

Een repository voor Gasunie

Bert Huizing b.huizing@gasunie.nl
Richard Jonkman richard.jonkman@rug.nl

In 2004 startte de Universiteitsbibliotheek Groningen het project NorthSee. Doel van dit project is samen met partners repositories in te richten op de gebieden Duurzaamheid, Energie en Milieu. Met de Nederlandse Gasunie is een repository gebouwd die fungeert als digitaal archief voor Energie en Aardgas. In dit artikel beschrijven Bert Huizing en Richard Jonkman het proces van idee tot oplevering van de repository.

Een belangrijke verandering in de wetenschappelijke informatiewereld is die van papier naar digitaal. Het verschijnsel repository¹ staat dus steeds meer in de belangstelling. Het opzetten van deze digitale archieven biedt veel voordelen. Toegang tot een digitaal document is overal, snel en simpel mogelijk via een computernetwerk of internet. Een belangrijk verschil tussen database en repository is dat documenten altijd full text beschikbaar zullen zijn in de repository. Verder wordt ook het opslaan gemakkelijker gemaakt door digitalisering. In plaats van magazijnruimte kan de academische bibliotheek steeds meer volstaan met het plaatsen en inrichten van een server. Tenslotte kunnen repositories ook gekoppeld worden aan andere systemen zoals kenniskaarten, die leiden naar curriculum vitae en de kennis van de auteurs.

Digitaal archiveren

De Nederlandse universiteiten hebben samenwerking² gezocht

in het project DARE (Digital Academic Repositories) en de repository is volop in opkomst als middel om output van wetenschappelijke arbeid digitaal te archiveren en toegankelijk te maken. Repositories richten zich dan op publicaties van de eigen universiteit (institutionele repository) of op een bepaald vakgebied of discipline. Binnen DARE lopen diverse repository-projecten waarin de Universiteitsbibliotheek Groningen (UB) initieert, coördineert en samenwerkt met diverse partners.

Het DARE-project NorthSee is in het voorjaar van 2004 gestart als initiatief van het Centrum voor Energie en Milieu (IVEM) van de Rijksuniversiteit Groningen. NorthSee richt zich op het creëren van repositories in drie disciplines: 'duurzaamheid', 'energie' en 'milieu'. Partners³ in dit project zijn naast IVEM ook o.a. de Universiteit Twente en de Nederlandse Gasunie. Doel van de samenwerking is het opslaan van relevante documenten in eigen institutionele repositories.

De afdeling DB (Digitale Bibliotheekvoorzieningen) van de UB leidt NorthSee. De DB houdt zich bezig met ICT-innovatie⁴ en bouw en beheer van repositories.

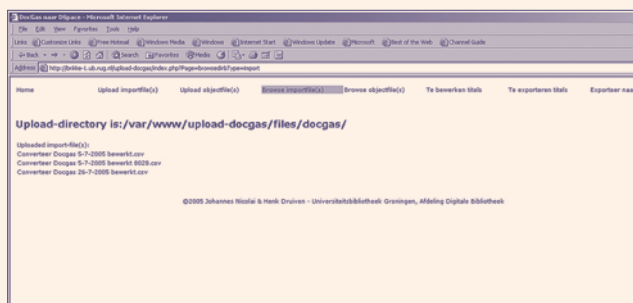
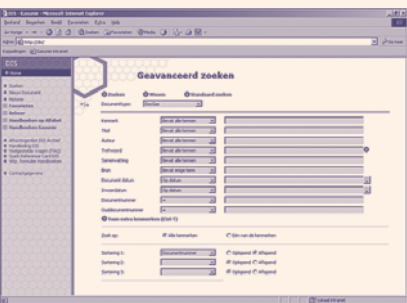
Voor de bouw van de Gasunie institutionele repository heeft de DB samengewerkt met het Energy Knowledge Center (EKC) van Gasunie Engineering & Technology. Het EKC is bibliotheek, informatie- en kenniscentrum van Gasunie Engineering & Technology. Het EKC heeft in het project

een belangrijke rol gespeeld als leverancier van content en klankbord. Sinds medio januari 2006 is de repository functioneel en beschikbaar over ruim 1300 documenten. De bouw ging echter niet zonder horten en stoten voordat de repository live kon gaan.

Gasgerelateerd

Het bouwen van de repository startte met het in kaart brengen van de hoeveelheid geschikte Gasuniedocumenten. Er moest een selectie gemaakt worden van ongeveer 17000 documenten, opgeslagen in het documentinformatiesysteem van de Gasunie genaamd Docgas. Docgas kan via verschillende zoekingen doorzocht worden. Het herbergt niet alleen interne Gasuniedocumenten, maar ook tijdschriftartikelen en externe onderzoeksrapporten. Natuurlijk zijn alle documenten in meer of mindere mate gerelateerd aan de gasindustrie. Alle documenten worden van metadata voorzien en opgeslagen onder een uniek nummer, het zogenaamde Docgasnummer. Metadata is informatie die aan een object wordt toegevoegd en bestaat in dit geval uit gegevens over het document zelf, zoals titel, auteur enz. Tijdschriftartikelen en andere externe documenten worden op dezelfde manier verwerkt, maar dienen vaak eerst gedigitaliseerd te worden.

De selectie vond plaats door het uitvoeren van gerichte zoekacties in de Docgasdatabase op basis van de trefwoorden 'energie', 'milieu' en 'duurzaamheid'⁵.



Dit leverde uiteindelijk een resultaat op van zo'n 1300 bruikbare documenten, die allen full text opvraagbaar zijn in de formaten PDF, TIFF, Word en Excel.

De metadata en het bijbehorende full text-document, ook wel *object-file* genoemd, zijn als aparte bestanden opgeslagen en kunnen ook als zodanig vanuit het systeem worden opgevraagd.

Nu de selectie had plaatsgevonden, moest er een manier bedacht worden om de metadata en object-files vanuit Docgas naar buiten toe te transporteren. Dit was noodzakelijk omdat deze gegevens anders niet overzichtelijk gepresenteerd zouden kunnen worden en dus onbruikbaar voor het verdere verwerkingsproces. Gelukkig biedt Docgas de mogelijkheid via de functie 'overzichten' een Excel-file aan te maken, waarin alle metadata van een document kan worden vastgelegd. Zo is het mogelijk om na een zoekactie, de gevonden documenten op het scherm aan te vinken en de metadata en vervolgens te transporteren. De object-files van de aangevinkte documenten kunnen in een aparte map worden geplaatst en zijn herkenbaar via het unieke Docgasnummer.

Conversieproblemen

Na de documentselectie rees de vraag hoe de Excel-files met metadata vanuit de Gasunie-omgeving zonder conversieproblemen konden worden overgezet naar de Gasunie-repository⁷ van de RUG. Hierbij was het van cruciaal belang dat de Docgasmetadata

naadloos aansloot op het gebruikelijke metadata-protocol in DSpace⁶. Dit is het open source repository-systeem in gebruik van de RUG waarmee gebruikers digitale bestanden kunnen ondervangen, archiveren, beschrijven en indexeren. DSpace zou de Docgasmetadata anders niet herkennen.

Als snel werd duidelijk dat er verschillen waren en dus moest er een oplossing gevonden worden. Dit resulteerde in het uitvoeren van een *mapping* tussen de metadata van Docgas en die van DSpace. Hierbij werd gecontroleerd welke velden overeenkwamen en welke niet. Bij afwijkingen werd een synoniem gezocht. Zo werd het veld 'auteur' in Docgas omgezet naar 'contributor.author' in DSpace. Nadat de metadata van beide systemen op één lijn waren gezet, werd het tijd om een oplossing te vinden voor de daadwerkelijke overdracht.

De 1300 Docgasdocumenten, bestaande uit metadata in een Excel-file en bijbehorende objectfiles in een aparte map, stonden nu klaar om getransporteerd te worden naar de Gasunie-repository.

Op basis van de uitgevoerde mapping werd door Johannes Nicolaï en Henk Druiven van de afdeling Digitale Bibliotheekvoorzieningen een conversieprogramma ontwikkeld, dat in vier stappen de overdracht mogelijk maakte. Allereerst werd via een FTP-programma een upload mogelijk gemaakt van de metadata en objectfiles naar het RUG-netwerk. Met dit File Transfer Protocol (FTP) kunnen netwerk-

gebruikers bestanden uitwisselen. Zo is het mogelijk om bestanden van de pc-harde schijf te uploaden naar een andere computer op het internet. Vervolgens werden de volgende stappen uitgevoerd:

Stap 1: Upload van de importfile, de Excel-file met metadata in txt. Formaat.

Stap 2: Upload van de Objectfiles.

Stap 3: Conversie van de importfile. Het gelijkstellen van de metadata-protocollen.

Stap 4: De export naar DSpace.

Met de laatste stap werden de Importfiles en Objectfiles samengevoegd op basis van het unieke Docgasnummer en getransporteerd. Dit resulteerde in een vloeiende samenvoeging van de metadata en de full text-documenten in de Gasunie-repository.

Improvisatievermogen

Het project Gasunie repository was een leerzame ervaring. Wij hadden te maken met een uitzonderlijke situatie, aangezien de wetenschappelijke output die aangeleverd moest worden reeds lag opgeslagen in documentinformatiesysteem. De uitdaging lag in het gelijktrekken van de verschillende metadata-protocollen en het uiteindelijke transport van de documenten vanuit de Gasunie-omgeving naar het RUG-netwerk. Met flink wat improvisatievermogen en een plezierige samenwerking met de afdeling Digitale Bibliotheekvoorzieningen, is het gewenste resultaat uiteindelijk toch bereikt.

noten

- 1 Saarti, Jarmo (2005) From printed world to a digital environment: the role of the repository libraries in a changing environment. In: Library Management Vol. 26, No. 1-2, 2005 pp. 26-31
- 2 Klugkist, Alex & Feijen Martin (2004) DARE: een universitair digitaal archief. In: Pictogram 4 (augustus/september) pp. 13-15.
- 3 www.rug.nl/bibliotheek/catalogibestanden/elpubrug/dare/northsee
- 4 Hollander, Frank den & Piersma, Kristien (2005) Mooie elektronische diensten. In: Pictogram 1 (februari/maart) pp. 3-6
- 5 Projectvoorstel Northsee: Een repository-based webplatform op het gebied van Sustainability, Energy en Environment in Noord Nederland
- 6 www.dspace.org
- 7 ir.ub.rug.nl