

Het E-depot: digitale archivering en de bouw van het Depot van Nederlandse Elektronische Publicaties

Trudi Noordermeer ¹

Hoofd Nederlandse Bibliografie
Koninklijke Bibliotheek
Den Haag
trudi.noordermeer@kb.nl
<http://www.kb.nl>

Inleiding

De Koninklijke Bibliotheek (KB) in Den Haag heeft als nationale bibliotheek van Nederland een depotfunctie. Dat betekent dat de KB 'alles' verzamelt wat in Nederland uitkomt, in uitgestrekte depots opslaat en goed bewaart voor het nageslacht. Deze depotcollectie mag alleen in de KB worden ingezien. De KB is ook een wetenschappelijke bibliotheek met de nadruk op humaniora, in het bijzonder geschiedenis, taal en cultuur van Nederland. De KB bezit enkele tientallen bijzondere collecties, die zij in de loop der tijd door schenkingen, legaten en aankoop heeft verworven. Daaronder bevinden zich unieke middeleeuwse handschriften, maar ook strips, kookboeken en affiches.

Vanaf het midden van de jaren tachtig ontvangt de KB ook elektronische publicaties voor de depotcollectie. Het begon met diskettes en later onder andere CD-Is, CD-ROMs, online elektronische tijdschriften van wetenschappelijke uitgevers en rapportliteratuur, de 'grijze literatuur' van universiteiten en overheidsinstellingen. Het Internet bevat ook veel publicaties die voldoen aan de selectiecriteria van het depot.

Vanaf 1993 realiseerde de KB zich dat elektronische publicaties belangrijk cultureel erfgoed zijn en dat ze net als de gedrukte publicaties een plaats moesten krijgen in het depot en de Nederlandse Bibliografie, de 'Brinkman's Cumulatieve Catalogus'. Deze nationale bibliografie publiceert de bibliografische beschrijvingen, de metadata, over alle publicaties die in het depot zijn opgenomen. Het Algemeen Bestuurs College van de KB besliste in 1995 dat elektronische publicaties ook in het depot opgenomen moesten worden: de start van het DNEP, Depot van Nederlandse Elektronische Publicaties. De nieuwe naam van de operationele dienst is E-depot.

¹ Met dank aan Raymond van Diessen, Betty Nieuwenburg, Johan Steenbakkers en Titia van der Werf

Problematiek en strategie van de KB

Er verschijnen steeds meer elektronische publicaties op allerlei gebied. En het is dus de taak van een nationale bibliotheek om ook dit deel van het cultureel erfgoed duurzaam te bewaren voor studie en onderzoek en ook op lange termijn toegankelijk te houden.

Deponering en bewaring van elektronische publicaties brengen nieuwe problemen met zich mee. Ten eerste is het veel moeilijker vast te stellen of er sprake is van een Nederlands product. Het geografische criterium biedt bij elektronische publicaties vaak weinig houvast. Daarom moeten er internationaal andere depotafspraken worden gemaakt. Een veel groter probleem is dat op lange termijn digitale gegevens ontoegankelijk worden door de invoering van nieuwe apparatuur, programmatuur, gegevensdragers, standaarden en formaten. Wie gebruikt nog een Commodore-64? Waar zijn de tekstverwerkers Wordstar en WordPerfect 4.0 gebleven? Wie kan nog oude DOS programma's gebruiken? De NASA kan tapes met belangrijke onderzoeksgegevens uit de jaren zeventig niet meer lezen. En in de Verenigde Staten zijn de gegevens van een bevolkingsonderzoek uit het begin van de jaren zeventig voorgoed verloren gegaan. Wie heeft nog een disk drive waarmee grote floppies gelezen kunnen worden – en als men de computer al heeft, staan de gegevens er dan nog wel op? Dit is de problematiek waar we voor staan. Als we niets doen dan zal veel digitaal erfgoed uit deze tijd verloren gaan. Dit is een wezenlijk verschil met archivering van papieren erfgoed. Bij papier is 't beste als het met rust wordt gelaten, maar digitaal erfgoed zal alleen blijven bestaan als we er conserverende handelingen op toepassen.

De KB volgt een tweesporige aanpak. Ten eerste doet de KB directe ervaring op in de praktijk; zij heeft daarmee inmiddels nationaal en internationaal een vooraanstaande positie verworven. Ten tweede bevordert de KB, met medewerking van uitgevers, in internationaal verband onderzoek naar oplossingen voor lange-termijn-toegang tot elektronische publicaties.

Onderzoek en projecten

Vanaf 1996 tot 2000 voerde de KB onderzoeksprojecten uit om meer inzicht te krijgen in de problematiek van digitale archivering. Het begon met het project *Depot van Nederlandse Elektronische Publicaties voor Innovatie van Wetenschappelijke Informatievoorziening*, DNEP-IWT, waarin voor het eerst het gehele proces van selectie, acquisitie, bibliografische beschrijving, technische metadata, duurzame opslag en beschikbaarstelling werd geprobeerd in een experimentele omgeving. Dit project toonde duidelijk aan welke vraagstukken er zijn en op basis daarvan werd verder gewerkt. Het CERBERUS project onderzocht diverse strategieën voor het beschikbaar houden van digitale informatie zoals migratie (het overzetten van de gegevens van de ene naar de andere drager) conversie (overzetting van het ene naar het andere platform of van het ene programma naar het andere) en emulatie (een techniek waarbij men de publicatie niet verandert, maar men 'simuleert' op nieuwe computers de oude omgeving). Dit leverde een basis voor verder onderzoek door IBM. In het Europese project BIBLINK, *Linking Publishers and National Bibliographic Services* werd geëxperimenteerd met het geautomatiseerd uitwisselen van Dublin Core metadata tussen uitgevers en nationale bibliotheken. Het Europese project *Networked European Deposit Library (Nedlib)* NEDLIB leverde een model voor een digitaal archief gebaseerd op de standaard OAIIS, de *Open Archival Information System*. Dit model is inmiddels is voorgedragen aan de ISO organisatie en het wordt een officiële ISO standaard. De resultaten van de projecten zijn beschikbaar via de website van de KB <http://www.kb.nl> De projecten zijn alfabetisch gerangschikt onder 'Kenniscentrum' en dan 'Projecten & samenwerking'.

Overeenkomsten met uitgevers

Samenwerking met de uitgevers is van groot belang voor het slagen van een digitaal archief. Nederland heeft een depot op vrijwillige basis en de contacten met de uitgevers zijn uitstekend, omdat het deponeren gebeurt op basis van goede afspraken en vrijwilligheid. Bijna alle landen in de wereld hebben een wet voor deponering, maar in Nederland bestaat vanaf 1974 een depot op vrijwillige basis. Als eerste in de wereld sloot de KB in 1996 en 1997 overeenkomsten met grote uitgevers (Elsevier Science, Kluwer Academic Publishers, SDU Uitgevers) over de opslag en bewaring van onder andere elektronische wetenschappelijke tijdschriften. Sindsdien slaat de KB alle ruim 720 tijdschriften van Kluwer Academic Publishers op, circa 350 tijdschriften van Elsevier Science met Nederlands imprint, en de Opmaat bestanden van SDU Uitgevers. Het aantal tijdschriften van Elsevier Science wordt in 2002 uitgebreid tot het totale pakket van ongeveer 1.500 tijdschriften. De KB is dan wereldwijd het eerste officiële archief voor de elektronische publicaties van Elsevier Science. Alle gedeponeerde tijdschriften kunnen lokaal (*on site*) worden geraadpleegd. De KB streeft naar een breder gebruik van het gedeponeerde materiaal op basis van overeenkomsten met de uitgevers.

Het beleid van de KB is mede gebaseerd op de regeling met het Nederlands Uitgeversverbond betreffende 'deponering van elektronische publicaties in de Koninklijke Bibliotheek'. De regeling, die voor onbepaalde tijd geldt, is in werking getreden op 1 juli 1998 en zal in 2002 worden geëvalueerd.

Naast online wetenschappelijke tijdschriften verzamelt de KB dus allerlei andere soorten elektronische publicaties, die bijvoorbeeld gepubliceerd zijn op CD-ROM, diskette, tape en het Internet. De drager is uiteindelijk niet relevant – het gaat om de informatie.

De KB is zelf ook uitgever van 'webpublicaties' en die zullen ook worden opgenomen in het E-depot. Het gaat dan om grootschalige digitaliseringsprojecten, zoals Het Geheugen van Nederland, Digitale Atlas Geschiedenis, Middeleeuwse Verluchte Handschriften et cetera. Informatie over de digitaliseringsprojecten van de KB vindt u ook via de website <http://www.kb.nl> De prent van Frederik Hendrik werd gedigitaliseerd in het kader van het project Digitale Atlas Geschiedenis, waarbij oorspronkelijk zeventiende eeuwse materiaal uit het Rijksmuseum in Amsterdam en de KB wordt gepubliceerd op de website <http://www.digitalatlasgeschiedenis.nl>



Afbeelding 1: Frederik Hendrik, 1647, Rijksmuseum Amsterdam, NG-793

Archivering van webpublicaties

Aangezien de commerciële uitgevers in Nederland zorg dragen voor de bulk van belangrijke uitgaven heeft de KB tot nu toe prioriteit gegeven aan de archivering van *online* wetenschappelijke tijdschriften. In de Scandinavische landen, de VS en in Australië is men al gestart met webarchivering. Daarbij moet onderscheid worden gemaakt tussen het verzamelen van afzonderlijke webpublicaties en het bewaren van de hele webomgeving, inclusief de publicatiecontext en webfunctionaliteit.

Ook de KB gaat de komende jaren op experimentele basis webpublicaties archiveren. Er wordt een verzamelbeleid ontwikkeld en met het oog daarop worden selectiecriteria vastgesteld. Om praktische ervaring op te doen wordt een testbed ingericht. Mogelijke strategische samenwerkingpartners zijn onder meer de Library of Congress en nationale bibliotheken in Scandinavië. Webarchivering vergt afstemming met de selectiecriteria die voor het E-depot zijn vastgesteld.

Kenniscentrum Metadata

Voor digitale archivering zijn bibliografische beschrijvingen en metadata onmisbaar, maar ook meer in het algemeen voor het vinden van de juiste publicaties en informatie in bibliotheeksystemen en op het Internet. In internationaal verband worden steeds nieuwe (de facto) standaarden, methoden en technieken ontwikkeld. De KB concentreert de bestaande uitgebreide kennis over metadata, catalogisering, bibliografische beschrijvingen en formaten, conversies en XML binnen een Kenniscentrum Metadata.

DNEP-p : het pilot project

Sinds 1998 heeft de KB een pilot-project opgezet, DNEP-p. Deze pilot is gebaseerd op IBM Digital Library and Tivoly Storage Management software. Het testmagazijn omvat dus naast publicaties die digitaal ontstaan zijn ook gescande kopieën afkomstig van het digitaliseringprogramma van de KB. Laadmodules zijn ontwikkeld voor verschillende types bestanden die in het systeem ingevoerd worden. Ook zijn interfaces gemaakt met de catalogus van de KB, het reeds bestaande zoekstelsel KB-catalogus en KB-Launch. De implementatie van het testmagazijn en de integratie in de bestaande werkwijze en processen van de Koninklijke Bibliotheek betekent een continue inzet van mensen en middelen.

Stand van zaken november 2001 :

Totaal gebruikte opslagcapaciteit	830 GB
Online tijdschriften (1087 titels - 550.000 artikelen)	230 GB
Scans van digitaliseringsprojecten	600 GB

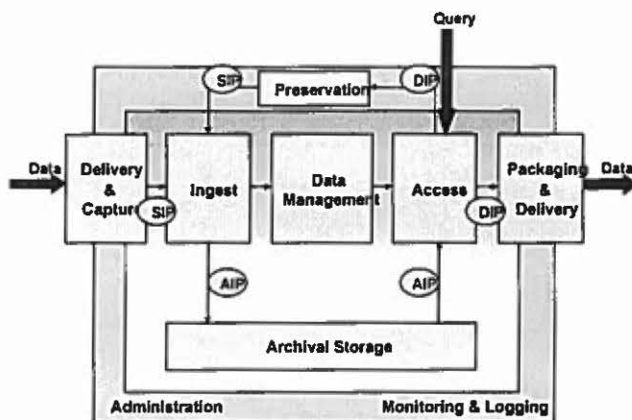
DNEP-i: Implementatie nieuw depotsysteem voor elektronische publicaties

In het bovengenoemde pilot project heeft de KB veel ervaring opgedaan met acquisitie, opslag en beschikbaarstelling van elektronische tijdschriftartikelen en scans van digitaliseringsprojecten. Dit systeem heeft echter geen functionaliteit voor lange termijn toegang, dat wil zeggen honderd jaar en meer. Daarom begon de KB in 1999 een Europese aanbestedingsprocedure voor een nieuw digitaal archief met lange termijn toegankelijkheid. IBM is uiteindelijk gekozen voor de ontwikkeling en implementatie van een nieuw depotsysteem, dat uniek is in de wereld. Daarbij wordt aangesloten bij de richtlijnen die de KB als coördinator van het project *Networked European Deposit Library (Nedlib)* samen met andere Europese nationale bibliotheken heeft opgesteld. De KB werkt nauw samen met de British Library, die een vergelijkbaar traject volgt als de KB. Het project DNEP-Implementatie (Depot van Nederlandse Elektronische Publicaties) levert eind 2002 het depotsysteem van de KB op; dit systeem vormt het technische hart van het Elektronische Depot (E-depot). Het ministerie van Onderwijs Cultuur en Wetenschappen stelt vanaf 2003 per jaar ruim 1,1 miljoen Euro beschikbaar voor de operationele kosten van dit E-depot. Het verder operationaliseren daarvan vergt de komende jaren een grote inspanning van de KB.

Het nieuwe systeem is geschikt voor de opslag van circa 8 miljoen publicaties (circa 12 Terabyte). De capaciteit kan worden uitgebreid tot maximaal 340 TB, genoeg voor de opslag van ruim 200 miljoen publicaties. Met het oog op de noodzakelijke verdere ontwikkeling en uitbreiding van het systeem richt de KB een experimentele omgeving in voor voortgaand onderzoek naar technieken en procedures voor lange-termijn-bewaring en -toegang. Net als in de afgelopen periode zal worden gezorgd voor de internationale inbedding van het onderzoek en voor een nauwe samenwerking met belangrijke ICT-spelers.

E-depot: het systeem

Within any electronic deposit a number of base processes can be identified as defined within the OAIS model



IBM / KB Long Term Preservation Study 2



Figuur 1: Open Archival Information System: basis voor het E-depot

Figuur 1 toont het schema van de KB implementatie van het OAIS model, dat de basis is van het depotsysteem. Dit Open Archival Information System, dat wereldwijd wordt gezien als het beste model voor een digitaal archief. Het is oorspronkelijk ontwikkeld door de NASA, die enorme hoeveelheden heterogene data van de ruimtevaartprogramma's moesten opslaan. Het model heeft een functionaliteit voor selectie, acquisitie, ontvangst, opslag en beschikbaarstelling van publicaties. Er zijn tevens modules voor digitale duurzaamheid en administratieve functies zoals data management en rapportages. De verdere technische uitwerking is ook beschikbaar via www.kb.nl in diverse rapportages.

Wat betekent het nieuwe digitale archief voor de KB-organisatie?

De KB verzamelt dus al enkele jaren, naast gedrukte publicaties, ook elektronische publicaties. Dat gebeurde tot nu toe min of meer op experimentele basis, maar binnenkort wordt het Elektronische Depot ofwel het E-depot op volwassen schaal operationeel. Ter voorbereiding daarvan realiseert de KB met steun van IBM een depotsysteem voor het voor de lange-termijn bewaren van de gedeponeerde elektronische publicaties. Op dit moment, april 2002, is er goede voortgang geboekt. De eerste oplevering van het depotsysteem heeft plaatsgevonden en inmiddels werkt de Hoofdafdeling Verwerking Publicaties op proef met de laadfunctionaliteit die is ontwikkeld. In juli 2002 zullen de andere twee systeemfuncties - het leveren en het beheer - door IBM aan de KB worden overgedragen om te worden getest. Meer en meer gaat nu de aandacht uit naar het inbedden van het depotsysteem in de KB. Dit is nodig om te zorgen dat het systeem straks goed past binnen de KB-organisatie en -infrastructuur.

Het depotsysteem dat de KB samen met IBM bouwt vormt het technische hart van het E-depot. Het E-depot is uiteraard meer dan alleen een systeem. Het betreft hier ook het gebruikmaken van functies van de diverse bibliotheeksystemen in de KB voor verwerken en leveren van elektronische publicaties. Voorbeelden van dergelijke functies zijn: acquisitie, catalogisering, maar ook publiekstoegang tot de catalogus en levering van publicaties. Het uitwerken van de opzet en het realiseren van het E-depot gebeurt onder verantwoordelijkheid van de Hoofdafdeling Verwerking Publicaties, maar deels ook van Hoofdafdeling Publieksdiensten en ICT.

Het nieuwe depotsysteem moet op een gegeven moment door de KB medewerkers gebruikt gaan worden. Voordat dat echter goed kan, moet de KB nog een aantal omliggende activiteiten regelen. Om het systeem in te bedden binnen de KB organisatie zijn de afgelopen tijd al allerlei acties gestart. Dit zijn acties zoals het aanpassen van gekoppelde systemen (bijvoorbeeld de KB-Catalogus, waarin een extra veld moet komen als link met het e-depot-systeem), ontwikkelen van omliggende applicaties zoals een generator (soort nummer-automaat) voor NBN's (National Bibliographic Number, dat als unieke sleutel tussen KB-systemen moet gaan dienen), het inrichten van de omliggende netwerkomgeving van het systeem (bijvoorbeeld een digitale ontvangstruimte voor elektronische publicaties), het organiseren dat medewerkers weten wat ze met ontvangen elektronische publicaties moeten doen, maar ook het organiseren en ontwikkelen van gebruikersregels en beveiligde toegang tot de opgeslagen elektronische publicaties. Al deze acties worden de komende tijd als Programma Inbedding E-depot uitgevoerd.

De KB gebruikte zoals gezegd reeds enige jaren het IBM Library systeem als systeem voor opslag en beschikbaarstelling van de elektronische tijdschriften van Elsevier, Kluwer en SDU. Het oude systeem, DNEP-p, is een aantal jaren geleden opgezet om als KB te leren wat het betekent om met elektronische publicaties te werken. De opgedane kennis wordt gebruikt bij de huidige ontwikkeling van het operationele systeem voor het E-depot. Het is de bedoeling om een gedeelte van de inhoud van het DNEP-p in 2003 geleidelijk over te brengen naar het nieuwe E-depot. In het DNEP-p-systeem zitten ook allerlei andere digitale materialen, zoals resultaten van digitaliseringsprojecten. Er wordt dus uiteindelijk een deel van de inhoud, namelijk alle e-depot publicaties naar het e-depot overgeheveld. De rest gaat naar andere KB-systemen toe die sterk op snelle presentatie zijn gericht. Het DNEP-p wordt daarna mogelijk voor andere doelen ingezet. Deze migratie is een mooi voorbeeld van de activiteiten die nodig zijn om computergegevens op lange termijn beschikbaar te houden, als men niet beschikt over een digitaal archief.

Tot slot

Het depotsysteem zal vanaf 2003 operationeel zijn en gebruikt worden door de medewerkers van de KB. Op de website van de KB kunt u de komende jaren steeds nieuwe informatie vinden over de praktische ervaringen die de KB opdoet met digitale archivering.