



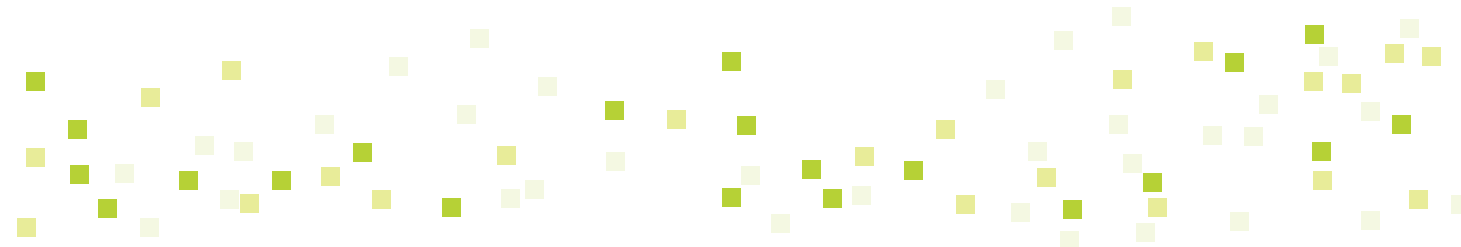
Digitale innovatie voor het erfgoed

CATCH

CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS



Digitale innovatie voor het erfgoed





Inhoud

CATCH en CATCHPlus	4-5
Voorwoord	6-7
Inleiding	8-11
Facts & Figures	12-13
SCRATCH4All	14-17
Workspaces	18-19
CHoralPlus	20-23
WitchcraftPlus	24-27
MuSeUMPlus	28-31
ZieOok en UPR	32-35
Multiply	36-39
STITCHPlus	40-41
OpenSKOS	42-43
DocumentChecker	44-45
Annotation Repository	46-47
Persistent Identifiers	48-49
Op weg naar een erfgoedmakelaar	50-53
Organisatie	54-55
Meer informatie	56-57

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

CATCH

Cultureel erfgoed vormt ons collectieve geheugen. Het is overal om ons heen. CATCH doet onderzoek en ontwikkelt wetenschappelijke methoden en technieken op het snijvlak van geesteswetenschappen en informatica. Met de resultaten worden erfgoedcollecties beter toegankelijk voor publiek, professionals en wetenschappers. De Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) financiert het CATCH-programma. Daarbij werken de NWO-gebieden Exacte Wetenschappen en Geesteswetenschappen samen.

Centraal in CATCH staan innovatie, samenwerking en overdraagbaarheid. Aan elk project ligt een vraag vanuit het erfgoed (in latere projecten ook wel vanuit de geesteswetenschappen) ten grondslag. Een onderzoeksteam bestaat uit een promovendus, een postdoc en een wetenschappelijk programmeur. Deze teams opereren binnen de muren van de erfgoedinstelling (*'embedded research'*).

In totaal zijn sinds 2004 18 projecten gefinancierd.

CATCHPlus

De onderzoeksprojecten van CATCH leveren naast wetenschappelijke output ook pilots en prototypen op. Om deze geschikt te maken voor daadwerkelijk gebruik in het erfgoed is CATCHPlus uitgevoerd. CATCHPlus ontving subsidie van het interdepartementale Programma Implementatie Agenda ICT-beleid (PRIMA), het ministerie van Onderwijs, Cultuur en Wetenschap en de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). Naast de doorontwikkeling van de pilots, zijn er ook erfgoedbrede diensten ontwikkeld en zijn er activiteiten ontplooid om stukjes van de infrastructuur voor het digitale erfgoed te ontwikkelen. Dit alles draagt bij aan meer samenwerking en grotere samenhang in het erfgoed om daardoor de collecties beter te ontsluiten en toegankelijk te maken.

CATCHPlus heeft van 2009 tot 2012 gelopen.



Voorwoord

Hans Bennis

Directeur Meertens Instituut (KNAW)

Voorzitter Raad van Toezicht CATCHPlus

In uw handen houdt u een overzicht van de resultaten van CATCHPlus. Dit boekje is een catalogus met informatie over nieuwe software en diensten die zijn ontwikkeld voor de erfgoedbranche. De basis voor deze toepassingen werd gelegd in het NWO-onderzoeksprogramma CATCH, waarin wetenschappers uit de informatica en de geesteswetenschappen samenwerken met medewerkers van verschillende erfgoedinstellingen. Binnen CATCHPlus zijn de veelbelovende demo's van CATCH doorontwikkeld tot volwaardige producten die breed inzetbaar zijn in de Nederlandse erfgoedwereld.

CATCH staat voor *Continuous Access To Cultural Heritage*, continue toegang tot het cultureel erfgoed. Veel van ons culturele erfgoed is inmiddels gedigitaliseerd of beschreven in databases. Theoretisch is het dus mogelijk deze schat aan data onbeperkt te doorzoeken en relaties te leggen die eerder niet zichtbaar waren. In de praktijk blijkt deze mogelijkheid echter beperkt, door het bestaan van verschillende standaarden en de begrensde mogelijkheden van bestaande technologieën. De software en diensten die bin-

nen CATCHPlus zijn ontwikkeld, brengen hier verandering in. Zowel voor mensen die werkzaam zijn bij erfgoedinstellingen als voor de bezoekers van deze instellingen. Ik hoop dan ook dat u zich zult laten verrassen en inspireren door de applicaties die in dit boekje beschreven worden. Elke toepassing is voorzien van een inhoudelijke beschrijving, soms aangevuld met de persoonlijke ervaring van een gebruiker uit de erfgoedwereld. Vervolgens vindt u een link naar een website waarop u meer informatie over de applicatie kunt vinden.

CATCHPlus heeft een begin gemaakt met het gemeenschappelijk ontsluiten van cultureel erfgoed voor verschillende doelgroepen. Wij zijn blij te kunnen vaststellen dat deze onderneming een vervolg krijgt, onder andere binnen het netwerk Cultureel Erfgoed van het innovatieplatform Creatieve Industrie. Ook komt er een erfgoedmakelaar die als schakel gaat fungeren tussen de wereld van het cultureel erfgoed en de ontwikkeling van nieuwe technologie. Hierdoor wordt u in staat gesteld om ook straks op een innovatieve manier kennis te maken met delen van ons erfgoed.

Ik wens u allereerst veel leesplezier en voor straks veel genoeg bij een tocht door de fascinerende wereld van het digitale erfgoed.



Inleiding

Paul Doorenbosch

*Koninklijke Bibliotheek
Voorzitter CATCHPlus-stuurgroep
Vice-voorzitter dagelijks bestuur CATCH*

Patricia Alkhoven

*Meertens Instituut
Projectleider CATCHPlus*

Het fenomeen CATCH

Inzet van IT is al meer dan dertig jaar gebruikelijk in het erfgoedbeheer en de erfgoedontsluiting. Digitaliseren, beheersdatabases, websites, games e.d. zijn normale onderdelen van het erfgoedbedrijf geworden. In de beginjaren veelal uitgevoerd door erfgoedbeheerders die lol hadden in computers ('in het land der blinden...') of door werkzaamheden uit te besteden aan externe bedrijven. Hiermee zijn prachtige resultaten behaald, zeker daar waar de IT-toepassingen sterk leken op de traditionele processen.

Maar de hoeveelheden digitale data werden steeds groter en het besef begon door te dringen dat veranderingen in de erfgoedketen (van analoog naar digitaal) misschien wel

vroegen om de inzet van nieuwe methodieken en nieuwe kennis. Rond de eeuwwisseling werden plannen ontwikkeld om de kennis en methoden uit de informatica in te zetten voor het erfgoed. De verwachting was dat daaruit voor beide partijen nieuwe mogelijkheden zouden ontstaan. Een eerste plan dat werd ingediend bij het aardgasbatenfonds haalde het net niet, maar NWO Exacte Wetenschappen en NWO Geesteswetenschappen zagen de grote potentie van die samenwerking. Het oorspronkelijke plan werd omgebouwd naar een meer wetenschappelijke insteek, waarbij nog steeds de vraag vanuit het erfgoed voorop bleef staan. Vanaf 2004 ging het programma CATCH (*Continuous Access to Cultural Heritage*) lopen. Eerst met zes projecten die in het programmaplan gedefinieerd waren, daarna met nog vier projecten die in vrije competitie tot stand kwamen. Sindsdien zijn er nog twee rondes met elk vier projecten gerealiseerd.

Bijzonder aan de projecten was niet alleen dat ze vanuit een erfgoed-, of later een humanioravraag tot stand kwamen, maar dat er met '*embedded research teams*' werd gewerkt. Door de onderzoekers op dagelijkse basis aan de koffiemachine te ontmoeten, kon er informeel kennisuitwisseling plaatsvinden en kon de bewustwording van de mogelijkheden die de informatica biedt, worden uitgebouwd.

De eerste tien CATCH-projecten zijn wetenschappelijk zeer succesvol afgesloten. Promoties en een lange lijst hoog-

waardige wetenschappelijke artikelen. Maar ook pilot-applicaties, algoritmen, databases, standaarden en technieken die in het erfgoed nieuw waren. Nieuw, maar waarvan voor alle betrokkenen duidelijk was dat ze de erfgoedsector verder zouden helpen. NWO en het overheidsfonds PRIMA zagen mogelijkheden om de maatschappelijke relevantie van het onderzoek te verzilveren (NWO) en om verder te bouwen aan een infrastructuur voor het digitale erfgoed (de overheid). Dit gevoegd bij de wens van de erfgoedinstellingen om de pilotapplicaties om te bouwen naar tools voor de praktijk, leidde tot het vervolgproject CATCH-Plus. Startkapitaal was de erfenis van de wetenschappelijke projecten, enkele infrastructuurwensen van OCW, geld van het PRIMA-fonds, geld en support van NWO en de in-kind inzet van medewerkers van de erfgoedinstellingen. Het was toen 2009.

In het CATCHPlus-project hebben twee instellingen een belangrijke rol in de organisatie gespeeld: het eerste jaar het Instituut voor Beeld en Geluid en de resterende periode het Meertens Instituut. Zij hebben het projectbureau gehuisvest en de verantwoordelijkheid namens de stuurgroep voor het project genomen. En het resultaat mag er zijn.

Op een kleine uitzondering na hebben alle projecten de eindstreep gehaald. Ze zijn soms wat bijgesteld, in een enkel geval heeft een project een andere 'eigenaar' gekregen, er zijn enkele nieuwe erfgoedinstellingen ingestapt.

Hobbels en afwijkingen die ieder project moet nemen. De aanvankelijk wat lastige start door de gecompliceerde financiering, is in de laatste twee jaar glorieus overwonnen. Nu ligt er een suite aan applicaties, soms klein en tot één instelling beperkt, soms groot en generiek over de hele erfgoedsector heen. Er zijn belangrijke vorderingen gemaakt bij de invoering van de persistente identifier-infrastructuur. Er is een centraal repository voor vocabulaires. Er is een annotatieruimte. Er ligt een aanzet voor een make-laarsfunctie die de duurzaamheid van applicaties moet gaan bevorderen. Maar bovenal ligt er een scala aan interessante tools voor het erfgoed.

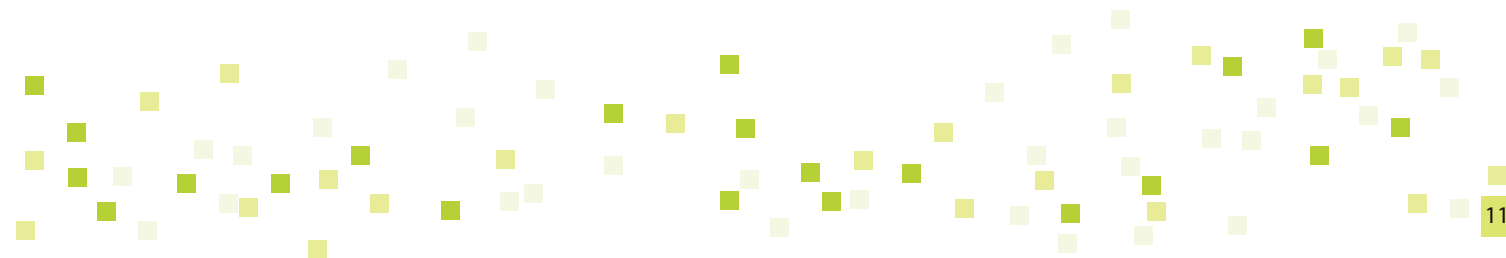
Toch is het gelukkig niet gebleven bij alleen harde project-resultaten. Het huwelijk tussen erfgoed en informatica bleek op zakelijk en op menselijk vlak een gouden greep. Het erfgoed kreeg kennis van de mogelijkheden die de informatica biedt, en de informatica maakte kennis met de onmetelijke rijkdom aan data en semantiek in het erfgoed. Mensen vonden elkaar met hun nieuwe kennis, ook op andere vlakken. Instellingen pasten soms hun organisatie aan toen ze zich bewust werden van de nieuwe kansen.

De synergie van erfgoed, informatica en in toenemende mate de geesteswetenschappen heeft ook bijgedragen aan de ontwikkeling van de digital humanities. Het idee dat je door bundeling van mensen en disciplines krachtiger wordt om innovaties en innovatief onderzoek

te doen, zien we nu overal opdoemen. Samenwerking van ogenschijnlijk zeer verschillende disciplines. Het blijft ook mensenwerk. De deelnemers in de CATCH-Plus-projecten worden allemaal in dit boekje genoemd. Wat we niet moeten vergeten is dat een klein groepje mensen in 2002 in een hotel aan een snelweg in Den Bosch een impasse in de gesprekken tussen het erfgoed en de verschillende informaticapartijen konden doorbreken. Zij schreven de aanzet tot het CATCH-programma. Alice Dijkstra, Paul Doorenbosch, Annemarie Bos, Elco Bruinsma, Jaap van den Herik, Frank van Harmelen, Mark Kas, Martin Kersten, Jos

Taekema en Peter Sigmond. Ere wie ere toekomt. CATCH heeft een verleden, maar CATCH heeft vooral een toekomst. Niet per se als programma, maar wel door de mensen die hebben meegedaan en door de materiële en immateriële resultaten die hun weg zullen vinden in de samenwerking tussen erfgoed, informatica en geesteswetenschappen voor een steeds effectiever en zinvoller gebruik van het erfgoed.

Oktober 2012



Facts & Figures

CATCHPlus staat voor Continuous Access To Cultural Heritage Plus.

CATCHPlus bouwt voort op het lopende NWO-onderzoeksprogramma CATCH.

CATCHPlus liep van 2009 tot en met 2012.

In CATCHPlus zijn verschillende CATCH-demo's doorontwikkeld tot breed inzetbare applicaties.



CATCHPlus heeft geleid tot 9 verschillende applicaties bij erfgoedinstellingen en 5 diensten voor de hele erfgoedsector.

Het projectbureau was vanaf november 2010 (o.l.v. Patricia Alkhoven) gevestigd in het Meertens Instituut. Hennie Brugman hield als technisch coördinator zicht op de ontwikkeling van applicaties.

Het projectbureau werkte onder supervisie van een Stuurgroep (o.l.v. Paul Doorenbosch) en een Raad van Toezicht (o.l.v. Hans Bennis).

Een begeleidingscommissie (o.l.v. Marco de Niet) zag toe op de voortgang binnen de projecten. Zij bracht *site visits* aan de verschillende betrokken erfgoedinstellingen.

Betrokken bij CATCHPlus waren 11 erfgoedinstellingen:

Amsterdam Museum, Rijksmuseum, Meertens Instituut, Naturalis, Gemeentemuseum Den Haag, Koninklijke Bibliotheek, Nationaal Archief, Instituut voor Beeld en Geluid, Rijksdienst voor het Cultureel Erfgoed, Theaterinstituut Nederland, Gemeentearchief Rotterdam.

Daarnaast waren er 6 kennisinstellingen betrokken: Digitaal Erfgoed Nederland (DEN), Universiteit van Groningen, Universiteit van Amsterdam, Vrije Universiteit, Universiteit van Tilburg, Universiteit Twente.

De totale projectkosten bedroegen 3,1 miljoen euro.

De financiers van CATCHPlus zijn:

Programma Implementatie ICT Agenda (PRIMA), Ministerie van Economie, Landbouw en Innovatie (ELI)

Ministerie van Onderwijs, Cultuur en Wetenschap (OCW)

Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)

De betrokken erfgoedinstellingen

Kennisinstellingen





SCRATCH4ALL

Script Analysis Tools for the Cultural Heritage

Digitale monnik ontcijfert ingescande handschriften

Erfgoedinstellingen hebben de laatste jaren veel tijd en geld besteed aan het digitaliseren van hun archieven. Gedrukte teksten zijn met optische tekenherkenning redelijk goed te doorzoeken, maar voor geschreven teksten is dat zo goed als onmogelijk. SCRATCH4All brengt daar verandering in.

Ingescande teksten zijn nog niet doorzoekbaar voor een computer. Die ziet de scan namelijk als een afbeelding en weet niet dat er letters opstaan. Met optische tekenherkenning (OCR) is de afbeelding om te zetten in letters, maar deze techniek werkt alleen bij gedrukte teksten. De resultaten van optische tekenherkenning zijn zeker niet foutloos, maar is meestal wel goed genoeg om de teksten te kunnen doorzoeken met een zoekapplicatie. Handgeschreven teksten bevatten te veel variatie om door OCR herkend te worden. Niemand schrijft een letter op precies dezelfde manier als zijn buurman en zelfs binnen de teksten van één auteur is dezelfde letter lang niet altijd hetzelfde. Wetenschappers aan de Rijksuniversiteit Groningen hebben daarom Monk ontwikkeld, software die hand-

geschreven teksten doorzoekbaar kan maken. In het CATCH-project SCRATCH is een gebruikersinterface ontwikkeld rondom Monk, waarmee het Nationaal Archief ingescande handschriften kan doorzoeken.

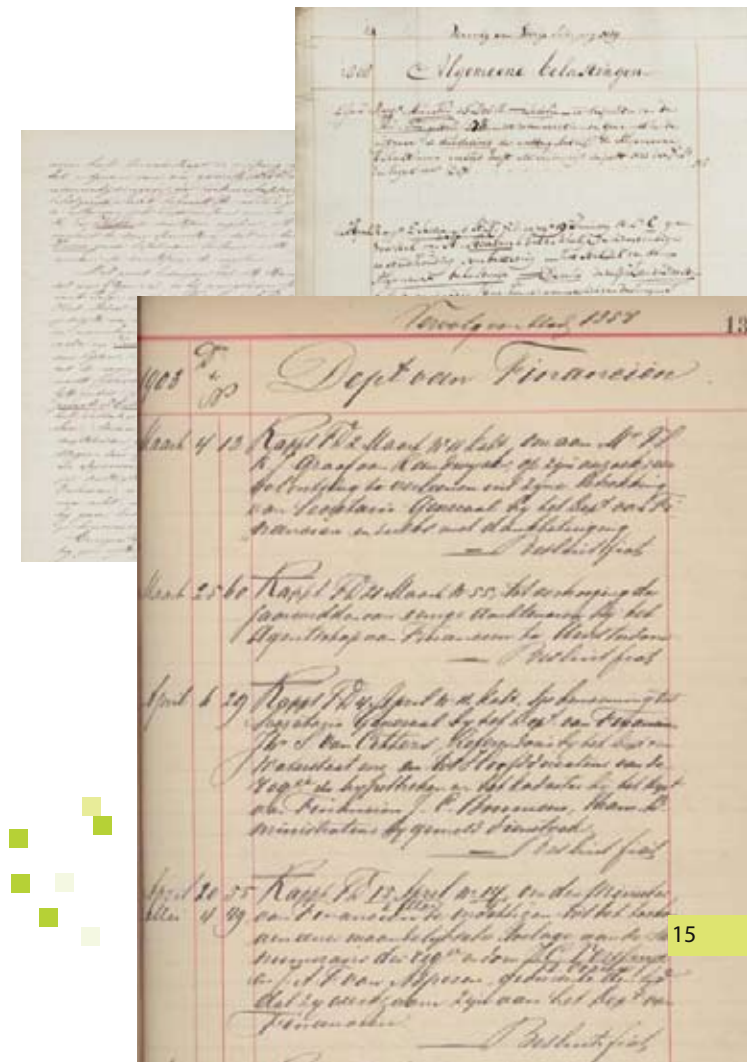
Het Nationaal Archief gebruikt Monk om het archief van het Kabinet der Koningin doorzoekbaar te maken. De indices van dit archief zijn gedurende lange tijd door dezelfde persoon geschreven.

Dat is erg belangrijk, want Monk wordt getraind voor één specifiek handschrift. Het systeem heeft minimaal vijf voorkomens van hetzelfde woord nodig om het de zesde keer zelf te kunnen herkennen.

Een grote database van hetzelfde handschrift is daarom essentieel voor een goed resultaat.

Volledig automatische herkenning van handschriften is met de in SCRATCH4All ontwikkelde software niet mogelijk, maar de archieven zullen wel dusdanig doorzoekbaar worden dat gericht en efficiënt verder zoeken op basis van de resultaten mogelijk is.

De SCRATCH4All-gebruiker, zowel professioneel als amateur, kan de resultaten van Monk verder bewerken via de Workspaces (zie blz. 18-19). Door de ingescande handschriften handmatig te annoteren, zullen de resultaten van Monk steeds beter worden.





**Henny van Schie,
Nationaal Archief:**

“Binnen het Nationaal Archief is Monk gebruikt om het handgeschreven archief van het Kabinet der Koningin te ontsluiten. De indices van dit archief zijn vele jaren geschreven door dezelfde persoon, dus deze waren een zeer geschikte kandidaat voor Monk.

SCRATCH4All heeft mijn werk als archivaris niet direct veranderd. Van een archivaris wordt zekerheid en volledigheid verwacht: als je iets niet vindt, mag het er ook niet zijn. Je kunt dan niet zeggen: ‘Jammer dat ik het niet voor je heb kunnen vinden, maar mijn tools hebben het niet kunnen traceren.’ Zelf blijf ik dus veroordeeld tot de handmatige werkwijze uit de 19e eeuw met alfabetische klappers en indices.

Maar voor onderzoekers is het gebruik van Monk binnen de werkomgeving van SCRATCH4All zeker interessant. Nadat Monk een handschrift heeft geleerd –hier gaat het nodige werk inzitten– zijn de archieven veel directer voor de onderzoeker toegankelijk. Hij kan dan archieven raadplegen die hij anders niet zou inzien, omdat de metadata te weinig houvast bieden.

Monk zal geen 100% betrouwbare resultaten opleveren,

omdat hij niet kan garanderen alle woorden goed herkend te hebben. Gebruikers zullen in de SCRATCH4All-interface de resultaten steeds moeten bevestigen of verbeteren en zo Monk moeten helpen om het handschrift steeds beter te leren.

Toch past SCRATCH4All goed bij het Nationaal Archief, omdat het lastig bevroagbare, chronologische archieven helpt ontsluiten. Het geeft een goede indicatie waar je relevante informatie kunt vinden. Op de lange termijn kan Monk er zeker toe bijdragen dat het onderzoek sneller en soepeler verloopt.”



Projectteam SCRATCH4All

Marika Ceunen (*Stadsarchief Leuven*)
 Rolf Fokkens (*Target Holding*)
 Petra Helwig (*Nationaal Archief*)
 Yvette Hoitink (*Nationaal Archief*)
 Edzer Lawerman (*Nspyre*)
 Jean-Paul van Oosten (*Rijksuniversiteit Groningen*)
 Minne Oostra (*de Ontwikkelfabriek*)
 Harry Romijn (*Groninger Archieven*)
 Henny van Schie (*Nationaal Archief*)
 Lambert Schomaker (*Rijksuniversiteit Groningen*)
 Petra Schoen (*Nationaal Archief*)
 Anco Westra (*Nspyre*)
 Michiel van Wijngaarden (*Gelders Archief*)
 Rees Williams (*Rijksuniversiteit Groningen*)

Projectteam SCRATCH

Jacques Bogaarts (*Nationaal Archief*)
 Cathy Jager (*Nationaal Archief*)
 Fons van der Laan (*Rijksuniversiteit Groningen*)
 John Nerbonne (*Rijksuniversiteit Groningen*)
 Henny van Schie (*Nationaal Archief*)
 Lambert Schomaker (*Rijksuniversiteit Groningen*)
 Tijn van der Zant (*Rijksuniversiteit Groningen*)
 Sveta Zinger (*Rijksuniversiteit Groningen*)

Interesse?

Wilt u meer weten over SCRATCH4All?
 Surf dan naar www.catchplus.nl/scratch4all



Workspaces

Samenwerken in een digitale werkomgeving

In een webgebaseerde wereld waar verschillende systemen en gebruikers samenwerken aan één taak is vaak tijdelijke opslag van gegevens nodig. Meestal moet de gebruiker aan de ene kant gegevens uploaden of invoeren en aan de andere kant de bewerkte data controleren of (na)bewerken. Workspaces biedt een digitale werkomgeving waarin verschillende collega's deze werkzaamheden samen uit kunnen voeren.

SCRATCH4All, gepresenteerd op de vorige pagina's, is hier een goed voorbeeld van. De gebruiker beschikt over digitale scans van een handschrift. Deze zal hij moeten uploaden naar de applicatie, zodat deze het handschrift kan analyseren. Vervolgens kan de gebruiker het resultaat verder bewerken en zo nodig corrigeren. De Workspaces zijn hier onmisbaar voor de gebruiker.

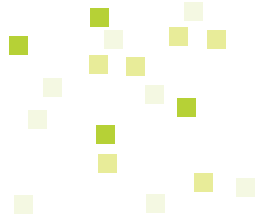
Workspaces biedt nog veel meer voordelen voor erfgoedinstellingen. De digitale werkomgeving kan toegang bieden aan verschillende applicaties, allemaal overzichtelijk in hetzelfde systeem. De Workspaces zijn flexibel in te richten naar wens van de instelling. Verschillende collega's (zowel binnen als buiten de instelling) kunnen via het platform

samenwerken aan collecties die nog in ontwikkeling zijn. De instelling kan precies aangeven welke gebruiker of welk systeem toegang krijgt tot welke informatie uit de collectie.

De Workspaces zijn ook interessant voor de ontwikkelaars van applicaties voor de digitale-erfgoedwereld. Zij kunnen de ingebouwde modules voor bijvoorbeeld inloggen, autorisatie of uploaden eenvoudig hergebruiken en aansluiten op hun nieuwe, zelfontwikkelde applicaties.

Projectteam

Hennie Brugman (*Meertens Instituut*)
Rolf Fokkens (*Target Holding*)
Minne Oostra (*Target Holding*)



Interesse?

Wilt u meer weten over Workspaces?
Surf dan naar www.catchplus.nl/workspaces





CHoralPlus

Access to oral history

Zoeken in spraak

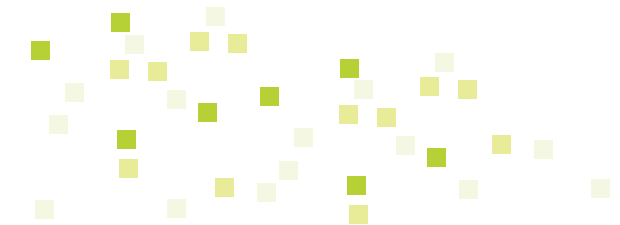
Zoeken in digitale tekst is vrij eenvoudig met ctrl+F. Zoeken naar een fragment in een geluids- of video-opname is een stuk moeilijker. Volledige transcripties zijn doorgaans niet aanwezig en de informatie uit meta-data is beperkt. Automatische spraakherkenning maakt het mogelijk deze bestanden toch te doorzoeken.

Spraakherkenning kan gesproken Nederlands omzetten naar geschreven tekst. Door de getranscribeerde tekst te koppelen aan de tijds codering van de opname, zijn fragmenten op woordniveau terug te vinden in de bestanden. Deze techniek werkt al langere tijd vrij goed bij ideale opname-omstandigheden: geen ruis en één spreker die duidelijk spreekt.

In het CATCH-project CHoral hebben onderzoekers van de Universiteit Twente onderzocht hoe audiobestanden getranscribeerd kunnen worden die zijn opgenomen onder minder ideale omstandigheden. Veel archieven hebben dergelijke opnamen in hun beheer, met onprofessionele sprekers, meerdere sprekers per opname en/of omgevingsgeluiden. Denk bijvoorbeeld aan interviews of radioprogramma's van regionale zenders.

De software die is ontwikkeld in CHoral werkt stapsgewijs. Eerst scheidt hij de spraak van de achtergrondgeluiden. Vervolgens clustert hij de fragmenten van de verschillende sprekers. Deze probeert hij automatisch te herkennen en ten slotte wordt de getranscribeerde tekst opgelijnd met de opname, zodat precies duidelijk is waar in de opname wie wat zegt.

Binnen CATCHPlus is een webinterface ontwikkeld waarmee erfgoedinstellingen gebruik kunnen maken van de ontwikkelde spraakherkenningssoftware. Instellingen die gebruik maken van deze toepassing kunnen via hun eigen account inloggen en files uploaden. De resultaten van de spraakherkenner worden vervolgens via mail naar de gebruiker gestuurd.





**Mies Langelaar,
Stadsarchief Rotterdam:**

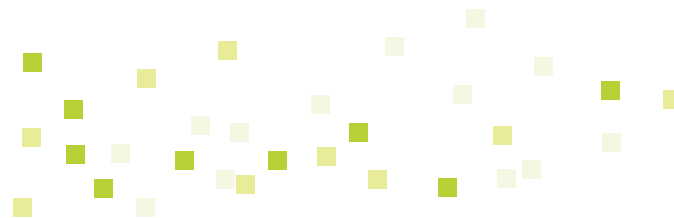
"Binnen het Stadsarchief is de spraakherkenningstool gebruikt voor het ontsluiten van interviews met overlevenden van het bombardement van Rotterdam. Deze interviews zijn te beluisteren op de *Brandgrens*-website. Daarnaast is in samenwerking met RTV Rijnmond gewerkt aan de ontsluiting van radio-materiaal uit de periode 1980-1986. De toepassing hierbij is tweeledig: het materiaal kan doorzocht worden door de belangstellende onderzoeker, maar daarnaast wordt het geïndiceerde materiaal ook gebruikt door RTV Rijnmond zelf, waar het als 'naslagmateriaal' gebruikt wordt door de huidige programmamakers.

CHoralPlus heeft meerwaarde voor het archief, omdat beeld- en geluidsmateriaal dat voorheen alleen door het bekijken/beluisteren van veel materiaal kon worden ontsloten, nu op betrekkelijk eenvoudige wijze geïndexeerd beschikbaar kan worden gesteld, waardoor er gericht in het materiaal is te zoeken. De metadata zijn te beperkt om een volledig beeld van de opname te geven, waardoor relevante data voor onderzoekers en andere geïnteresseerden op de plank bleven liggen.



De resultaten van de spraakherkenning zijn niet foutloos, daarvoor is er te veel variatie tussen de verschillende sprekers en opnames. Maar de resultaten zijn wel goed genoeg om geautomatiseerd zoeken in de opname mogelijk te maken. Bovendien zal de transcriptie van CHoralPlus steeds beter worden naarmate de gebruikers de uitkomsten bevestigen of indien nodig verbeteren.

Al met al is het Stadsarchief erg tevreden met de functionaliteiten van CHoralPlus, omdat het voorheen slecht toegankelijke beeld- en geluidscollecties helpt ontsluiten. De automatische transcripties geven een goede indicatie waar je relevante informatie kunt vinden."



Projectteam

Marijn Huibregts (*XMI Cross Media Interaction*)
Franciska de Jong (*Universiteit Twente*)
Mies Langelaar (*Stadsarchief Rotterdam*)
Roeland Ordelman (*XMI Cross Media Interaction*)
Jantje Steenhuis (*Stadsarchief Rotterdam*)

Projectteam CHoral

Wilma van Giesbergen (*Stadsarchief Rotterdam*)
Mies Langelaar (*Stadsarchief Rotterdam*)
Willemijn Heeren (*Universiteit Twente*)
Franciska de Jong (*Universiteit Twente*)
Roeland Ordelman (*Universiteit Twente*)
Jantje Steenhuis (*Stadsarchief Rotterdam*)
Thijs Verschoor (*Universiteit Twente*)
Laurens van der Werff (*Universiteit Twente*)

Interesse?

Wilt u meer weten over CHoralPlus?
Surf dan naar www.catchplus.nl/choralplus





WITCHCRAFTplus

*What Is Topical in Cultural Heritage:
Content-based Retrieval Among Folksong Tunes*

Online zoeken in melodieën

Volksliedjes worden van generatie op generatie doorgegeven. Sommige liederen zijn al meer dan vijf eeuwen oud. Elke nieuwe generatie verandert ongemerkt de melodie van deze liedjes, waardoor verschillende varianten van dezelfde liedjes ontstaan. Een speciale melodieën-zoekmachine kan deze variaties helpen opsporen.

Liederens maken deel uit van het 'immaterieel erfgoed'. Wereldwijd zijn volksliederen verzameld, zowel in grote databases met geluidsopnamen als in tekstuele beschrijvingen van de liederen. De bekendste Nederlandse liedcollectie *Onder de Groene Linde* berust bij het Meertens Instituut en maakt deel uit van de Nederlandse Liederensbank. Deze databank ontsluit de metadata van een aantal liedcollecties, maar had een manco: de muzikale inhoud van de liederen kon niet doorzocht worden.

In het CATCH-project WITCHCRAFT is een zoekmachine ontwikkeld die bepaalt in welke mate melodieën op elkaar lijken. Hiermee is het mogelijk om een onbekende melodie te identificeren of om verschillende varianten van een liedje te vinden tussen duizenden andere melodieën.

De ontwikkelde software helpt onderzoekers bovendien beter begrijpen op welke manieren melodieën met elkaar verwant kunnen zijn.

In WITCHCRAFTplus werd deze zoekmachine uitontwikkeld en geschikt gemaakt voor gebruik online. Onderzoekers, muzikanten en andere gebruikers kunnen zoeken op variaties van melodieën in de Nederlandse Liederensbank, onder meer door zelf een MIDI-file te uploaden of een melodietje in te tikken. Hiernaast werd een online muziek-editor gebouwd, waarmee de collectiespecialisten van het Meertens Instituut en andere gebruikers klinkende of op scan zichtbare muziek kunnen invoeren in doorzoekbare muzieknotatie. Deze muzieknotatie kan in verschillende formaten worden geüp- en gedownload zodat zoveel mogelijk gebruikers, van onderzoekers tot muzikanten, het materiaal kunnen (her-)gebruiken.



**Martine de Bruin,
Meertens Instituut:**

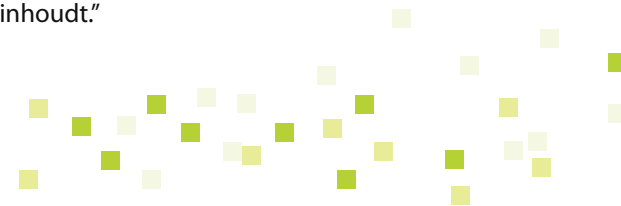
“Waarschijnlijk heeft iedereen wel eens ervaren hoe moeilijk het is een melodie te benoemen als je je alleen de muziek – en niet de tekst – herinnert. Voor onderzoekers naar liedcultuur was dit een dagelijks terugkerend probleem. Er werden vaak nieuwe teksten geschreven op bestaande melodieën, maar vind dan maar eens uit welke melodie is hergebruikt. Meer dan twee eeuwen lang is geprobeerd om melodieën opzoekbaar te maken, bijvoorbeeld door ze in kaartenbakken te ordenen. Als dat al lukte – dit is beduidend lastiger dan alfabetiseren – bestond altijd nog de grote beperking dat alleen op het beginnetje van een melodie kon worden gezocht. Daardoor was men uiteindelijk altijd aangewezen op het eigen geheugen.

Op dat laatste hebben we nu een belangrijke aanvulling. Met de kersverse melodieënzoekmachine is het mogelijk op allerlei manieren in muziek te zoeken en de eerste mooie resultaten zijn dan ook al binnen. Voor het eerst zijn we niet meer afhankelijk van de tekst of contextgegevens. Toch zijn niet al onze wensen vervuld: de zoekmachine smaakt ook naar meer. Dat betreft zowel de data als de



techniek. Een zoekmachine is alleen succesvol als er voldoende data zijn waarin gezocht kan worden. Voor WITCHCRAFT is gekozen voor het handmatig transcriberen van melodieën; dat levert de beste resultaten op, maar is wel zeer arbeidsintensief. We hebben zelf al ruim 7.500 melodieën ‘gedaan’, maar zouden ons corpus graag aanmerkelijk vergroten. Ook wordt er naar aanleiding van het WITCHCRAFT-project alweer nagedacht over andere algoritmes die de zoekfunctionaliteit zouden kunnen verbeteren. Daarnaast zijn er plannen om een Europees samenwerkingsproject te starten – melodieën komen vaak in meerdere landen voor en melodieënzoekmachines bieden eindelijk een kans migraties te volgen.

WITCHCRAFT heeft ons ook op een ander punt geholpen. Om een zo goed mogelijk algoritme te ontwikkelen, werd aan de onderzoekers gevraagd welke technieken zij gebruiken om overeenkomst tussen melodieën vast te stellen. Deze inzichten werden verwerkt en getest in de zoekmachine in wording. Het gevolg was dat niet alleen een optimaal algoritme ontstond, maar ook dat de onderzoekers zelf veel beter voor ogen staat wat een dergelijke gelijkenis inhoudt.”



Projectteam WITCHCRAFTplus

Martine de Bruin (*Meertens Instituut*)
Louis Grijp (*Meertens Instituut*)
Peter van Kranenburg (*Universiteit Utrecht, Meertens Instituut*)
Bouke Versteegh (*Meertens Instituut*)
Lysander Vogelzang (*Meertens Instituut*)
Ellen van der Grijn Santen (*Meertens Instituut*)
m.m.v. Frans Wiering (*Universiteit Utrecht*)
Chiel Arends (*Doelmatica*)
Timen van de Berg (*OGD*)

Projectteam WITCHCRAFT

Frans Wiering (*Universiteit Utrecht*)
Remco Veltkamp (*Universiteit Utrecht*)
Louis Grijp (*Universiteit Utrecht, Meertens Instituut*)
Anja Volk (*Universiteit Utrecht*)
Jörg Garbers (*Universiteit Utrecht*)
Peter van Kranenburg (*Universiteit Utrecht*)

Interesse?

Wilt u meer weten over WITCHCRAFTplus?
Surf naar www.catchplus.nl/witchcraftplus





MuSeUMPlus

MULTiple-collection SEarching Using Metadata Plus

Verbeterd zoeken met MuS en Geméén

Een complete database is een utopie. Altijd zijn er wel gegevens die in het verkeerde veld staan, inconsequent zijn of tekstuele of inhoudelijke onjuistheden bevatten. MuSeUMPlus maakt het mogelijk deze ‘vervuilde’ databases te doorzoeken en vervolgens de zoekresultaten te bewerken in een online werkomgeving.

De meeste erfgoedinstellingen beheren hun collectie in gestructureerde, specialistische databases. Deze informatiesystemen beschikken vaak over eigen zoekformulieren waarmee binnen bepaalde zoekvelden (auteur, datum, titel, etc.) naar specifieke termen gezocht kan worden. Als de relevante informatie niet in het juiste veld staat, zal het zoekformulier deze dus ook niet vinden. Zo loopt de gebruiker grote kans relevante items uit de collectie mis te lopen. In het CATCH-project MuSeUM is onderzocht hoe het best gezocht kan worden in ‘vervuilde’ databases. Via de traditionele, gestructureerde zoekmethodiek of via een Google-achtige, ongestructureerde benadering die alle aanwezige tekst meeneemt? De eindconclusie was dat zowel specialistische gebruikers als doorsneegebruikers de beste resultaten krijgen uit een combinatie van vrije en gestructureerde zoekmethoden.

De resultaten uit MuSeUM zijn in MuSeUMPlus verder ontwikkeld tot bruikbare toepassingen. Zoekmachine MuS maakt het mogelijk verschillende databases tegelijkertijd te doorzoeken, zowel binnen de aangegeven velden als in de overige aanwezige tekst. Ook onderlinge relaties binnen en tussen databases neemt *MuS* tot op zekere hoogte mee in zijn aanpak.

De erfgoedinstelling behoudt de controle over de presentatie van de gevonden items uit de collectie om te voorkomen dat externe gebruikers toegang krijgen tot gevoelige informatie. In de meest behouden vorm worden slechts de PID-codes (Persistent Identifiers) getoond. De gebruiker weet dan alleen dat de betreffende bron wellicht relevante informatie bevat en zal voor meer informatie persoonlijk het instituut moeten benaderen.

Naast MuS heeft het projectteam ook *Geméén* ontwikkeld. *Geméén* is een online werkomgeving waarin gebruikers in werkgroepverband o.a. de resultaten uit *MuS* kunnen beheren en bewerken. De zgn. *O_og werkgroepen* zijn voorbeelden van een toepassing van *MuS* waarbij alleen objectbeschrijvingen worden getoond. Hiermee kunnen bijvoorbeeld relevante objecten voor een specifiek thema of een bepaalde tentoonstelling verzameld worden. De door de werkgroep uitgevoerde zoekopdrachten worden continu geüpdatet, zodat veranderingen in de collectie direct in de zoekresultaten zichtbaar zijn.



Vincent de Keijzer,
Gemeentemuseum
Den Haag:

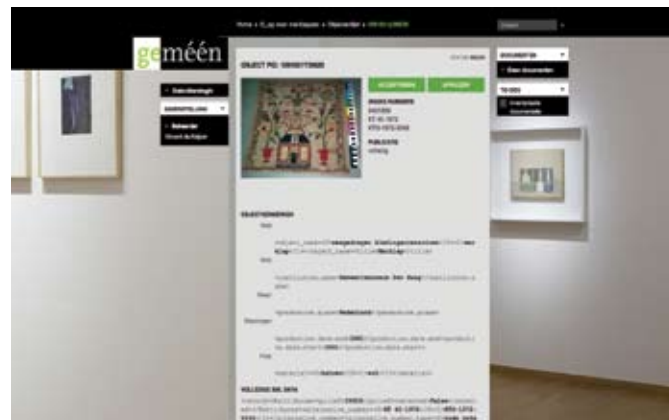
“Stel je bent een liefhebber van merklappen, ex-librissen, modelboten of Delfts aardewerk. Of je verzamelt de vroege Haagse School, werken van kunstenaars uit Arnhem of afbeeldingen van schoenen. Hoe weet je dan

wat er van je gading is in de Nederlandse museumcollecties? Je zou denken dat je met zoeken in Google een heel eind kan komen, in ieder geval voor een eerste indruk. Dat is echter helemaal niet het geval. Blijkbaar zijn onze museumschatkamers zelfs voor de allesvinder Google stevig afgeschermd. Er zit niets anders op dan zelf verschillende musea aan te schrijven met verzoek om inzage van de collectie.

Bevlogen verzamelaars laten zich niet afschrikken door deze hoge drempels, maar steeds meer gebruikers weigeren zich neer te leggen bij deze situatie. Zij zien de museumcollecties als openbaar bezit waar zij ongehinderd toegang toe willen hebben. Musea wijzen op inhoudelijke, technische en organisatorische problemen die het grootschalig beschikbaar stellen van informatie over de collectie verhinderen. Hoe onterecht misschien ook, veel gebruikers zien dit als

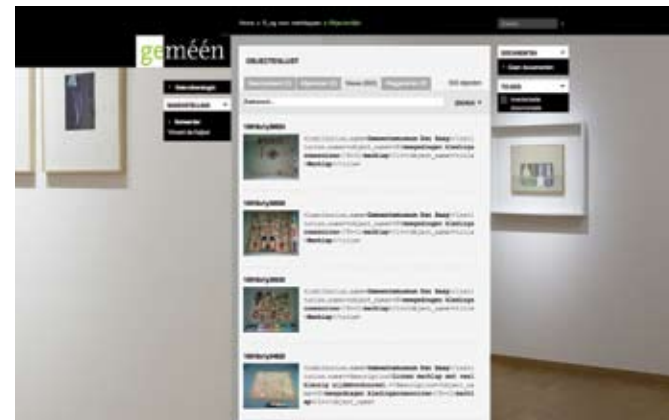
ouderwets protectionisme en willen zelf beoordelen of de gegevens in de museale systemen relevant zijn voor hen of niet.

MuSeUMPlus legt de basis voor een oplossing die voorzichtige musea én ongeduldige gebruikers tegemoet komt. Musea kunnen snel toegang verschaffen tot hun informatie én controle houden over wat zij ervan willen presenteren. Liefhebbers en verzamelaars hebben de mogelijkheid om alle informatie te doorzoeken en relevante objecten te verzamelen en naar eigen inzicht te bewerken. In ieder geval worden zij geholpen met het antwoord op hun vraag: waar vind ik merklappen, ex-librissen, modelboten, Delfts aardewerk, vroege Haagse School, werken van kunstenaars uit Arnhem of afbeeldingen van schoenen?”



Projectteam MuSeUMPlus

Hanno Lans (*Datascape*)
Frodo Schering (*platform Gemeen*)
Jur de Vries (*Triquanta*)
Volkan Florchinger (*Rhizom*)
Vincent de Keijzer (*Gemeentemuseum Den Haag*)
Marijn Koolen (*Universiteit van Amsterdam*)



Projectteam MuSeUM

Avi Arampatzis (*Universiteit van Amsterdam*)
Jaap Kamps (*Universiteit van Amsterdam*)
Marijn Koolen (*Universiteit van Amsterdam*)
Vincent de Keijzer (*Gemeentemuseum Den Haag*)
Nir Nussbaum (*Universiteit van Amsterdam*)
Maarten de Rijke (*Universiteit van Amsterdam*)



Interesse?

Wilt u meer weten over MuSeUMPlus?
Surf dan naar www.catchplus.nl/museumplus





UPR en ZieOok

User Profile Repository en ZieOok

Aanbevelingen op basis van een persoonlijk profiel

Een bezoeker die een schilderij van de Munttoren wel kan waarderen, heeft misschien ook interesse in andere stadsgezichten van Amsterdam. De applicatie ZieOok kan de bezoeker op basis van zijn persoonlijke profiel suggesties doen voor andere items uit de collectie.

In het CATCH-project CHIP (*Cultural Heritage Information Presentation*) onderzochten wetenschappers van de Technische Universiteit Eindhoven en het Telematica Instituut hoe bezoekers van het Rijksmuseum persoonlijke aanbevelingen konden krijgen met behulp van semantische technieken. De demo's uit dit project zijn binnen CATCH-Plus als uitgangspunt genomen voor de ontwikkeling van nieuwe tools die inmiddels gebruikt worden door zowel het Amsterdam Museum als het Theater Instituut Nederland.

De *User Profile Repository* (UPR) slaat de persoonlijke profielen van gebruikers op. Naast persoonsgegevens kan zo'n profiel ook beschrijvingen en verwijzingen naar voorkeursobjecten bevatten. Deze voorkeursobjecten heeft de gebruiker eerder gewaardeerd door ze wel of niet aan de favorieten toe te voegen. Daarnaast houdt de UPR ook

statistieken bij van de bezoekfrequentie van de gebruiker aan een bepaald object.

De bezoeker heeft uiteraard zelf toegang tot zijn eigen profiel. Hij kan zich via een webapplicatie registreren en inloggen met bijvoorbeeld zijn account van Facebook, Google of Twitter. Vervolgens kan hij persoonlijke ratings geven aan verschillende items uit de collectie op de website van de erfgoedinstelling. De bezoeker bepaalt zelf welke erfgoedinstellingen hij toegang verleent tot zijn profiel, zodat hij gepersonaliseerde diensten aangeboden kan krijgen. Zijn profiel is voor deze diensten altijd anoniem.

De UPR is verbonden met ZieOok, het aanbevelingsplatform dat is ontwikkeld door het Instituut voor Beeld en Geluid. Websitebezoekers kunnen erfgoedinstellingen die gebruikmaken van ZieOok toegang verlenen tot hun persoonlijke profiel. Op basis van dat profiel kan de instelling gepersonaliseerde aanbevelingen doen uit de eigen collectie. Ook kan de erfgoedinstelling zien hoe items uit de collectie bezocht en gewaardeerd worden. In principe is het zelfs mogelijk aanbevelingen te doen, die collectie- en instellingsgrenzen overschrijden.





**Judith van Gent,
Amsterdam Museum:**

“Twee jaar geleden heeft het Amsterdam Museum de collectie als digitaal depot online gezet. Hoewel de data soms incompleet zijn, hebben we ervoor gekozen om de héle collectie via het internet toegankelijk te maken.

Daarbij stimuleren wij het

gebruik en hergebruik van onze data.

De belangrijkste reden hiervoor is dat we vinden dat publieke collecties toegankelijk moeten zijn voor het publiek. Daarnaast is de collectie zowel historisch als kunsthistorisch internationaal van belang. Iedereen moet de objecten kunnen vinden en gebruiken. Omdat slechts 20% van de collectie fysiek toegankelijk is, biedt online openstelling van de collectie een goed alternatief om onze objecten zichtbaar te maken.

Nu onze collectie online toegankelijk is, zoeken we naar manieren om het gebruik ervan aantrekkelijker te maken. We zoeken naar toepassingen waarmee de collectie nog toegankelijker wordt.

Daarnaast behoort het stimuleren van innovatie en kennisdelen tot ons beleid en doen we daarom mee aan verschillende innovatieve projecten waarin onze data worden gebruikt. ZieOok past daarom perfect in onze doelstellingen.

Op onze verhalenwebsites *Geheugen van Oost*, *Buurtwinkels* en *Hart van Amsterdam Museum* wordt het lezen van verhalen gestimuleerd door aanbevelingen naar andere verhalen in die sites te doen. Wij hebben de indruk dat deze functionaliteit daadwerkelijk het gebruik van de sites stimuleert. Dit is de reden dat ZieOok ons aanspreekt, daar deze applicatie vergelijkbare functionaliteit aanbiedt voor onze collectie online. Met de aanbevelingen van ZieOok wordt de bezoeker uitgenodigd om ook andere objecten uit de collectie te leren kennen. Wanneer ook digitale depots van andere erfgoedinstellingen aan ZieOok worden toegevoegd, wordt de functionaliteit nog interessanter. Op dat moment zal het gebruik van verschillende collecties over en weer worden gestimuleerd en zal de gebruikerservaring nog rijker worden.”



Projectteam

Patricia Alkhoven (*CATCHPlus Projectbureau*)
Hennie Brugman (*CATCHPlus Projectbureau*)
Mart Trautwein (*Gridline*)
Job Tiel Groenestege (*Gridline*)
Siem Vaessen (*Instituut voor Beeld en Geluid*)
Judith van Gent (*Amsterdam Museum*)
Marijke Oosterbroek (*Amsterdam Museum*)
Pim Luiten (*Theater Instituut Nederland*)
Roelof Brandsma (*Theater Instituut Nederland*)

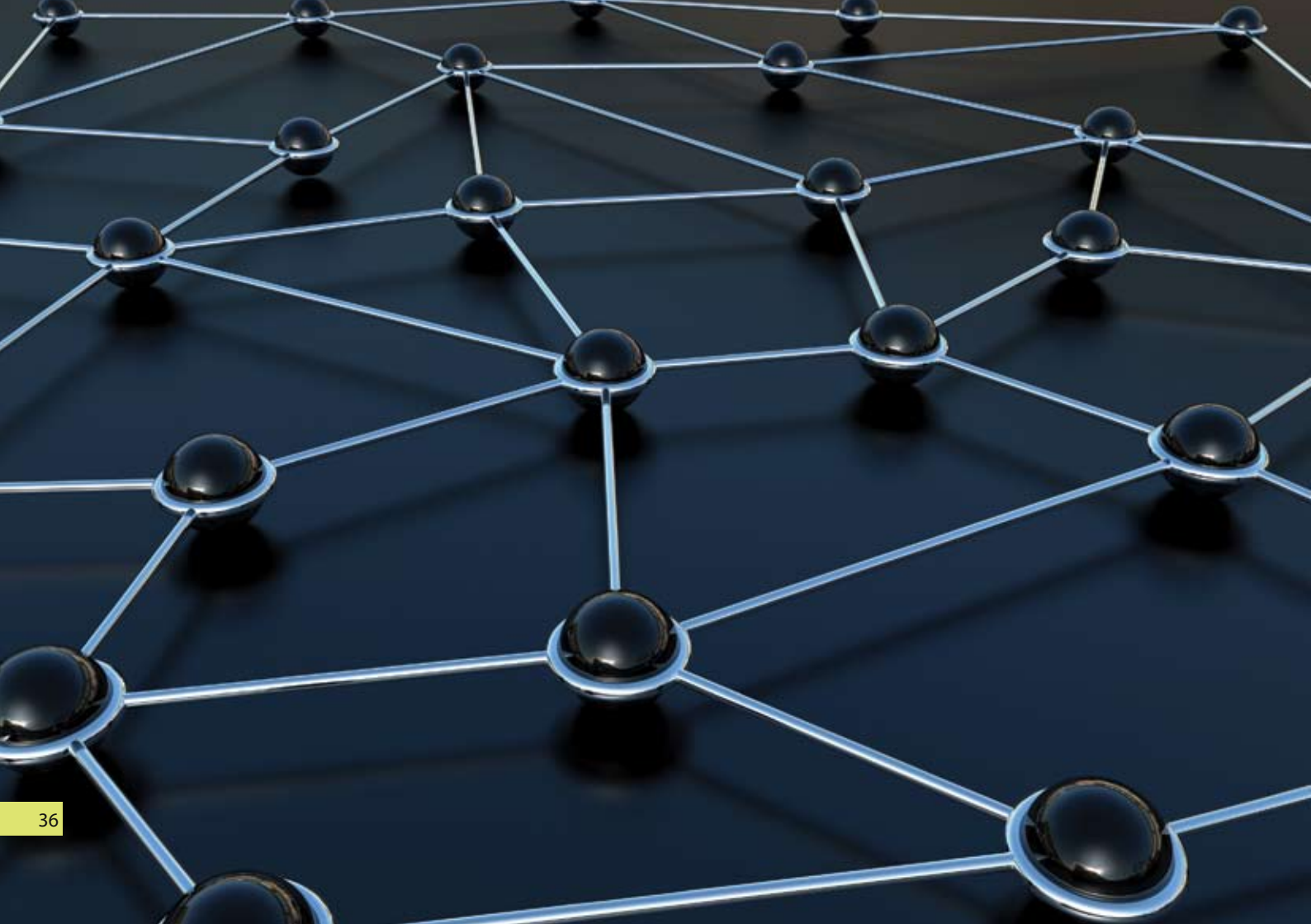
Projectteam CHIP

Lora Aroyo (*Technische Universiteit Eindhoven*)
Paul de Bra (*Technische Universiteit Eindhoven*)
Rogier Brussee (*Telematica Instituut*)
Peter Gorgels (*Rijksmuseum Amsterdam*)
Lloyd Rutledge (*Telematica Instituut*)
Peter Sigmond (*Rijksmuseum Amsterdam*)
Natalia Stash (*Technische Universiteit Eindhoven*)
Mettina Veenstra (*Telematica Instituut*)
Yiwen Wang (*Technische Universiteit Eindhoven*)

Interesse?

Wilt u meer weten over UPR en ZieOok?
Surf dan naar www.catchplus.nl/upr





Multiply

Trefwoordsuggesties op basis van automatische tekstanalyse

Archivarissen en documentalisten gebruiken trefwoorden om stukken uit hun collectie te beschrijven. Als de computer archiefmateriaal automatisch kan indexeren, kan deze ook zelf suggesties doen voor relevante trefwoorden en de documentalist veel werk uit handen nemen. Dat is precies wat Multiply doet.

Multiply bouwt voort op het CATCH-project CHOICE, dat een acroniem is voor *CHarting the informatiOn landscape employIng ContExt information*. Hierin is gewerkt aan een applicatie die documentalisten van het Instituut voor Beeld en Geluid ondersteunt bij het indexeren en vinden van audiovisuele documenten.

In Multiply is deze applicatie uitgewerkt tot software die documentalisten in staat stelt automatische trefwoordsuggesties te analyseren en te gebruiken in het beschrijvingsproces. De gesuggereerde trefwoorden zijn afkomstig uit de thesaurus van Beeld en Geluid, die ruim 150 duizend termen bevat. Handmatige selectie van trefwoorden is een erg tijdrovend en subjectief proces, dus het geautomatiseerde systeem kan de documentalist veel werk uit handen nemen. De thesaurus van Beeld en Geluid (de GTAA) is in de

context van Multiply aangesloten bij OpenSKOS (zie blz. 42-43), waardoor ook externe gebruikers kunnen beschikken over GTAA.

De applicatie baseert zijn trefwoordsuggestie op de resultaten van een automatische tekstanalyse van contextuele tekstdocumenten. Deze documenten zijn opgeslagen in een zogenaamde contextdatabase en bestaan voor Beeld en Geluid uit bijvoorbeeld omroepwebsites, wiki's en programmeergegevens. De applicatie kan ook suggesties doen op basis van tijdgebaseerde metadata, zoals ondertitelfiles of de uitkomst van automatische spraakherkennings-technologie.





Johan Oomen, Instituut voor Beeld en Geluid:

“Beeld en Geluid zet veel mankracht in voor het beschrijven van audiovisueel materiaal, zoals radio- en televisieprogramma’s. Het maken van deze beschrijvingen is tijdrovend en daardoor kostbaar. Tegelijk stroomt er jaarlijks meer en meer content

het digitale archief binnen en verwachten gebruikers veel meer fijnmazige toegang. Inzet van technologie die het beschrijvingsproces efficiënter kan maken, en betere toegang kan bieden is derhalve noodzakelijk.

Binnen Beeld en Geluid is gewerkt aan de annotatieomgeving, die MASS genoemd wordt: het *Media Archivist Support System*. Op basis van tekstuele input biedt het systeem automatische suggesties uit de GTAA-thesaurus. Documentaristen maken gebruik van de MASS-interface om suggesties te analyseren en de correcte trefwoorden toe te kennen. Deze semi-automatische manier van beschrijven versnelt het annotatieproces aanzienlijk.

In de huidige versie van MASS (oktober 2012) wordt gebruik gemaakt van ondertitelfiles. In de toekomst zal ook een koppeling gemaakt worden met de contextdatabase van Beeld en Geluid.

Verder zal de techniek die binnen Multiply ontwikkeld is een belangrijke rol spelen bij een in 2012 gestart initiatief om te komen tot een Nederlandse Cultuurthesaurus. Beeld en Geluid, de Rijksdienst voor het Cultureel Erfgoed en Naturalis hebben afgesproken de komende jaren de krachten te bundelen aangaande het ontwikkelen van techniek om thesauri op nationale schaal te beheren.”



Projectteam Multiply

Jaap Blom (*Instituut voor Beeld en Geluid*)
Hennie Brugman (*CATCHPlus Projectbureau*)
Vincent Huis in 't Veld (*Instituut voor Beeld en Geluid*)
Willem Melder (*Instituut voor Beeld en Geluid*)
Johan Oomen (*Instituut voor Beeld en Geluid*)
Roeland Ordelman (*Instituut voor Beeld en Geluid*)
Alma Wolthuis (*Instituut voor Beeld en Geluid*)
Picturae
Rijksdienst voor het Cultureel Erfgoed (RCE)

Projectteam CHOICE

Hennie Brugman (*Max Planck Instituut*)
Luit Gazendam (*Instituut voor Beeld en Geluid*)
Veronique Malaise (*Vrije Universiteit*)
Johan Oomen (*Instituut voor Beeld en Geluid*)
Mettina Veenstra (*Telematica Instituut*)

Interesse?

Wilt u meer weten over Multiply?
Surf dan naar www.catchplus.nl/multiply



STITCHPlus

Semantic Interoperability To access Cultural Heritage

Zinvol zoeken in bestaande collecties

Van oudsher werken veel erfgoedinstellingen vaak met meerdere thesauri om de eigen collectie zo goed mogelijk te ontsluiten. Door verschillende thesauri geautomatiseerd met elkaar te verbinden, maken de applicaties van STITCHPlus het mogelijk bestaande collecties zinvoller te doorzoeken.

Voor de computer is tekst niks meer dan een lange rij eentjes en nulletjes. Hij heeft geen wezenlijk begrip van de inhoud van de tekst of databases waarin erfgoedinstellingen hun collectie beheren. De laatste jaren begint hier geleidelijk verandering in te komen: wetenschappers ontwikkelen algoritmes waarmee de computer zelfstandig relaties en betekenis uit tekst kan afleiden. Deze semantische technologie maakt het mogelijk geautomatiseerd tekstbestanden te analyseren.

In het CATCH-project STITCH hebben wetenschappers van de Vrije Universiteit onderzocht hoe semantische technieken de verschillende thesauri en begrippenlijsten van de Koninklijke Bibliotheek geautomatiseerd konden verbinden. Dit project heeft onder andere geleid tot de ontwikkeling van de vocabulairebank OpenSKOS, waarover op de volgende pagina's meer is te lezen.

Het team van STITCHPlus is verder gegaan waar STITCH ophield. Ze hebben twee suites ontwikkeld. De eerste bevat een workflow waarin bestaande thesauri van hun oorspronkelijke locatie worden gekopieerd en omgezet naar SKOS, een universeel formaat waarvan ook de vocabulairebank OpenSKOS gebruikmaakt. Deze nieuwe thesaurus wordt zowel lokaal bij de erfgoedinstelling opgeslagen als toegevoegd aan OpenSKOS. Dit maakt verschillende collecties eenvoudiger met elkaar te verbinden.

De tweede suite bestaat uit een reeks applicaties die namen (van personen, locaties, bedrijven, etc.) kunnen herkennen in teksten en kunnen koppelen aan termen in SKOS-thesauri of aan externe bronnen als Wikipedia. De gebruiker kan deze automatisch gelegde koppelingen controleren en indien nodig ook corrigeren.

Projectteam STITCHPlus

René van der Ark (*Koninklijke Bibliotheek*)
Paul Doorenbosch (*Koninklijke Bibliotheek*)
Willem-Jan Faber (*Koninklijke Bibliotheek*)
Dirk Kramer (*Koninklijke Bibliotheek*)
Henk Matthezing (*Koninklijke Bibliotheek*)
Theo van Veen (*Koninklijke Bibliotheek*)

Projectteam STITCH

Paul Doorenbosch (*Koninklijke Bibliotheek*)
Marjolein van Gendt (*Vrije Universiteit*)
Frank van Harmelen (*Vrije Universiteit*),
Antoine Isaac (*Vrije Universiteit, Kon. Bibliotheek*)
Henk Matthezing (*Koninklijke Bibliotheek*)
Lourens van der Meij (*Vrije Universiteit, Kon. Bibliotheek*)
Stefan Schlobach (*Vrije Universiteit*)
Shenghui Wang (*Vrije Universiteit, Kon. Bibliotheek*)
Claus Zinn (*Max Planck Instituut*)
Dirk Kramer (*Koninklijke Bibliotheek*)

Interesse?

Wilt u meer weten over STITCHPlus?
Surf dan naar www.catchplus.nl/stitchplus





OpenSKOS

Een online platform voor thesauri

Het ontwikkelen en beheren van thesauri om collecties mee te beschrijven, is tijdrovend werk. En als deze dan eenmaal bestaan, zijn ze meestal niet geschikt voor verschillende software-tools en kosten kleine updates relatief veel technische aanpassingen. Het online platform OpenSKOS helpt deze problemen de wereld uit.

De laatste jaren worden steeds meer thesauri uit de geesteswetenschappen en de erfgoedwereld omgezet naar SKOS (*Simple Knowledge Organization System*). Deze standaard voor vocabulaires maakt gebruik van semantische webtechnieken en Linked Open Data, waardoor dergelijke kennisbronnen beter kunnen worden gedeeld, en waardoor er over het web naar kan worden gelinkt.

Binnen CATCHPlus is OpenSKOS ontwikkeld, een online platform voor thesauri in het SKOS-formaat. OpenSkos is het resultaat van een samenwerking tussen de Vrije Universiteit, de Rijksdienst voor het Cultureel Erfgoed en bedrijven als Adlib, Picturae en Trezorix. Onder andere het Nederlands Instituut voor Beeld en Geluid en het Nationaal Archief maken al gebruik van dit platform. OpenSKOS biedt veel voordelen voor zowel de ontwikkelaars als de gebruikers van thesauri.

De ontwikkelaars van thesauri kunnen eenvoudig de zelfontworpen vocabulaires uploaden en delen met andere gebruikers. Updates kan de ontwikkelaar zelf doorvoeren, zonder tussenkomst van derden. Concepten uit de eigen vocabulaires kunnen eenvoudig gekoppeld worden aan bestaande vocabulaires.

De gebruiker kan in OpenSKOS op zoek gaan naar vocabulaires die passen bij zijn eisen en wensen. Bestaande thesauri kunnen eenvoudig hergebruikt worden, waarmee een instelling veel ontwikkelkosten kan besparen. De thesauri uit OpenSKOS kunnen aangesproken worden door veel verschillende applicaties en zijn zo niet beperkt in hun toepassingen. De gebruiker werkt altijd met de laatste versie van de thesauri, omdat updates direct doorgevoerd worden.



Projectteam

versie 1

Hennie Brugman (*Meertens Instituut*)
Lourens van der Meij (*Vrije Universiteit*)

specificatie-fase

Patricia Alkhoven (*Meertens Instituut*)
Hennie Brugman (*Meertens Instituut*)
Bert Degenhart-Drenth (*Adlib*)
Mark Lindeman (*Picturae*)
Sander van der Meulen (*Trezorix*)
Hans Nederbragt (*Trezorix*)
ondersteund en gefaciliteerd door RCE

versie 2

Hennie Brugman (*Meertens Instituut*)
Mark Lindeman (*Picturae*)

Interesse?

Wilt u meer weten over OpenSKOS?
Surf dan naar www.catchplus.nl/openskos





DocChecker

Automatische trefwoordsuggestie

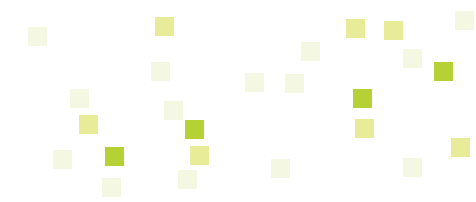
Concepten uit thesauri van erfgoedinstellingen komen vaak ook al voor in gearchiveerde tekstdocumenten. Archivarissen zijn vaak veel tijd kwijt om de juiste thesauri-termen te koppelen aan deze documenten. De DocChecker kan een deel van deze taak overnemen.

De DocChecker is ontwikkeld bij de Rijksdienst voor het Cultureel Erfgoed (RCE) voor tekstbestanden in verschillende formaten, zoals Word, PDF, html en Excel. Door nieuwe thesauri of nieuwe versies daarvan in SKOS-formaat te uploaden, kan de gebruiker zelf bepalen welke thesauri worden gebruikt. Daarnaast kunnen de thesauri in de DocChecker nog verrijkt worden met data uit andere databases, zoals synoniemen, verkleinwoorden en vervoegingen.

De gebruiker kan de DocChecker gebruiken in een web-service waarin hij profielen aan kan maken met daarin de thesauri die meegenomen moeten worden in de analyse. Vervolgens koppelt de DocChecker automatisch de concepten uit de thesauri aan de tekstfragmenten in de ingevoerde documenten. De DocChecker stelt de gebruiker relevante thesaurus-termen voor die hij kan valideren. De gematchte gegevens worden opgeslagen in een metadata-bestand waarin de links zijn opgenomen tussen het betreffende item en de geselecteerde concepten.

Projectteam

Antal van den Bosch (*Universiteit Tilburg*)
Hennie Brugman (*CATCHPlus Projectbureau*)
Kees Hendriks (*RCE*)
Dirk Houtgraaf (*RCE*)
Guus Lange (*RCE*)
Sander van der Meulen (*Trezorix*)
Hans Nederbragt (*Trezorix*)
Hans Paijmans (*Universiteit Tilburg*)
Carla Schulte (*RCE*)



Interesse?

Wilt u meer weten over DocChecker?
Surf dan naar www.catchplus.nl/docchecker





Annotatie Repository Service

Gestandaardiseerde annotaties online delen en hergebruiken

Bij annotaties denk je misschien in eerste instantie aan tekstuele beschrijvingen, maar in steeds meer gevallen kan het ook gaan om webpagina's, foto's of multimediale opnames. In dit project is gewerkt aan de uniformiteit van deze annotaties, zodat ze optimaal kunnen worden gebruikt in interne en externe samenwerkingsverbanden.

Dat het concept annotaties breder moet worden opgevat dan alleen tekstuele beschrijvingen, bewijzen de CATCHPlus-toepassingen uit dit boekje. In SCRATCH4All worden delen van gescande handschriften getranscribeerd. In CHoralPlus worden de uitkomsten van automatische spraakherkenning gekoppeld aan audio- of videofragmenten. En de DocChecker koppelt de uitkomsten van taalanalyse-algoritmen aan woorden in een tekst. Annotaties kunnen in de praktijk dus erg divers zijn.

De CATCHPlus Annotatie Repository en Service maken het mogelijk om al deze verschillende typen annotaties op uniforme wijze bij elkaar te brengen en te doorzoeken. Op deze manier kunnen gebruikers voortbouwen op elkaars resultaten, door ze als annotaties aan elkaar door te geven. De algoritmes van de DocChecker kunnen bijvoorbeeld

worden toegepast op de resultaten van de spraakherkenning van CHoralPlus of de handschriftherkenning van SCRATCH4All.

Daarnaast levert het zoeken in annotaties een resultaatlijst op met verwijzingen naar onderdelen van (online) erfgoedobjecten bij mogelijk verschillende instellingen.

Voor het standaardiseren van het formaat voor annotaties heeft CATCHPlus aansluiting gezocht bij het Open Annotation Collaboration, een internationaal samenwerkingsverband dat tracht te komen tot een generiek model voor annotaties. Het annotatie-model van dit consortium is webgebaseerd en sluit aan bij de Open Data-beweging. Met hen werkt CATCHPlus aan het model in wording.

Samen met Seecr werkte CATCHPlus aan de *Open Annotation Server*. Dit is een webservice waarin annotaties geüpload, doorzocht en uitgewisseld kunnen worden. De annotaties zijn bovendien online als data te gebruiken; daarnaast kunnen verschillende instellingen via een zogenaamde 'data harvester' annotaties uitwisselen.

Projectteam

Hennie Brugman (*CATCHPlus Projectbureau*)

Erik Groeneveld (*Seecr*)

Johan Jonkers (*Seecr*)

Open Annotation Collaboration



Interesse?

Meer over Annotatie Repository Service?
Surf dan naar www.catchplus.nl/annotation



PID's

Persistent Identifiers

Vertrouwen op unieke en blijvende identifiers

In de archiefsystemen van erfgoedstellingen wordt met unieke namen of nummers (identifiers) verwezen naar fysieke of digitale objecten uit de collectie. Ook thesaurus-termen en metadata-beschrijvingen hebben hun eigen identifier waarmee ze eenvoudig teruggevonden kunnen worden. Maar hoe garandeer je dat deze identifiers werkelijk uniek en persistent zijn?

Erfgoedinstellingen hanteren meestal verschillende identifier-systemen voor verschillende informatiesystemen. Alle objecten uit de collectie hebben hun eigen systeem, net als de thesaurustermen, de metadata-beschrijvingen en indien aanwezig ook de annotaties en de gebruikersprofielen. Binnen de eigen instelling zijn deze identifiers uniek, maar wat gebeurt er als de instelling gaat samenwerken met andere partijen? Of als de instelling gaat werken met een ander databasemanagementsysteem met een ander identifier-systeem?

Deze praktijkvoorbeelden tonen de noodzaak aan van de garantie dat identifiers werkelijk uniek zijn en in de loop der tijd niet zullen verouderen. Als een identifier verwijst

naar een object uit de collectie van een andere instelling, moet de gebruiker dit object altijd kunnen vinden, óók als de andere instelling verandert van databasesysteem of webserver. Het toenemende belang van internet en samenwerkingsverbanden als CATCHPlus hebben zo geleid tot de noodzaak om oplossingen voor Persistente Identifiers (PID's) te ontwikkelen.

De PID-oplossing die is ontwikkeld binnen CATCHPlus is gebaseerd op bestaande en veelbeproefde Handle-technologie. Handle biedt een wereldwijd-werkende 'resolver'-dienst aan: als je een persistente identifier naar de dienst opstuurt, wordt je automatisch doorgeleid naar de actuele URL die bij de identifier hoort. CATCHPlus heeft een webservice ontwikkeld waarin de gebruiker eenvoudig de koppeling tussen PID's en URL's kan beheren.

Het is natuurlijk van groot belang dat de resolver altijd bereikbaar is, anders is de collectie niet toegankelijk voor de gebruikers. De Handle-technologie garandeert deze toegang met meerdere (redundante) kopieën van de resolver-dienst. Rekencentrum SARA heeft een samenwerkings-overeenkomst met andere Europese rekencentra, zodat het mogelijk is PID's duurzaam te hosten en te resolvablen, zelfs als het object zelf of de eigenaar ervan niet meer bestaat. Momenteel start CATCHPlus een tweejarig project met een achttal grote instellingen en organisaties. Doel van dit project is om bij deze deelnemers het PID-systeem in te voeren als een op maat gemaakte dienst.

Projectteam

Hennie Brugman (*Meertens Instituut*)

Pieter van Beek (*SARA*)

Tibor Kalman (*GWDG*)

Daan Broeder (*Max Planck Instituut, namens CLARIN*)

Eric Auer (*Max Planck Instituut*)

Jan Pieter Kunst (*Meertens Instituut*)

en vele anderen waren/zijn betrokken

Interesse?

Wilt u meer weten over PID's?

Surf dan naar www.catchplus.nl/pids





Op weg naar een erfgoed-makelaar

CATCHPlus richtte zich, in tegenstelling tot CATCH, op valorisatie en verzilvering van onderzoeksresultaten. Met andere woorden, het bouwen en implementeren van tools en diensten om erfgoedcollecties beter toegankelijk te maken en houden.

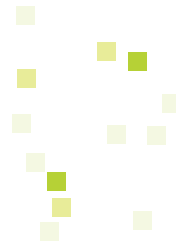
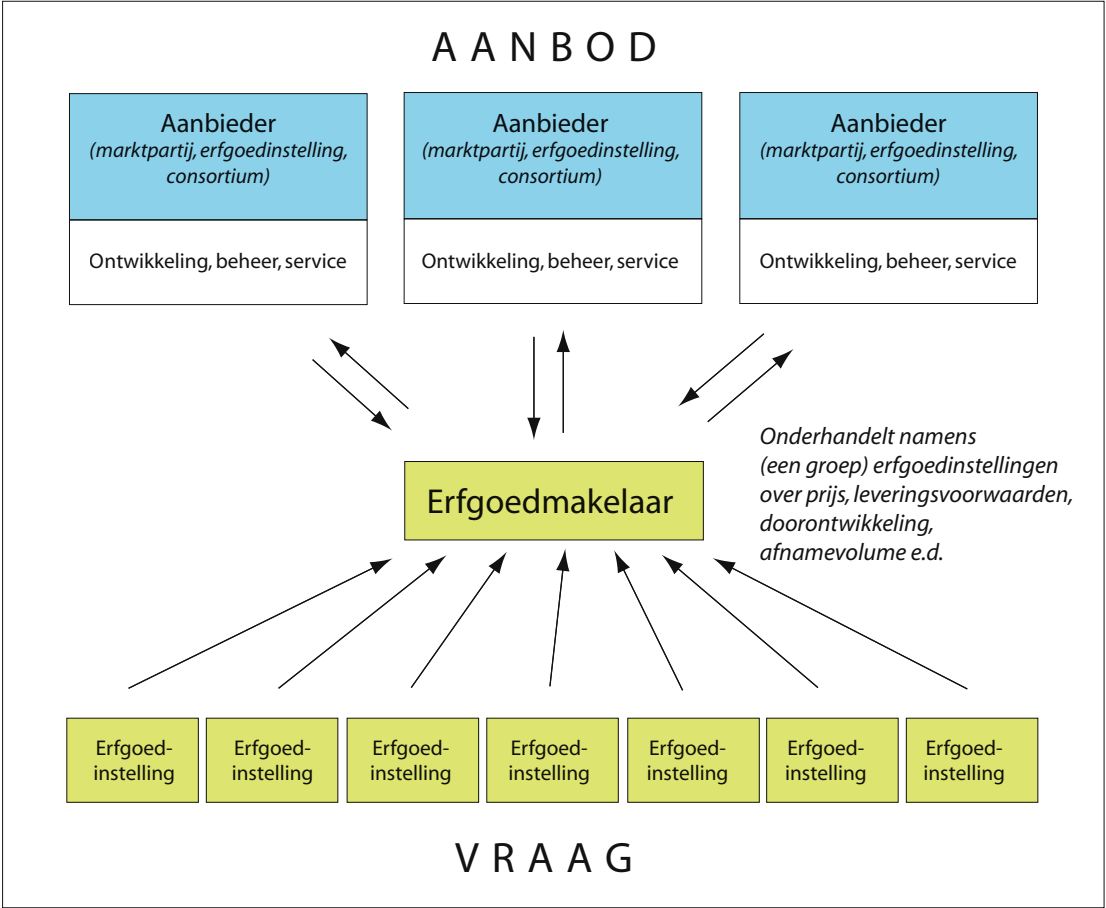
Om die valorisatie te realiseren werd van de betrokken erfgoedinstellingen een businessplan gevraagd voor de door hen ontwikkelde 'tool' of dienst. Hierin moesten zij ingaan op de wijze waarop de erfgoedinstelling de software dacht te beheren en exploiteren na afloop van het project.

Uit analyses van deze businessplannen en de markt blijkt dat CATCHPlus –net als de erfgoedsector in het algemeen– staat voor een belangrijk vraagstuk. Tal van erfgoedinstellingen experimenteren met en investeren in de ontwikkeling van (prototypes van) digitale producten en diensten waarmee de digitale infrastructuur voor het erfgoed wordt versterkt. Maar dat biedt nog geen garantie voor duurzaam beheer, aanbod en doorontwikkeling van die tools. De betrokken partijen ontbreekt het vaak aan kennis en capaciteit om duurzame borging –het op lange termijn op grote schaal beschikbaar houden van die producten en diensten– te realiseren, zo bleek binnen CATCHPlus. Partijen die deze taak wel op zich kunnen en willen

nemen, dienen zich niet als vanzelfsprekend aan. Andere erfgoedinstellingen en marktpartijen blijken vaak niet op de hoogte van het bestaan en van de waarde van de tools en spin-offs.

CATCHPlus ontbreekt het aan een partij die namens en in het belang van de erfgoedinstellingen een aanbieder kan vinden voor de door erfgoedinstellingen ontwikkelde tools en diensten. Een soort makelaar dus, die een link kan leggen tussen de ontwikkeling van kennis, prototypes en demo's, en de doorontwikkeling en exploitatie daarvan op de langere termijn. Hoewel er wel bedrijven actief zijn in CATCHPlus, is er tot nu toe geen organisatie naar voren getreden die zich actief toelegt op de vermarkting van projectresultaten die bij een erfgoedinstelling in beheer zijn of op de plank liggen.

Kennisland en DEN hebben daarom in nauwe samenwerking met het CATCHPlus Projectbureau een voorstel geschreven voor een erfgoedbrede oplossing voor de borging van de projectresultaten na afloop van het project. Hiermee wordt een bouwsteen gelegd voor een brede ICT-ondersteunende instelling, een *shared service*-organisatie voor het erfgoed, in navolging van de sectoren onderwijs (Kennisnet) en wetenschap (Surfnet). Hierin zouden gemeenschappelijke taken op basis van een nieuw, kostendekkend en indien succesvol, winstgevend, businessmodel kunnen worden samengebracht waarna de winst terugvloeit naar de sector.



Samen hebben we verschillende scenario's in kaart gebracht:

- Een tijdelijke makelaar die zich alleen met CATCHPlus bezighoudt, is wellicht het snelst en eenvoudigst te realiseren. Deze constructie sluit echter niet aan bij de visie en missie van de makelaar. Het ligt niet erg voor de hand om bij zo'n tijdelijke makelaar een aparte entiteit (in de vorm van aparte rechtspersoon of afdeling binnen een bestaande organisatie) in het leven te roepen.
- Een makelaar die breder werkt dan alleen voor CATCH-Plus vereist een goede coördinatie. Indien deze makelaar gestuurd wordt door een consortium van samenwerkende erfgoedinstellingen, zal de werkdruk voor dit consortium al snel te groot worden. In dit scenario zal snel behoefte ontstaan aan een zelfstandige makelaar in plaats van aan een samenwerkingsverband dat taken onderling verdeelt. Deze makelaar als zelfstandige entiteit kan aangestuurd worden door een aantal erfgoedinstellingen.

Al met al lijkt een Erfgoedmakelaar als zelfstandige entiteit met een breed mandaat, uiteindelijk het best in staat om het collectieve belang van de erfgoedsector te behartigen en daadwerkelijk een bijdrage te leveren aan een sterke infrastructuur voor digitale erfgoedproducten en -diensten. In het voorstel is een ingroeimodel voorgesteld dat de makelaar in staat stelt te starten met CATCHPlus-resultaten, en vervolgens te groeien naar een bredere aanpak. Hierbij worden de taken van de makelaar (deels) door samenwer-

kende erfgoedinstellingen en andere partijen uitgevoerd. Het concept van een Erfgoedmakelaar is uniek in de erfgoedwereld: het loopt vooruit op de ondernemende erfgoeddirecteur die instellingsoverstijgend denkt. Het idee heeft inmiddels dan ook ingang gevonden bij de Topsector Creatieve Industrie en is opgenomen in het werkplan van het innovatienetwerk CLICK Cultureel Erfgoed. De verwachting is dat de makelaar een belangrijke bijdrage kan leveren in het beschikbaar houden van de resultaten waarop nationaal en internationaal talloze spin-offs gebaseerd worden.

Dit stuk is gebaseerd op het rapport van Kennisland en DEN "Businessplan op hoofdlijnen. De Erfgoed Makelaar".

Projectteam Businessplannen

- Patricia Alkhoven (CATCHPlus Projectbureau)
- Thomas van Andel (Kennisland)
- Martijn Arnoldus (Kennisland)
- Hennie Brugman (CATCHPlus Projectbureau)
- Marco Streefkerk (DEN)
- In nauwe samenwerking met de CATCHPlus Stuurgroep*

Interesse?

Wilt u meer weten over de erfgoedmakelaar?
Surf dan naar www.catchplus.nl



Organisatie

Projectbureau

Patricia Alkhoven (*vanaf nov 2010*)
Hennie Brugman
Erica Renckens
Marius Snyders (*tot nov 2010*)
Julia Vytopil (*tot nov 2010*)

Hennie Brugman



Stuurgroep

Hans Bennis (*Meertens Instituut*)
Paul Doorenbosch, voorzitter (*Koninklijke Bibliotheek*)
Marco de Niet (*Stichting DEN*)
Eric Postma (*Universiteit van Tilburg*)
Rosemarie van der Veen-Oei (*NWO*)
Hans Westerhof (*Instituut voor Beeld en Geluid*)

Raad van Toezicht

Hans Bennis, voorzitter (*Meertens Instituut*)
Martin Berendse (*Nationaal Archief*)
Daan Broeder (*Max Planck Instituut*)
Hans Buurman (*Gemeentemuseum Den Haag*)
René Dekker (*Naturalis*)
Taco Dibbets (*Rijksmuseum*)
Paul Doorenbosch (*Koninklijke Bibliotheek*)
Jaap van den Herik (*CATCH bestuur*)
Dirk Houtgraaf (*Rijksdienst voor Cultureel Erfgoed*)
Bas Savenije (*Koninklijke Bibliotheek*)
Guus Schreiber (*Vrije Universiteit*)
Paul Spies (*Amsterdam Museum*)
Jantje Steenhuis (*Gemeentearchief Rotterdam*)
Louis Vertegaal (*NWO*)
Hans Westerhof (*Instituut voor Beeld en Geluid*)

Begeleidingscommissie

Marc van der Berg (*Universiteit van Tilburg*)
Frank Bosmans (*Tresoar*)
Renze Brandsma (*Universiteitsbibliotheek UvA*)
Nikola Eltink (*Van Goghmuseum*)
Ellen Fleurbaay (*Stadsarchief Amsterdam*)
Paul Gompes (*Muziek Centrum Nederland*)
Jaap van den Herik (*CATCH bestuur*)
Pim Luiten (*Theater Instituut Nederland*)
Marco de Niet, voorzitter (*Stichting DEN*)
Marijke Oosterbroek (*Amsterdam Museum*)
Boudewijn Ridder (*Nederlands Fotomuseum*)
Eddy Tulp (*Stadsarchief en Athenaeum Bibliotheek Deventer*)
Henk Vanstappen (*PACKED Brussel*)
Rosemarie van der Veen-Oei (*NWO*)
Obby Veenstra (*Fries Flim Archief*)
Reinier van 't Zelfde (*Rijksbureau voor Kunsthistorische Documentatie*)

Betrokken bedrijven

Pieter van Beek (*SARA*)
Bert Degenhart Drenth (*ADLIB*)
Gert-Jan van Dijk (*Target Holding*)
Bas Goedhart (*RDE*)
Erik Groeneveld (*Seecr*)
Thijs Janssen (*Seecr*)
Hanno Lans (*Gemeentemuseum, extern*)

Marc Lindeman (*Picturae*)
Hans Nederbragt (*Trezorix*)
Roeland Ordelman (*X-MI*)
Paul Stork (*Fabrique*)
Mart Trautwein (*Gridline*)
Kars Veling (*Q42*)
Tom Visser (*SARA*)



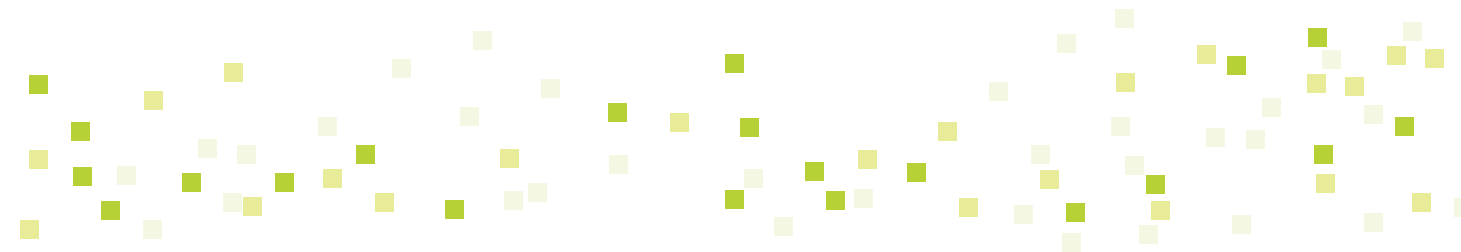
Begeleidingscommissie





Meer informatie

CATCHPlus	www.catchplus.nl
CATCH	www.nwo.nl/catch
SCRATCH4All	www.catchplus.nl/scratch4all
Workspaces	www.catchplus.nl/workspaces
CHoralPlus	www.catchplus.nl/choralplus
WitchcraftPlus	www.catchplus.nl/witchcraftplus
MuSeUMPlus	www.catchplus.nl/museumplus
ZieOok en UPR	www.catchplus.nl/upr
Multiply	www.catchplus.nl/multiply
STITCHPlus	www.catchplus.nl/stitchplus
OpenSKOS	www.catchplus.nl/openskos
DocumentChecker	www.catchplus.nl/docchecker
Annotation Repository	www.catchplus.nl/annotation
Persistent Identifiers	www.catchplus.nl/pids



Colofon

Redactie: Erica Renckens, *Tatataal*
Met dank aan alle betrokkenen bij CATCHPlus

Vormgeving: Jeannette van Bommel
Fotograaf: Bart van Vliet
Beeld: Shutterstock
Drukwerk: Zalsman, Zwolle

November 2012



Nederlandse Organisatie voor Wetenschappelijk Onderzoek



Ministerie van Onderwijs, Cultuur en
Wetenschap



Ministerie van Economische Zaken,
Landbouw en Innovatie

Cultureel erfgoed vormt ons collectieve geheugen. Het is overal om ons heen. CATCH doet onderzoek en ontwikkelt wetenschappelijke methoden en technieken op het snijvlak van geesteswetenschappen en informatica. Met de resultaten worden erfgoedcollecties beter toegankelijk voor publiek, professionals en wetenschappers. Binnen CATCHPlus zijn de veelbelovende demo's van CATCH doorontwikkeld tot volwaardige producten die breed inzetbaar zijn in de Nederlandse erfgoedwereld.