MODS).

Identificatie

Er bestaat wel enige weerstand in het veld tegen de basiseis van een persistente identificatie die vanuit de beschrijving herleidbaar is naar het originele record of object. De argumenten zijn van verschillende aard.

Praktisch zijn de objecten vaak niet (digitaal) toegankelijk. Strategisch willen collectiebeheerders de mogelijkheid om het digitale object te bekijken reserveren voor de presentatie op hun eigen website en, omgekeerd, erfgoedportalaanbieders de gebruiker op hun site houden en niet, via het volgen van de identificaties, van hun erfgoedportal zien verdwijnen.

Nog nergens wordt gewerkt met collectiebrede persistente, uitvoerbare identificatie. Er is wel een grote behoefte aan. Bij een aantal instellingen en projecten wordt gewerkt met URN. Kenners betwijfelen of het veld in staat is een sectorbrede organisatie van identificatieverstrekkers en resolversdiensten op te zetten en te onderhouden.

Toegang en interoperabiliteit

Op het gebied van interoperabiliteit zijn de minimale eisen (HTTP, XML en UTF-8) onomstreden en veelal al mogelijk, hoewel nergens in combinatie voor de totale collectie toegepast ten behoeve van vindbaarheid.

Slechts enkele instellingen hebben de toegangsprotocollen OAI-PMH en/of SRU geïmplementeerd. Waar dat wel het geval is, betreft het, op één uitzondering na, een beperkte dataset voor een specifiek doel. Eén instelling geeft aan dat meer algemeen streven naar webdiensten ook voldoet voor toegankelijkheid.

SRU is nog vrij onbekend vergeleken met OAI-MPH. Daarbij wordt de meerwaarde van SRU ten opzichte van OAI betwijfeld.

Mogelijke aanvullingen

EAD is een veelgenoemde standaard waarmee instellingen, zoals archieven, collecties ontsluiten. Hoewel omzetting van EAD naar Dublin Core mogelijk is, gaat daarbij zeer veel informatie verloren.

Een tweede aanvulling ten einde vindbaarheid op meerdere niveaus binnen erfgoed informatie mogelijk te maken is een standaard voor complexe objecten. De noodzaak van een standaard voor complexe objecten wordt door sommige gezien als een nuttige aanvulling op DE BASIS. Daarbij wordt wel erkend dat een dergelijke standaard nu als ondergrens bij digitalisering niet realistisch is.

Eén instelling acht Creative Commons (CC) een nuttige aanvulling op DE BASIS als ondergrens voor Digitale Right Management (DRM) en Intellectual Property Rights (IPR) dat speelt bij hergebruik van informatie.

Meerdere instellingen noemen thesauri als onmisbaar element om vindbaarheid over (erfgoedbrede) collecties effectief te maken. De Art and Architecture Thesaurus (AAT) wordt een aantal keren genoemd. Ook de behoefte aan een geografische referentiestructuur wordt een keer genoemd.

De wiki

Het idee van DEN om een wiki te gebruiken als middel voor kennisopbouw en discussie wordt algemeen gesteund. Aan de andere kant heeft geen enkele instelling actief bijgedragen aan de wiki. Ook hadden slechts enkele instellingen de inhoud van de wiki uitgebreid bestudeerd. De meeste instellingen keken vluchtig ter voorbereiding op het bezoek van DEN. Een paar instellingen gaven zelfs tijdens het bezoek aan niet van het bestaan van de wiki op de hoogte te zijn, ondanks de algemene communicatie erover via de gebruikelijke kanalen en de voorafgaande aan het bezoek gestuurde documentatie waarin naar de wiki wordt verwezen.

Desgevraagd denken de bezochte instellingen dat alleen een beperkt aantal (grotere) erfgoedinstellingen in staat zal blijken de inhoud van de wiki te begrijpen en toe te passen. Dat geldt ook voor informatiespecialisten zoals de museumconsulenten.

Tot slot

Uiteindelijk is het doel voor de bezochte instellingen om de geselecteerde standaarden en best practises in de praktijk te brengen. Daarvoor is de wens om de kennis in het ICT-register in het algemeen en DE BASIS in het bijzonder te kunnen gebruiken in de vorm van overzichtelijke checklists. Een instelling noemt scenario's als manier om de toepasbaarheid van standaarden te verduidelijken.

Bibliotheken hebben een voorsprong ten opzichte van andere sectoren op het gebied van vindbaar maken van de eigen collectie en die van anderen.

Aansprekende voorbeelden van toepassing van DE BASIS binnen het erfgoed, zoals bijvoorbeeld DareNet voor de academische bibliotheken, zijn nodig voor een beter inzicht in de meerwaarde van het gebruik van standaardde.

DE BASIS zou onderdeel moeten zijn van de digitale Canon, zoals die onlangs door een consortium van de meest prominente erfgoedinstellingen in het project "Nederlands Erfgoed: Digitaal" is voorgesteld.