



Whitepaper Voorkeursformaten

4 juni 2020

Auteurs: Sam Alloing, Valentijn Gilissen, Hans Laagland, Marjolein Steeman, Walter Swagemakers



netwerk
digitaal
erfgoed

Inhoudsopgave

Samenvatting	3
Leeswijzer	4
1. Aanleiding	5
2. Probleemstelling	6
Belang van duurzaamheidsbeleid	6
Digitale conserveringsformaten	6
Voorkeursformaten	7
Conclusie	9
3. Oplossing: Wegwijzer Voorkeursformaten	10
Wat is de Wegwijzer?	10
Wat kun je vinden in de Wegwijzer?	10
Register met informatie over formaten (mens-en machineleesbaar)	11
Beheer Wegwijzer	12
Woordenlijst	14
Bijlagen	17
Gerelateerde initiatieven	17
Voorkeursformaten bij knooppunten binnen NDE	18
Colofon	20

Samenvatting

Informatie en drager vormen samen een tandem. Informatie heeft een drager nodig, of de drager een vel perkament is, een glasnegatief, of een PDF-bestand.

Werk je met digitale informatieobjecten, dan ontkom je er niet aan ook meer te weten over de dragers van de informatie, namelijk de bestandsformaten. Bestandsformaten vragen om beleid van de instelling: welke garanties wil of kan de instelling afgeven ten aanzien van duurzaam conserveren. En hoe communiceert de instelling daarover met aanbieders (of gebruikers) van het materiaal.

Het Netwerk Digitaal Erfgoed (NDE) werkt aan een online hulpmiddel om instellingen te helpen beleid te formuleren rond bestandsformaten. Dit online hulpmiddel wordt de Wegwijzer Voorkeursformaten. De Wegwijzer bestaat uit twee onderdelen: een deel dat via uitleg en voorbeelden instellingen handreikingen wil bieden voor opstellen van beleid, en een deel dat dient als register met informatie over bestandsformaten.

De Wegwijzer wordt beheerd door een Expertisegroep Voorkeursformaten die bestaande kennis in nationaal verband bundelt, deelt en verder ontwikkelt.

Leeswijzer

Alle **vetgedrukte** woorden in de tekst worden verklaard in de Woordenlijst, die aan het eind van deze whitepaper is toegevoegd. De whitepaper maakt zo veel mogelijk gebruik van Nederlandstalige terminologie.

1. Aanleiding

“Help! Ik ontvang een WordPerfect 5.1 bestand. Welke mogelijkheden zijn er met een WordPerfect 5.1 bestand als het gaat om duurzame toegankelijkheid van de informatie in het bestand?”

“Is het compacte JPEG2000 bestand geschikt om mijn scans langdurig mee te bewaren of kies ik liever voor het grotere TIFF?”

Als je digitale informatieobjecten langdurig gaat bewaren, is het ook nodig meer te weten over de dragers van de digitale informatieobjecten, namelijk de bestandsformaten. Bestandsformaten zijn er in allerlei soorten en maten. De Koninklijke Bibliotheek telde onlangs bij een inventarisatie al meer dan 1300 verschillende bestandsextensies. Probeer dan maar eens door de bomen het bos te zien: welk bestandsformaat is het meest geschikt voor langdurig bewaren van de informatie? Welk bestandsformaat gebruik ik wel of niet in mijn beleid voor duurzaamheid?

Het Netwerk Digitaal Erfgoed (NDE) werkt aan een oplossing in de vorm van een Wegwijzer. Deze Wegwijzer helpt instellingen bij het maken van keuzes omtrent bestandsformaten als onderdeel van hun duurzaamheidsbeleid.

Met deze whitepaper brengen we de problematiek rond bestandsformaten in kaart, stellen we een aanpak voor het beschrijven en vaststellen van voorkeursformaten en het beleid erop voor en schetsen we de opzet van de Wegwijzer, waarin aanpak en actuele informatie over voorkeursformaten gegeven wordt en deze actueel gehouden wordt. Het doel van de whitepaper is te toetsen of deze oplossingen aansluiten bij de behoeften van de instellingen in het netwerk. Met zoveel mogelijk feedback wordt duidelijk of een Wegwijzer voor instellingen nuttig kan zijn. Bovendien kunnen we de voorgestelde aanpak voor beleid op voorkeursformaten en de opzet van de Wegwijzer verbeteren.

De whitepaper wordt actief verspreid over instellingen, platform en diverse netwerken. Daarnaast wordt de whitepaper naar het Engels vertaald om zo ook internationaal input op te halen.

Deze whitepaper is een initiatief van de NDE-projectgroep *Preservation Watch en Preferred Formats* die sinds 2019 actief is. Deze projectgroep wil realiseren dat zoveel mogelijk instellingen Preservation Watch en Preferred Formats (**voorkeursformaten**) hebben ingebed in hun eigen preserveringsstrategie. Deze whitepaper richt zich op het onderdeel voorkeursformaten. In de loop van het jaar verschijnt een vervolg op deze whitepaper dat ingaat op de opzet van een gezamenlijke Preservation Watch.

2. Probleemstelling

Belang van duurzaamheidsbeleid

Maar eerst het belang van duurzaamheidsbeleid voor een organisatie. De Wegwijzer Duurzaamheidsbeleid van het NDE zegt hier het volgende over:

“Steeds meer collecties van archieven, bibliotheken, media, musea en kennisinstellingen komen digitaal en online beschikbaar. Informatieverlies ligt constant op de loer. Hardware veroudert, bij het kopiëren van bestanden worden fouten gemaakt, oude bestandsformaten blijken niet meer leesbaar, wijzigingen worden niet voldoende gedocumenteerd, enzovoort. Zonder maatregelen gaan digitale collecties onherroepelijk verloren. Beleid helpt. Een organisatie met een helder duurzaamheidsbeleid spreekt zich zichtbaar uit voor behoud van en langetermijntoegang tot haar digitale collectie. Beleid verbindt de missie en visie van een organisatie met de taken die nodig zijn om duurzame toegang te waarborgen, en vormt zodoende de basis voor het ontwikkelen van een preservingsstrategie en het rechtvaardigen van investeringen.”¹

Onderdeel van duurzaamheidsbeleid is beleid op bestandsformaten. Zonder dat een instelling het door heeft, groeit het aantal bestandsformaten dat een instelling in gebruik heeft. Maar wat weet de instelling over deze bestandsformaten? Welke bestandsformaten komen in aanmerking om de informatie in het bestand langdurig te kunnen bewaren? Op welke bestandsformaten past een instelling actief preservering toe? En welke garanties kan een instelling bieden op langdurige afspeelbaarheid of bruikbaarheid van een bestandsformaat? Kortom, welke keuze maak je als instelling om het ene formaat wel te accepteren voor je digitaal archief en het andere formaat niet?

Digitale preservingsformaten

Kiest een instelling voor **bitpreservering**, dan is het voldoende om te garanderen dat het bestand ‘as is’ kan worden uitgeleverd. Dat wil zeggen, tot op bitniveau identiek aan het bestand zoals die bij ingest is aangeboden en eventueel in combinatie met metadata over de aard van het bestand, voor zover vooraf meegeleverd. In de meest basale variant is het zelfs niet nodig om iets te weten over het bestandsformaat, laat staan eisen te stellen aan het bestandsformaat. Zolang het bestand terugvindbaar is en met een *checksum* de integriteit kan worden aangetoond.

Virus checking, identificatie van het bestandsformaat, karakteriseren en valideren kunnen echter ook onderdeel zijn van bitpreservering. Over het algemeen zijn deze activiteiten dan gericht op het verzamelen en documenteren van kennis. Daarin kunnen verschillende niveaus worden onderscheiden.

Naarmate het archief meer kennis heeft, is het ook in staat meer garanties te verstrekken over de toekomstige bruikbaarheid van de bestanden door bijvoorbeeld gebruik te maken van preservingsstrategieën (denk hierbij aan migratie en emulatie). In dat geval is sprake van een overgang naar **functionele preservering**. Zo wordt er extra zorg gedragen in de toekomst voor toegankelijkheid van bestanden, zowel in de organisatie als voor de doelgroep. De manier waarop een instelling gebruikswaarde van de bestanden definieert, kan sterk verschillen. Dit is namelijk

¹ Wegwijzer Duurzaamheidsbeleid: <https://wegwijzerduurzaamheidsbeleid.nl>

afhankelijk van de doelgroep, de voor hen gangbare hard- en software en de mogelijkheden of beperkingen van de eigen infrastructuur van de instelling voor het beschikbaar stellen van het bestand. Instellingen kunnen met betrekking tot dezelfde formaten dus tot andere keuzes komen.

Voorbeeld: kennisniveaus bij de Koninklijke Bibliotheek

De Koninklijke Bibliotheek (KB) heeft door gebruik te maken van kennisniveaus een beleid rond bestandsformaten opgesteld. Voor dit beleid werd gekozen omdat er niet opgelegd kan worden aan uitgevers welk formaat de uitgevers moeten aanleveren aan de KB. Daarom worden kennisniveaus gebruikt. Met kennisniveaus duidt de KB aan welk niveau van ondersteuning wordt gegeven voor een bestandsformaat. Het is een groeipad van bitpreserving tot functionele preserving. Niet elk bestandsformaat hoeft daarbij op het hoogste niveau (functionele preserving) te belanden. De kennisniveaus zijn een communicatiemiddel om duidelijk te maken in welke mate het bestandsformaat door de KB wordt ondersteund.

De bestandsformaten waar de instelling functionele preserving voor inricht of wil inrichten worden ook wel **preserveringsformaten** genoemd. Deze bestanden (de *masters*) staan het hoogst in de hiërarchie van een digitaal archief. Om het gebruik te faciliteren kan een instelling andere bestanden hiervan afleiden. Soms worden deze afgeleiden *on demand* vervaardigd op moment van uitlevering voor de gebruiker, maar vaak ook worden ze op voorhand vervaardigd en met evenveel zorg gepreserveerd als de oorspronkelijke *master*.

Voorbeeld: afgeleide bestanden bij Nederlands Instituut voor Beeld en Geluid

Het Nederlands Instituut voor Beeld en Geluid (B&G) maakt ter ondersteuning van het bestellen van fragmenten kopieën met een lage resolutie (**proxybestanden**). Vanwege de volumes en investeringen die daarbij spelen, worden de proxybestanden in het archief onder specifieke condities bewaard.

Voorbeeld: afgeleide bestanden bij Eye Filmmuseum

Niet alle preserveringsformaten zijn geschikt voor rechtstreeks uitlevering of gebruik. In zo'n geval wordt er vaak een tussenformaat (**mezzaninebestand**) gemaakt. Deze praktijk komt veel voor bij media-archieven waar de preserveringsmasters door hun grootte en hun aard niet direct te gebruiken zijn. Dit gebeurt bijvoorbeeld bij het Eye Filmmuseum.

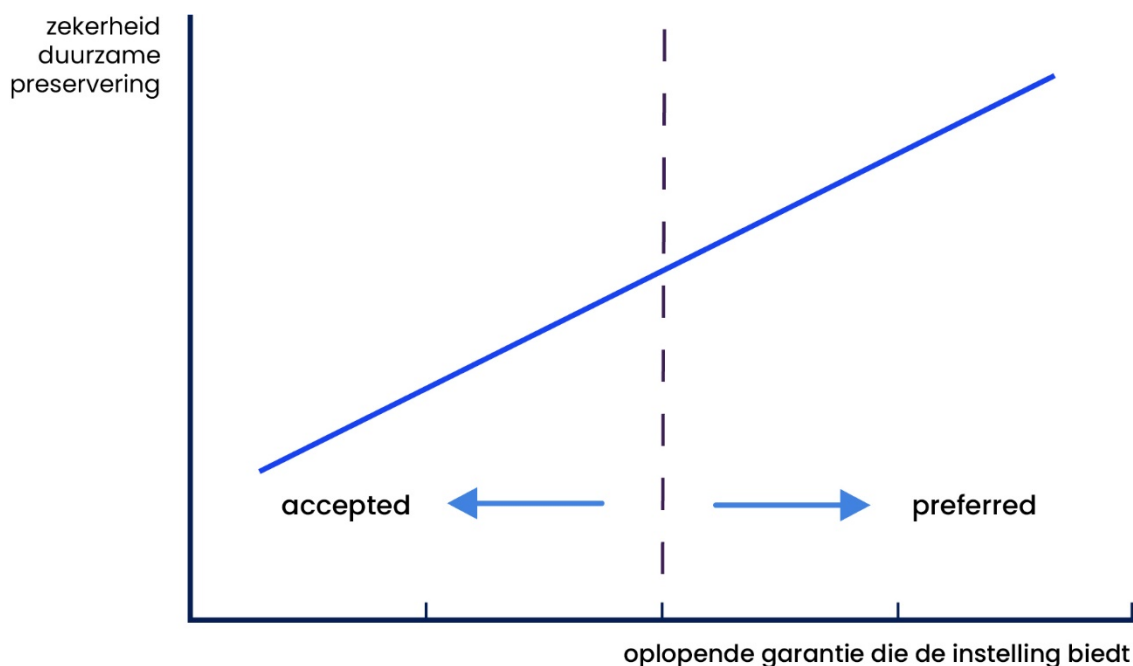
Voorkeursformaten

Afhankelijk van de manier waarop de instelling de preserving inricht, kan men ervoor kiezen bepaalde formaten aan te wijzen waarin de instelling digitaal materiaal bij voorkeur wenst te ontvangen. Deze voorkeursformaten worden ook wel "preferred formats" genoemd. Uit hetgeen hiervoor is gezegd over preserveringsformaten, zijn voor voorkeursformaten de volgende principes af te leiden:

- subjectiviteit: het aanwijzen van voorkeursformaten vindt plaats door de instelling vanuit de context en doelstelling van die specifieke instelling. Er bestaat niet zoiets als een objectief universeel lijstje voorkeursformaten

- tijdgebonden: onder invloed van technische en commerciële ontwikkelingen kan de status van een formaat veranderen. In de praktijk zal dit betekenen dat een instelling een formaat alsnog aanwijst als voorkeur (bijvoorbeeld wanneer er voldoende ondersteuning voor het formaat is gekomen) of dat een formaat het predikaat voorkeur verliest. In dat laatste geval zal de instelling ook moeten nadenken over de consequenties voor de bestanden die in dat formaat aanwezig zijn en de garanties die daarover zijn afgegeven. Dit veranderende karakter van voorkeursformaten zal in de gaten gehouden moeten worden in een **Preservation Watch**.
- garantie stellend: een voorkeursformaat garandeert altijd een bepaalde zekerheid ten aanzien van duurzame preservering. Doorgaans zal dat een garantie zijn waarbij de bestanden een grotere zekerheid genieten in termen van duurzaamheid. Maar in principe kan een instelling elke garantie aan een of meer voorkeursformaten koppelen. Bij een lagere zekerheid spreekt men in dit verband ook wel van **acceptabele formaten**.²
- communicatief: het opstellen van voorkeursformaten heeft als doel de leveranciers (en gebruikers) van archiefmateriaal te informeren over hoe het archief met bepaalde formaten omgaat. Soms heeft het een voorschrijvend karakter (dwingend) maar het kan ook volstrekt neutraal zijn (zoals de kennisniveaus van de KB).

Uit onderstaande figuur volgt dat naarmate de zekerheid over duurzame preservering groter is, men eerder spreekt van voorkeursformaten. Is deze zekerheid minder groot, dan kan het archief ervoor kiezen de bestanden als acceptabel formaat op te nemen, bijvoorbeeld omdat een bepaald formaat in de praktijk veel wordt gebruikt. Waar de breuklijn precies ligt, zal per instelling verschillen.



Figuur 1: relatie tussen eigenschappen formaat en geboden garantie door instelling

Op de verticale as staat de mate van zekerheid ten aanzien van duurzame preservering die geldt voor het formaat. In het algemeen is wel een aantal eigenschappen³ te noemen op basis waarvan deze zekerheid toeneemt:

² Handreiking voorkeursformaten NA: <https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-voorkeursformaten-nationaal-archief>

³ DANS Bestandsformaten: <https://dans.knaw.nl/nl/over/diensten/easy/toelichting-data-deponeren/voor-het-deponeren/bestandsformaten>

- algemeen gebruikt: het bestandsformaat mag niet obscuur zijn, want anders zijn er weinig tools beschikbaar en is er waarschijnlijk ook beperkte ondersteuning in de verschillende mogelijkheden van bijvoorbeeld migratie.
- open specificaties: als het formaat een (open) standaard of een open specificatie is, dan kan in het slechtste geval - wanneer de software niet meer ondersteund wordt door een software leverancier - toch nog met behulp van alternatieven ondersteuning geboden worden. Het is ook mogelijk om het bestandsformaat op een technisch niveau beter te begrijpen.
- software onafhankelijkheid: hierdoor is er minder gevaar dat een leverancier failliet gaat en er geen ondersteuning meer is voor het bestandsformaat en dat bepaalde functies niet meer uitgevoerd kunnen worden.

Op de horizontale as staat de oplopende garantie die de instelling biedt. Denk hierbij aan het eerder genoemde onderscheid tussen bitpreserving en functionele preserving en de verschillende niveaus die een instelling daarbinnen kan inrichten.⁴

De archiverende instelling bepaalt uiteindelijk zelf aan de hand van het preservingbeleid welke formaten men accepteert of prefereert en welke garanties men daarbij afgeeft. Elk formaat kan als het ware als een stip in de grafiek worden 'geplot'. De verwachting is dat hoe hoger de punt in de figuur, hoe eerder de instelling genegen is dit als een voorkeursformaat te bestempelen. Daarbij kunnen instellingen overigens aanvullende eisen stellen ten aanzien van een bestandsformaat. In het algemeen geldt dat de eisen waaraan het formaat moet voldoen in het rechter spectrum strikter worden. De linker helft van het spectrum kent navenant ruimere, algemene specificaties of eisen.

Dit kan inhouden dat instellingen ruimhartig een brede range aan te conserveren formaten accepteren, al of niet in combinatie met het gecontroleerd omzetten naar mezzanines of proxybestanden, zoals in de eerdere voorbeelden van B&G of Eye Filmmuseum. En er zijn archieven die de mogelijkheid hebben om vooraf met een archiefvormer of een digitaliseerder afspraken te maken over de aan te leveren formaten. Binnen deze afspraken kunnen dan ook digitale objecten worden geweigerd die niet aan de eisen of specificaties van de archieven voldoen.

Conclusie

In de hierboven al genoemde Wegwijzer Duurzaamheidsbeleid wordt aangeraden om als onderdeel van functionele preserving een bestandsformatencheck⁵ te doen. De bestandsformatencheck is een inventarisatie van de bestandsformaten in een digitaal archief. Daarbij wordt ook gekeken naar de mogelijkheden tot afspeelbaar of bruikbaar houden van deze bestandsformaten op de lange termijn. Vervolgens zullen instellingen de uitkomsten van de check moeten vertalen naar eigen duurzaamheidsbeleid, inclusief de communicatie over de omgang met deze bestandsformaten. De afweging die instellingen hierbij moeten maken is complex, is context afhankelijk en vergt onderzoek.

Op dit moment voeren instellingen dit onderzoek individueel uit. Instellingen zullen niet altijd in staat zijn om hun duurzaamheidsbeleid op dit punt goed uit te werken. Er is behoefte aan samenwerking bij onderzoek naar formaten en het kunnen profiteren van elkaars kennis.

⁴ Zie ook levels-of-digital-preservation van de NDSA en dan in het bijzonder met betrekking tot Content: <https://ndsa.org/publications/levels-of-digital-preservation/>

⁵ Wegwijzer Duurzaamheidsbeleid > Bestandsformatencheck: <https://wegwijzerduurzaamheidsbeleid.nl/artikelen/bestandsformatencheck/>

3. Oplossing: Wegwijzer Voorkeursformaten

De Wegwijzer Voorkeursformaten wil instellingen helpen om keuzes te maken in het opstellen van beleid rond (voorkeurs)formaten. Dit doet de Wegwijzer op twee manieren: aan de ene kant biedt de Wegwijzer inzicht in (on)zekerheid omtrent duurzame preservering van bestandsformaten. Aan de andere kant geeft de Wegwijzer bij wijze van een register een overzicht van verschillende formaten met daarbij aanvullende informatie. De Wegwijzer wordt gerealiseerd in het NDE-project *Preservation Watch en Preferred Formats*. Een (nog op te richten) nationale Expertisegroep Voorkeursformaten zal de Wegwijzer up to date houden.

Wat is de Wegwijzer?

De Wegwijzer wil geen universeel lijstje van voorkeursformaten geven, maar een hulpmiddel zijn voor instellingen om hun eigen bestandsformatenbeleid op te stellen op basis van eigen ervaringen en kennis en op basis van ervaringen en kennis van andere instellingen. Zo hoeven instellingen niet het wiel opnieuw uit te vinden, maar kunnen ze verder bouwen op bestaande kennis en ervaringen. Daarnaast is het belangrijk dat andere instellingen kunnen bijdragen aan de uitbouw van de Wegwijzer. Instellingen kunnen via de Wegwijzer een bijdrage leveren aan de verdere ontwikkeling van kennis op (voorkeurs)formaten door kennis en ervaringen te delen. Hiermee wordt de Wegwijzer een open omgeving.

Wat kun je vinden in de Wegwijzer?

De Wegwijzer:

- legt uit wat voorkeursformaten, preserveringsformaten, **gebruiksformaten**, acceptabele formaten en preserveringsniveaus zijn
- geeft voorbeelden van acceptabele formaten en voorkeursformaten bij instellingen van het netwerk
- geeft uitleg waarom bepaalde formaten gekozen worden als voorkeursformaten of waarom ze tot een bepaald preserveringsniveau behoren
- biedt instellingen aan om zelf op basis van de Wegwijzer hun eigen bestandsformatenbeleid vast te leggen, inclusief voorkeursformaten.

De tool beoogt zoveel mogelijk bestaande informatie te hergebruiken en te bundelen. Als bron voor bestandsinformatie worden onder andere **PRONOM**⁶ en **Wikidata**⁷ gebruikt. Er wordt een aantal eigenschappen getoond, zoals: is dit formaat een (open) standaard of open specificatie, is het formaat wijdverspreid, zijn er meerdere applicaties die het bestandsformaat ondersteunen en zijn er ontwikkelaars of leveranciers die het bestandsformaat ondersteunen met hun applicaties. Daarnaast wordt ook aangegeven welke instellingen van het netwerk een voorkeursformaat hanteren, onder welke restricties en eventueel in combinatie met welke gebruiksformaten. Er wordt ook een

⁶ PRONOM: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

⁷ Wikidata: <https://www.wikidata.org> en <https://wikidp.org>

verbinding gemaakt met **COPTR**. De Wegwijzer ontwikkelt zich zo tot een gezamenlijke kennisbank, en dat is een basisvoorwaarde bij de evolutie naar een nationaal preservation netwerk.

Als toegang kunnen bestandsformaten getoond worden met bestandsformaatnaam en Pronom Identifier (ook wel afgekort als **PUID**)⁸. Voor sommige bestandsformaten is deze identificatie niet voldoende, maar moeten ook de eigenschappen van bestandsformaten gebruikt worden voor de definitie van voorkeursformaten, bijvoorbeeld bit rate bij videoformaten. Voor het toevoegen van deze eigenschappen wordt de verbinding met Wikidata gebruikt. Ook versies van bestandsformaten kunnen van belang zijn.

Ook moet inzichtelijk zijn welke eigenschappen het bestandsformaat ondersteunt: dit noemen we de **essentiële kenmerken**. Een goed voorbeeld om dit toe te lichten, is tabellen of spreadsheets. Biedt het bestandsformaat mogelijkheden om tabellen op te slaan met inkleuring van cellen en gebruik van formules of de opties om de gegevens te sorteren of te filteren.

In de Wegwijzer zal een stappenplan te vinden zijn om te komen tot beleid op voorkeursformaten. Dit stappenplan is gebaseerd op de al eerder genoemde bestandsformatencheck uit de Wegwijzer Duurzaamheidsbeleid⁹, waaraan twee stappen zijn toegevoegd, specifiek gericht op voorkeursformaten:

1. inventariseren van de bestandsformaten in een digitaal archief.
Afhankelijk van de software die gebruikt wordt voor het digitaal archief kan hiervoor de bestandextensie gebruikt worden of zijn bestandsidentificatietools handige hulpmiddelen.
2. voorkeursformaten vaststellen¹⁰
3. kenmerken van bestandsformaten vaststellen
4. procedure opstellen valideren tijdens opname in het digitaal archief
5. procedure opstellen valideren voor migratie
6. procedure opstellen valideren na migratie

Register met informatie over formaten (mens-en machineleesbaar)

Zoals gesteld in de Aanleiding beoogt de Wegwijzer zowel inzicht in kennis rondom bestandsformaten als ook een overzicht van verschillende formaten te geven.

In het kort: het register wordt een inventarisatie van (gangbare) bestandsformaten met per bestandsformaat een beschrijving van de duurzaamheid. Deze beschrijving betreft niet alleen technische implicaties maar ook organisatorische implicaties, met eraan toegevoegd het gebruik van een bestandsformaat door de betreffende sector op basis van een bepaald kennisniveau over het bestandsformaat.

Uit het overzicht van formaten moeten gebruikers helder kunnen aflezen met wat voor bestandsformaat zij te maken hebben: typering en classificatie, een technische beschrijving, of het bestand open is gespecificeerd of het bestand gelieerd is aan bepaalde software.

Het moet duidelijk zijn waar een bepaald bestandsformaat toe dient. Ook moet inzichtelijk zijn welke eigenschappen het bestandsformaat ondersteunt: de essentiële kenmerken. Ook moet duidelijk zijn welke software in staat is om een bepaald formaat te openen. Als een bestand meerdere versies

⁸ Als voorbeeld kan gekeken worden naar het register van DANS (<https://dans-labs.github.io/formats/>). Zie ook Bijlage 1

⁹ Wegwijzer Duurzaamheidsbeleid: <https://wegwijzerduurzaamheidsbeleid.nl/>

¹⁰ Zie hiervoor paragraaf Voorkeursformaten in Hoofdstuk 1 Probleemstelling

kent, moet helder worden wanneer het verschil in versies van belang is voor de duurzame toegankelijkheid.

Door de integratie met PRONOM en Wikidata wordt het register met voorkeursformaten niet alleen voor de mens leesbaar maar ook voor de machine. De Wegwijzer kan systematisch aangeropen worden (machineleesbaar) voor nadere koppeling aan interne en externe databases en kennisbanken. Hieronder valt ook de koppeling met het NDE Termennetwerk.¹¹

Essentieel bij een laagdrempelig te gebruiken register is streven naar een bruikbaar niveau van **granulariteit**. Bedenk hierbij als voorbeeld de bestandsextensies TIFF en ZIP. In de meest ruwe granulariteit betreft dit respectievelijk een bestandsformaat voor het opslaan van afbeeldingen en een compressieformaat waar een of meerdere bestanden in worden gepakt. Op dit basale niveau bestaat het risico om essentiële informatie te missen, bijvoorbeeld eigenschappen die van een bepaalde versie afhankelijk zijn. Een zoekopdracht in de database van Wikidata op de extensies TIFF of ZIP toont een zeer hoge granulariteit: hier geeft een zoekopdracht de resultaten voor alle bestaande versies en typen van deze formaten. De Wegwijzer zoekt een middenweg in deze granulariteit.

Een beschrijving van alle subtypes van een formaat biedt weinig meerwaarde op bestaande registers zoals Wikidata en zal de gebruiker van de Wegwijzer geen bruikbare ingang bieden. Het ligt meer voor de hand om de informatie over TIFF op het algemene niveau van de .tif-extensie te vinden dan om gericht informatie te zoeken op een PRONOM-nummer van een specifieke versie zoals 'fmt/353'.

Het uitgangspunt van het register in de Wegwijzer moet daarom op een algemeen niveau komen te liggen. Informatie over verschillende versies of van versies afhankelijke adviezen moeten binnen het algemene niveau helder worden gemaakt; gedetailleerde technische informatie met betrekking tot bestaande versies kan door middel van het leggen van koppelingen naar PRONOM worden verzameld.

Beheer Wegwijzer

Kennis over bestandsformaten is in Nederland niet gebundeld over de sectoren heen, los van afzonderlijke initiatieven tussen instellingen of knooppunten.¹² Deze bundeling is wel wenselijk: het behalen van een bepaald kennisniveau op bestandsformaten is intensief en tijdrovend. Een netwerkvoorziening die gebruik maakt van internationale kennis over bestandsformaten, overbrugt de afstemming tussen sectoren.

Om deze overbrugging tot stand te brengen, zullen we bestaande kennis niet alleen bundelen via een Wegwijzer maar ook in de vorm van een (op te richten) *Expertisegroep Voorkeursformaten*. Hierin wordt de kennis en expertise met betrekking tot voorkeursformaten in nationaal verband gedeeld en verder ontwikkeld. De expertisegroep:

- deelt kennis, ervaring en aanpak
- draagt bij aan deskundigheidsbevordering
- zorgt voor afstemming tussen sectoren en instellingen op basis van een gemeenschappelijke agenda
- draagt bij aan het beheer en verdere ontwikkeling van de Wegwijzer

¹¹ NDE Termennetwerk: <https://www.netwerkdigitaalervoed.nl/kennis-en-voorzieningen/digitaal-erfgoed-bruikbaar/termennetwerk/>

¹² Knooppunten binnen NDE zijn DANS KNAW, Koninklijke Bibliotheek, Nationaal Archief, Nederlands Instituut voor Beeld & Geluid en Rijksdienst voor Cultureel Erfgoed

Via de Wegwijzer Voorkeursformaten en de Expertisegroep draagt het NDE bij aan het vergroten van kennis aan de informatieprofessional: van de gemeente die haar bouwvergunningen wil digitaliseren en de scans duurzaam toegankelijk wil maken, tot aan de onderzoeker die er zeker van wil zijn dat de door hem of haar gedeponeerde onderzoeksdata ook over een langdurige periode beschikbaar zijn voor reproduceerbaarheid van het onderzoek.

Woordenlijst

Verklarende woordenlijst voor onderwerpen over digitale preservering, en in het bijzonder voorkeursformaten. De woordenlijst gebruikt Nederlandstalige terminologie met de Engelstalige begrippen tussen haakjes om ook (vervolg)zoekacties op internet met de Engelstalige termen mogelijk te maken.

Acceptabele formaten (accepted format)

Dit zijn formaten die door de instelling worden opgenomen in het digitale archief onder een bepaalde (minimale) set garanties voor wat betreft de bruikbaarheid, toegankelijkheid en robuustheid op de lange termijn.

Bitpreservering (bit preservation)

Bit preservering omvat de werkzaamheden die nodig zijn om de bitstreams (de oorspronkelijke volgorde van nullen en enen) intact en leesbaar te houden.

(bron: <https://wegwijzerduurzaamheidsbeleid.nl/artikelen/bit-preservering/>)

COPTR

Community Owned digital Preservation Tool Registry (COPTR) beschrijft tools die bruikbaar zijn voor lange termijn digitale preservering.

(https://coptr.digipres.org/Main_Page)

Essentiële kenmerken (significant characteristics/ significant properties)

De essentiële kenmerken van een informatieobject zijn die eigenschappen die het uiterlijk, gedrag, de kwaliteit en bruikbaarheid bepalen en die moeten worden gepreserveerd om het digitale object in de tijd toegankelijk en betekenisvol te houden.

(<https://lerenpreserveren.nl/topic/essentiele-kenmerken/>)

Functionele preservering (functional preservation)

Functionele preservering is bedoeld om – door de tijd heen – de duurzame toegankelijkheid van digitale bronnen te waarborgen, door maatregelen te nemen om informatieverlies door technologische veranderingen te voorkomen.

(<https://wegwijzerduurzaamheidsbeleid.nl/artikelen/functionele-preservering/>)

Gebruiksformaat

Met gebruiksformaat wordt bedoeld raadpleegformaat. Het betreft een afgeleide van de preservingsmaster.

Granulariteit (granularity)

In de context van Voorkeursformaten wordt met 'granulariteit' bedoeld: het niveau van indeling van bestandsformaten en mate van detail in beschrijving. Op de meest basale granulariteit wordt een bestandsformaat afgaande aan de extensie genoemd. De meest ruime granulariteit beschrijft elke versie van het formaat in uitgebreid detail.

Mezzaninebestanden

Zie: [Proxybestand](#)

Preserveringsformaat (preservation format)

Preserveringsformaten zijn bestandsformaten waar de instelling functionele preservering voor inricht of wil inrichten. Deze bestanden (de *masters*) staan het hoogst in de hiërarchie van een digitaal archief.

Preservation watch

Preservation Watch betreft het volgen van vernieuwingen op het gebied van opslagmedia, bestandsformaten, manieren van presentatie en andere (technologie)wijzigingen om zo de authenticiteit, bruikbaarheid en duurzame toegang van digitale collecties te kunnen garanderen. Naast technologie wijzigingen kunnen andere veranderingen in de omgeving van het digitale archief een bedreiging vormen voor de duurzame toegankelijkheid. Denk daarbij aan budgetten, veranderende gebruikerseisen, etc.

(bron: https://wegwijzerduurzaamheidsbeleid.nl/artikelen/preservation_watch/)

PRONOM

PRONOM (<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>) is een database van The National Archives met informatie over alle – bij hen – bekende bestandsformaten en de ondersteunende softwareproducten. Het bevat informatie over softwareproducten en de bestandsformaten die door deze producten gelezen kunnen worden. Je kunt zoeken op bestandsformaat, software, softwareleverancier en levenscycli (je kunt zoeken welke softwareproducten op een bepaald moment in de tijd ondersteund werden).

(bron: <https://lerenpreserveren.nl/topic/duurzame-bestandsformaten/>)

Proxybestand (proxy file)

Proxybestanden zijn afgeleid van de bestanden waarop preservering wordt uitgevoerd ten behoeve van langdurig bewaren (preserveringsmasters of archiefmasters). Proxybestanden worden gebruikt voor doeleinden zoals raadplegen of leveren van de inhoud. Deze files hebben een relatief lage resolutie en zijn niet geschikt voor hergebruik.

Naast proxybestanden zijn er ook de **mezzaninebestanden**. Mezzaninebestanden of mezzanines zijn tussenformaatbestanden die worden afgeleid van de preserveringsmaster. Dit tussenformaat is feitelijk identiek is aan de preserveringsmaster en functioneren in de praktijk als een soort tweede preserveringsmaster.

(bron: <https://publications.beeldengeluid.nl/pub/387>)

PUID

PUID staat voor PRONOM Unique IDentifier. Een PUID is een persistente en unieke identifier naar een record in PRONOM database.

Voorkeursformaat (preferred format)

Een voorkeursformaat is het bestandsformaat waarin een digitaal archief – dat zich richt op duurzame toegang – digitale objecten bij voorkeur opneemt.

(bron: <https://lerenpreserveren.nl/woordenlijst/voorkeursformaat/>)

Wikidata

Wikidata (<https://www.wikidata.org>) is een gezamenlijk bewerkte kennisdatabase die is opgezet om Wikipedia te ondersteunen. Het is de gemeenschappelijke bron voor bepaalde types gegevens, bijvoorbeeld geboortedata, die veel gebruikt worden in artikelen op Wikipedia en elders.

Wikidata for Digital Preservation is een portaal dat specifiek gericht is raadplegen van gestructureerde data en toevoegen van gestructureerde data over software en bestandsformaten aan Wikidata:

(bron: <https://www.wikimedia.nl/pagina/wikipedia-en-meer>)

(bron: <https://wikidp.org/about>)

Bijlagen

Gerelateerde initiatieven

DANS Lab op GitHub: Preferred Formats

Vanuit DANS wordt een GitHub opgezet¹³ voor het samenbrengen van expertise, kennis en richtlijnen rondom voorkeursformaten op internationaal niveau. Het doel van dit platform is om tot een collaboratieve, multi-instelling infrastructuur te komen, om informatie met betrekking tot voorkeursformaten in te voegen, bij te werken, en gestructureerd en gestandaardiseerd te exporteren naar andere systemen. Het platform moet aansluiten bij bestaande voorzieningen en biedt de mogelijkheid voor experts van verschillende instellingen om hun expertise in bepaalde formaten in te brengen. Algemene informatie vanuit PRONOM, Wikidata zal ook in de GitHub verzameld worden. Het levert daardoor een informatiemodel op met objectieve/ karakteristieke onderdelen per formaat: naam formaat, extensies, relatie, type formaat, significante eigenschappen, openheid, ondersteuning, verwijzing naar documentatie en kennisniveau, identifier, etc. Elke instelling die iets over het formaat te zeggen heeft, kan daar aan gekoppeld worden of zelf een bijdrage aanleveren vanuit perspectief van die instelling. De informatie uit GitHub kan machine-leesbaar doorlinken naar externe bronnen zoals PRONOM maar ook het overzicht van Digital Preservation Coalition van tooling (COPTR). Voor gebruikers kan de inhoud uit GitHub mens-leesbaar gerendered worden naar HTML of via SKOS worden geëxporteerd.

Een greep uit (inter)nationale initiatieven

Er bestaat al veel kennis rondom formaten en de specificaties daarvan. Hieronder een greep uit enkele (inter)nationale initiatieven:

- Wikidata, i.h.b. Wikidata for Digital Preservation¹⁴
- PRONOM¹⁵
- Library of Congress > format descriptions¹⁶
- U.S. National Archives > Tables of File Formats¹⁷
- Preforma (PREservation FORMAts for culture information/e-archives) > Project Overview¹⁸
- Archaeology Data Service / Digital Antiquity > Guides to Good Practice¹⁹
- Stadsarchief Rotterdam > Bestandsformaten E-depot²⁰
- Digital Preservation Coalition > Knowledge Base: File Formats²¹

¹³ DANS Lab op GitHub: <https://dans-labs.github.io/formats/>

¹⁴ <https://www.wikidata.org> en <https://wikidp.org/>

¹⁵ <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

¹⁶ <https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>

¹⁷ <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>

¹⁸ <http://www.preforma-project.eu/project.html>

¹⁹ <https://guides.archaeologydataservice.ac.uk/g2gpwiki/>

²⁰ <https://stadsarchief.rotterdam.nl/diensten/e-depot/bestandsformaten-e-depot/>

²¹ <https://www.dpconline.org/knowledge-base/tags/109-file-formats>

Voorkeursformaten bij knooppunten binnen NDE

Hoe doen anderen “het”? Welke keuzes maken andere instellingen bij hun beleid op (voorkeurs)formaten. Als “voorproefje” op de Wegwijzer staat hieronder beknopt het beleid op (voorkeurs)formaten van vier van de vijf knooppunten binnen het NDE: DANS KNAW, Koninklijke Bibliotheek, Nationaal Archief en Nederlands Instituut voor Beeld & Geluid.

DANS KNAW

Preferred formats zijn de bestandsformaten waarvan DANS het vertrouwen heeft dat deze op de langere termijn de beste garanties bieden qua bruikbaarheid, toegankelijkheid en duurzaamheid. Het deponeren van onderzoeksdata in preferred formats zal zonder meer door DANS worden geaccepteerd. [...] Als algemene richtlijn stelt DANS dat de bestandsformaten die het beste geschikt zijn voor duurzaamheid en toegankelijkheid op de lange termijn:

- veel worden gebruikt
- open specificaties hebben
- onafhankelijk zijn van specifieke software, ontwikkelaars of leveranciers

In de praktijk blijkt het niet altijd mogelijk om formaten te gebruiken die voldoen aan al deze kenmerken.²²

Koninklijke Bibliotheek

Het Digitaal Magazijn bevat heel wat bestandsformaten, maar de controle en identificatie van deze bestandsformaten is beperkt. [...] Daarom is het belangrijk om beter de bestanden in het Digitaal Magazijn te controleren en te identificeren.

Om dit te kunnen doen, is het belangrijk om bestandsformaten beter te kunnen identificeren, controleren, valideren, extraheren van technische metadata, Hiervoor is meer kennis over bestandsformaten nodig in de praktijk. Door het grote aantal bestandsformaten, is het niet mogelijk om dit onmiddellijk voor alle bestandsformaten te doen. Daarom wordt er met kennisniveaus gewerkt. Zo kan duidelijk gecommuniceerd worden over de kennis van bestandsformaten en welke formaten opgeslagen worden en welke bewaard worden aan de designated community.

Per bestandsformaat wordt de status bijgehouden en wordt ook bijgehouden welke stappen nog gezet moeten worden om naar het volgende kennisniveau te stijgen:

1. Kennisniveau ‘opgeslagen bestandsformaat’
Bij opgeslagen bestandsformaten worden enkel checksumcontroles uitgevoerd om te kijken of er geen bit rot optreedt. Er is weinig bekend over het formaat aangezien het bestandsformaat niet geïdentificeerd kan worden.
2. Kennisniveau ‘geïdentificeerd bestandsformaat’
Een geïdentificeerd bestandsformaat heeft een PRONOM ID en er is dus ook een tool die bestanden kan herkennen en een PRONOM ID kan toevoegen. Het is de basisstap voor het duurzaam bewaren. Het is een eerste stap richting functional preservation, maar is nog onvoldoende om van functionele preservation te spreken. Maar is meer dan enkel bit preservation, bit preservation.
3. Kennisniveau ‘gekend bestandsformaat’
Bij een gekend bestandsformaat is goede bewaring mogelijk aangezien de KB de resultaten van de identificatie, validatie en technische metadata extractie kan interpreteren en richtlijnen

²² Bron: <https://dans.knaw.nl/nl/over/diensten/easy/toelichting-data-deponeren/voor-het-deponeren/bestandsformaten>

heeft opgesteld voor elk formaat. [...] Met deze informatie kan het bestand als duurzaam bewaard beschouwd worden en wordt functional preservation uitgevoerd.²³

Nationaal Archief

Het Nationaal Archief krijgt digitale informatie voor het e-Depot het liefst aangeleverd in ‘open’ formaten. Naast deze voorkeursformaten onderscheidt het Nationaal Archief ook acceptabele formaten.

- voorkeursformaten: dit zijn de ‘open’ formaten. Daarnaast kan het zijn dat voor bepaalde bestandstypes zogeheten ‘industry standards’ zijn die massaal worden gebruikt en voldoende gedocumenteerd zijn.
- acceptabele formaten: dit zijn formaten die niet (volledig) ‘open’ zijn, maar die toch acceptabel zijn voor het Nationaal Archief. De reden hiervoor kan zijn dat het Nationaal Archief (of andere (inter)nationale instanties) ervaring heeft met een bepaald formaat, en er een goede preservationstrategie voor heeft.

Een ‘open standaard’ omvat volgens het Forum Standaardisatie een laagdrempelige beschikbaarheid van documentatie, geen hindernissen op basis van intellectuele eigendomsrechten (zoals patent royalties), inspraakmogelijkheden en onafhankelijkheid en duurzaamheid van de standaardisatie-organisatie.²⁴

Nederlands Instituut voor Beeld en Geluid

Nederlands Instituut voor Beeld en Geluid wijst een beperkt aantal formaten aan voor duurzame preservering. Voor deze formaten kan het Digitaal Archief welomschreven duurzaamheidsgaranties afgeven. De organisatie hanteert vaststaande criteria voor het bepalen van deze preserveerbare formaten: het moet gaan om een goed gedocumenteerde industriestandaard die werkt op courante software, zoals die in het audiovisuele domein wordt gebruikt. Het formaat moet kunnen worden geïndexeerd binnen de technische- en catalogusinfrastructuur van Beeld en Geluid, zodat er afgeleide bestanden van gemaakt kunnen worden ten behoeve van viewing en uitlevering. De standaard moet dus ondersteuning bieden aan media-gerelateerde functionaliteit zoals tijdcode, ondertiteling en metadata. Daarnaast moet het formaat met courante transcodingssoftware kunnen worden getranscodeerd naar andere, gangbare formaten én er moeten, met behulp van de standaard analysesoftware, kwaliteitsanalyses op uitgevoerd kunnen worden.²⁵

²³ Bron: Interne richtlijn bestandsformaten Koninklijke Bibliotheek

²⁴ Bron: <https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-voorkeursformaten-nationaal-archief>

²⁵ Bron: <https://publications.beeldengeluid.nl/pub/387>

Colofon

Sam Alloing
Digital Preservation Officer - Koninklijke Bibliotheek

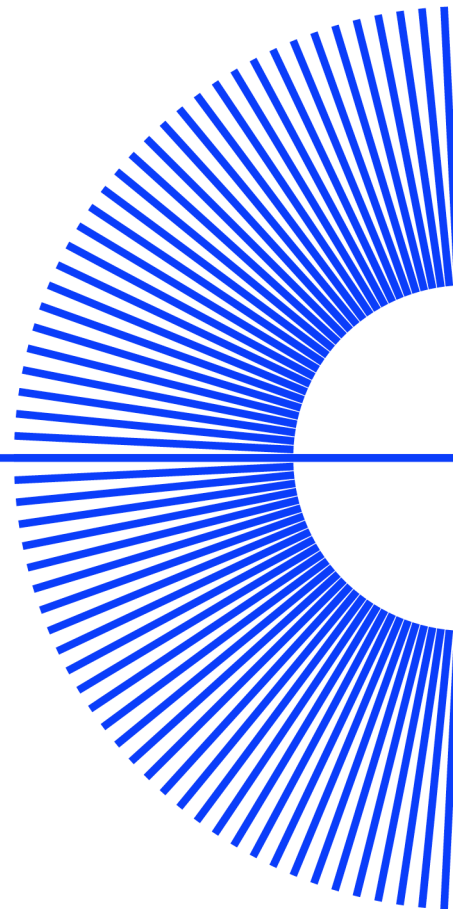
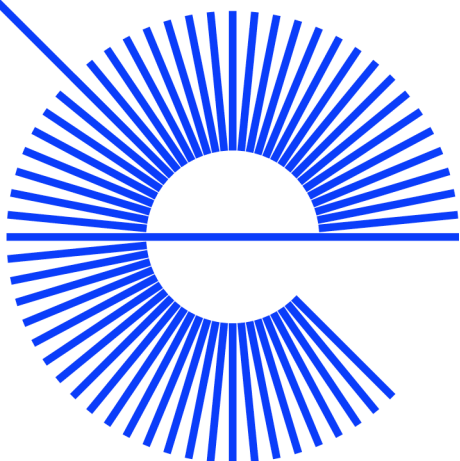
Valentijn Gilissen
Senior Data Manager, Preservation Officer, Data Steward - Data Archiving and Networked Services

Hans Laagland
Opdrachtregisseur Duurzame Toegankelijkheid - Tresoar, Regionaal Historisch Centrum

Marjolein Steeman
Information Officer Preservering - Nederlands Instituut voor Beeld en Geluid

Walter Swagemakers
Senior Projectleider Collectie - Eye Filmmuseum

Dit is een uitgave van het Netwerk Digitaal Erfgoed, juni 2020.
Meer informatie is te vinden op www.netwerkdigitaal erfgoed.nl
Reacties zijn welkom via: info@netwerkdigitaal erfgoed.nl



**netwerk
digitaal
erfgoed**