

CATCH Plus

Projectplan voor PRIMA-aanvraag

Continuous Access To Cultural Heritage Plus

Korte Titel	CATCH Plus		
Aanvangsdatum	December 2008	Einddatum	December 2011
Aanvrager	Ministerie van OCW		
Contactinformatie	Robert van der Zwan (OCW/MLB) – tel: 070- 4124259		
Partnerorganisatie(s)	Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO)		
Deelnemende erfgoedinstellingen	Gemeentearchief Rotterdam Gemeentemuseum Den Haag Instituut Beeld en Geluid Koninklijke Bibliotheek Meertens Instituut Nationaal Archief Naturalis Rijksmuseum Amsterdam Rijksdienst voor Archeologie, Cultuurlandschap en Monumenten (RACM)		
Datum	06 november 2008		
Versie	PRIMA-projectvoorstel CATCH Plus-v9.1 20081106.doc		
Auteur(s)	Hennie Brugman (Max Planck Institute for Psycholinguistics) Paul Doorenbosch (Koninklijke Bibliotheek) Mark Kas (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) Rosemarie van der Veen-Oei (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) Met commentaar en bijdragen van: Hans Bennis (Meertens Instituut) Jaap van den Herik (Tilburg Centre for Creative Computing) Vincent de Keijzer (Gemeentemuseum Den Haag) Marco de Niet (Digitaal Erfgoed Nederland) Johan Oomen (Instituut voor Beeld en Geluid)		

Inhoud

1	Samenvatting.....	3
2	CATCH Plus	7
	2.3 Werkpakketten.....	10
	2.3.1 Werkpakket 1 Centrale diensten.....	10
	2.3.2 Werkpakket 2 Toolontwikkeling.....	11
	2.3.3 Werkpakket 3 Framework en interoperabiliteit	13
	2.3.4 Werkpakket 4 Onderhoud en duurzaamheid.....	15
	2.3.5 Werkpakket 5 Projectmanagement	16
	2.4 Organisatie van CATCH Plus	17
	2.4.1 De organisatiestructuur.....	17
	2.4.2 Fasering	19
	2.5 Technische architectuur	19
	2.6 Positionering van Catch Plus in Nederland en daarbuiten.....	22
3	Relatie tot ICT-agenda	23
4	Intellectueel eigendom	24
	Bijlage 1 Belanghebbendenanalyse.....	26
	Bijlage 2 Risicoanalyse.....	28
	Bijlage 3 Overzicht begroting en liquiditeitsbehoefte.....	29
	Bijlage 4 Evaluatieplan	32
	Bijlage 5 Overzicht CATCH Plus-consortium partners	33
	Bijlage 6 Uitwerking centrale diensten (Werkpakket 1).....	34

1 Samenvatting

Erfgoedinstellingen en wetenschappelijke informaticaonderzoeksgroepen hebben de afgelopen jaren in het onderzoeksprogramma Continuous Access to Cultural Heritage (hierna: CATCH) van NWO samengewerkt om technieken en methoden te ontwikkelen die het zoeken en navigeren in cultureel-erfgoeddata optimaliseren. De resultaten zijn opzienbarend en worden nationaal en internationaal erkend. Het is nu tijd om te oogsten, om deze onderzoeksresultaten in de praktijk in te zetten. CATCH Plus zet de demonstrators en pilots¹ die in CATCH ontwikkeld zijn en de kennis die is opgedaan, om in software (diensten en tools). Deze kunnen worden ingezet in het primaire werkproces van de erfgoedinstellingen en in de interactie met het publiek. Kortom: CATCH Plus maakt de slag van wetenschappelijke resultaten naar betrouwbare productieverversies.

Deliverables van CATCH Plus

De demonstrators en pilots uit CATCH-projecten (2005-2010) kunnen, mits doorontwikkeld, een bijzonder nuttige en innovatieve rol vervullen in het primaire proces van de deelnemende erfgoedinstellingen en in de beschikbaarstelling van het cultureel erfgoed aan het publiek. De architectuur daarvoor is een servicegeoriënteerde architectuur (SOA) met gestandaardiseerde interfaces, waardoor de services met minimale voorkennis gebruikt kunnen worden en in combinatie met andere applicaties ingezet kunnen worden.

In CATCH Plus worden twee bedrijfsklare centrale diensten² gebouwd op basis van de demonstrators waarin gemeenschappelijke aspecten van de CATCH-projecten bij elkaar zijn gebracht. Deze diensten leveren op hun beurt output die gebruikt wordt in de generiek inzetbare tools, die op hun beurt de uitbouw zijn van de in CATCH ontwikkelde pilots. Om de diensten en tools eenvoudig en stabiel in onderling verband te kunnen inzetten, wordt een framework ontwikkeld en worden voorwaarden gecreëerd die de interoperabiliteit tussen diverse erfgoedcollecties ondersteunen. Ook worden de voorwaarden en de kaders gecreëerd voor de duurzaamheid van de CATCH Plus-resultaten.

De software van CATCH Plus wordt openbaar gemaakt via een open-sourceorganisatie (SourceForge, OSOSS). Voor de erfgoedsector en geïnteresseerde eindgebruikers komt er een website die toegang geeft tot de resultaten van CATCH Plus.

Werkpakket 1: centrale diensten

CATCH Plus levert twee centrale softwarediensten op, die zijn gebaseerd op de CATCH-demonstrators waarin de gemeenschappelijke aspecten van de CATCH-onderzoeksprojecten bij elkaar zijn gebracht. De output van deze centrale diensten wordt gebruikt in de tools die in werkpakket 2 worden ontwikkeld. Deze twee diensten liggen aan de basis van veel erfgoedactiviteiten: het ontsluiten van informatie in onderlinge samenhang en het toevoegen van kennis aan objecten voor beter begrip en gebruik daarvan.

1. Een repository of vocabularies, waarin thesauri als de GTAA³, AAT⁴, ICONCLASS⁵, GOO⁶ en andere vocabulaires (gecontroleerde termen- en conceptenverzamelingen) op uniforme wijze en onderling gerelateerd beschikbaar

¹ Onder 'pilot' wordt verstaan de softwareapplicaties die in de projecten van CATCH zijn ontwikkeld om de mogelijkheden en de resultaten van het onderzoek in dat project tastbaar en zichtbaar te maken. De 'demonstrators' in CATCH zijn applicaties waarin de gemeenschappelijke kennis van alle CATCH-projecten is samengebracht om te laten zien wat de integratie van die kennis kan opleveren.

² Het begrip 'dienst' wordt gebruikt als er niet uitsluitend sprake is van softwareapplicaties, maar ook van menselijke organisatie. Uiteraard gaat het in CATCH Plus om geautomatiseerde diensten, die slechts een minimum aan menselijk ingrijpen vragen.

³ GTAA (Gemeenschappelijke Thesaurus Audiovisuele Archieven) is een ontsluitingssysteem in gebruik bij Beeld en Geluid en het Nederlands Filmmuseum

worden gesteld. Hierbij worden hulpmiddelen ontwikkeld om bestaande vocabulaires om te zetten naar standaardvormen, en om begrippen in de verschillende vocabulaires aan elkaar te koppelen.

2. Een annotation- & recommendation-dienst, die het mogelijk maakt (semi-)geautomatiseerd inhoudelijke metadata toe te kennen aan digitale objecten (audio, video, teksten, beeldmateriaal, muziek, et cetera) en aan segmenten van dergelijke objecten (woorden, elementen binnen beelden, scènes). Hierbij wordt onder meer gebruik gemaakt van de vocabulaires uit de repository of vocabularies.

Werkpakket 2: toolontwikkeling

Bovengenoemde centrale diensten ondersteunen de specifieke tools, die voortbouwen op de pilots van het CATCH-onderzoeksprogramma. Deze tools worden ontwikkeld als zelfstandige webservices. Voorbeelden zijn:

- Tool voor de automatische verrijking en correctie van gegevens met behulp van machinelearning-technieken en van de geautomatiseerde annotation-dienst;
- Tool voor de ontwikkeling van gelaagde annotaties (annotaties op basis van bron A koppelen aan annotaties van bron B);
- Rankingtool voor het rangschikken van concepten of documenten op grond van clustering in een vocabulaire.

Werkpakket 3: framework en interoperabiliteit

Om de softwarediensten en tools in onderling verband te kunnen gebruiken wordt een geformaliseerde workflow-managementmodule ontwikkeld (framework). Voorwaarde voor geïntegreerd gebruik is dat de content (erfgoedcollecties) interoperabel is en de verwijzingen naar objecten persistent zijn. Hiervoor kent CATCH Plus de volgende deliverables:

1. Ontwikkeling van een eenvoudige, gestandaardiseerde en expliciete workflow-managementmodule om diensten, tools en applicaties te verbinden met de servers die het eigenlijke werk doen of data leveren. Het wordt zo voor erfgoedinstellingen eenvoudiger om zelf bouwstenen bij te dragen, en om zelf applicaties samen te stellen op grond van al bestaande bouwstenen en services;
2. Activering van OAI-PMH en DC of vergelijkbare standaards, waar die in huidige lokale systemen voorhanden zijn (Adlib, TMS). Ontwikkeling van application programming interfaces (API) voor overige partijen om de metadata van de deelnemende partijen via een standaardformaat en met standaardprotocollen te kunnen doorzoeken, harvesten en presenteren;
3. Invoering van een resolverstructuur voor persistente identifiers bij de deelnemende instellingen, waarbij andere erfgoedinstellingen eenvoudig kunnen aansluiten. Nodig voor de Digitale Collectie Nederland en de samenwerking van genoemde diensten en tools. Tevens sluit dit aan bij een expliciete wens van de Directie Cultureel Erfgoed van OCW.

⁴ De AAT (Art and Architecture thesaurus) is een breed opgezette internationale thesaurus, eigendom van Getty Research Institute, en onder licentie in een Nederlandse vertaling in Nederland en Vlaanderen in gebruik. De AAT richt zich vooral op kunstgeschiedenis en architectuur, maar wordt ook op andere gebieden toegepast. Hij bestaat in vele talen en is daardoor een instrument dat uitstekend meertaligheid kan ondersteunen.

⁵ ICONCLASS is an international subject-specific classification system; it is a hierarchically ordered collection of definitions of objects, persons, events and abstract ideas that can be the subject of an image. Art historians, researchers and curators use it to describe, classify and examine the subject of images represented in various media such as paintings, drawings and photographs.

⁶ GOO (Gemeenschappelijke Onderwerpsontsluiting) bestaat uit een numerieke classificatie (NBC) en een trefwoordthesaurus (GTT). In principe omvat de GOO alle wetenschapsgebieden, maar het zwaartepunt ligt bij de humaniora en de rechtswetenschappen/sociale wetenschappen. Het wordt door een groot aantal Nederlandse universiteitsbibliotheken onderhouden en gehanteerd in het centrale catalogussysteem van OCLC-Europa.

Werkpakket 4: duurzaamheid en onderhoud

Om de ontwikkelde software ook voor de toekomst veilig te stellen en verder te ontwikkelen, wordt aangehaakt bij de landelijke infrastructuur die in opbouw is in het cultureel erfgoed. CATCH Plus levert een bijdrage aan de verdere ontwikkeling en implementatie van deze infrastructuur. Dit gebeurt door de volgende activiteiten:

1. Het ontwikkelen van één of meerdere exemplarische businessplannen voor het duurzame onderhoud van de software. Hierbij wordt gekeken naar bestaande modellen voor publiek-private samenwerking. Het bedrijfsleven heeft in een oriëntatieronde positief gereageerd op eventuele participatie;
2. Het identificeren van één of meerdere kenniscentra binnen het erfgoed, die het eigenaarschap en functioneel beheer van gemeenschappelijk ontwikkelde tools op zich kan/kunnen nemen. Hiervoor wordt aangesloten bij bestaande zwaartepunten in expertise.⁷

Samenhang binnen de CATCH Plus-architectuur

- De twee centrale diensten maken gebruik van elkaar, en werken samen met één of meerdere tools. Ze kunnen ook worden gevoed door de resultaten uit de tools;
- Zowel de centrale diensten als de specifieke tools werken op basis van metadata, digitale objecten of informatie uit de centrale diensten;
- De metadata en digitale objecten blijven op de instellingen. Door het gebruik van persistente identifiers blijven de relaties op lange termijn bestaan;
- De tools hebben als resultaat ook verrijking van de bestaande metadata op de instellingen.

Wat wil CATCH Plus daarnaast bereiken

- De expertise die in CATCH is opgebouwd, wordt met het bedrijfsleven gedeeld. Hierdoor kan die kennis geïntegreerd worden in commerciële producten. Zeven bedrijven⁸ hebben al expliciet belangstelling getoond. Door gezamenlijk de softwareontwikkeling ter hand te nemen worden standaardisatie en interoperabiliteit verder verbeterd. CATCH Plus levert een substantiële bijdrage aan het verder oplossen van de zogenoemde innovatieparadox - in Nederland wordt bovengemiddeld gepresteerd op het gebied van fundamenteel onderzoek, maar in sommige sectoren minder als het gaat om de doorontwikkeling van onderzoek naar praktische toepassingen;
- De samenwerking binnen de erfgoedsector wordt naar een hoger plan gebracht en de interoperabiliteit van collecties wordt vergroot. Dit zijn belangrijke stappen op weg naar een integrale Digitale Collectie Nederland. Tevens geeft het Nederland een goede positie binnen het Europeana-netwerk;
- De toegankelijkheid van het cultureel erfgoed wordt sterk verbeterd. Toegankelijkheid van het erfgoed versterkt de kenniseconomie. Dat levert zowel een belangrijke culturele en maatschappelijke meerwaarde als een economische. De creatieve industrie⁹ kan in haar productontwikkeling moeiteloos gebruik maken van het culturele werkkapitaal. Gezien de omvang van de educatieve sector en de wens om die op digital leest te schoeien, ligt daar een omvangrijk toepassingsgebied en dus economisch potentieel;

⁷ Digitaal Erfgoed Nederland (DEN), de sectoroverschrijdende organisatie voor standaarden, kwaliteitsbewaking, en kennisuitwisseling, heeft interesse getoond om deze rol voor de centrale diensten op zich te nemen. Deze centrale diensten worden gezien als mogelijk complementair aan de huidige dienstverlening van DEN.

⁸ Het gaat om: CIT, Deventit, Furore, Q42, Seek You Too, Trezorix en Zylab, waarmee in de voorbereidingsfase is gesproken.

⁹ Beschrijving van de 'creative class' in de research agenda van het Innovatieplatform IIP-CREATE: http://83.98.156.43/~iipcreate/wp-content/uploads/2008/05/q_a_sra.pdf

- Het open en innovatieve karakter van CATCH Plus uit zich in de ontwikkeling van open-sourcesoftware, gemeenschappelijk gebruikte tools en interoperabele data. Zowel de erfgoedsector, de wetenschappelijke informatica als commerciële en niet-commerciële partijen krijgen kansen om een blijvende, synergetische relatie met elkaar aan te gaan en deel te worden van een nieuw gezamenlijk netwerk.

Duur en budget van CATCH Plus

Het CATCH Plus-project loopt van december 2008 tot december 2011. De totale kostenraming voor het implementatieproject is € 3.293.326. De bijdrage van de consortiumpartners (erfgoedinstellingen, kennisinstellingen en bedrijfsleven) en van NWO en OCW is € 1.646.663. De eigen bijdrage is 50% van de totale kostenraming. De gevraagde financiering van PRIMA bedraagt € 1.646.663. De doorlooptijd van CATCH Plus is 3 jaar.

2 CATCH Plus

Erfgoed is een essentieel onderdeel van de cultuur van een gemeenschap. Het gaat om schilderijen, teksten, liederen, foto's, films, maar ook om gebouwen en landschappen, rituelen en tradities. Samen weerspiegelen deze objecten onze culturele identiteit. Ze zijn de bron van onze kennis over heden en verleden, ze zijn onderdeel van het maatschappelijke discours, ze vormen het basismateriaal voor onderzoek naar historie en maatschappij en ze bieden vermaak en ontspanning. Door het cultureel erfgoed digitaal te ontsluiten komt onze culturele rijkdom binnen het bereik van iedereen, want het leven speelt zich in toenemende mate online af. De interesse van mensen in cultureel erfgoed kan zo worden aangewakkerd én versterkt. Educatie in ruime zin speelt daar een niet te verwaarlozen rol bij. De overheid maakt zich al jaren sterk voor de opbouw van de Digitale Collectie Nederland en investeert er structureel in.

Simpelweg digitaal maken is echter niet voldoende. Er moet in het materiaal gezocht en genavigeerd kunnen worden en er moeten toepassingen mee gemaakt kunnen worden. Het gaat om vele miljoenen voorwerpen en ontelbare stukjes informatie. Erfgoedinstellingen en wetenschappelijke informaticaonderzoeksgroepen hebben de afgelopen jaren in het onderzoeksprogramma CATCH van NWO samengewerkt om technieken en methoden te ontwikkelen die het zoeken en navigeren optimaliseren. De resultaten zijn opzienbarend en worden nationaal en internationaal erkend. Het is nu tijd om te oogsten, om deze onderzoeksresultaten in de praktijk in te zetten. CATCH Plus zet de demonstrators en pilots¹⁰ die in CATCH ontwikkeld zijn en de kennis die is opgedaan, om in software en tools. Deze kunnen worden ingezet in het primaire werkproces van de erfgoedinstellingen en in de interactie met het publiek. Kortom: CATCH Plus maakt de slag van wetenschappelijke resultaten naar betrouwbare productieverversies.

Kern van CATCH Plus is het valoriseren van en voortbouwen op de resultaten die verkregen zijn binnen het informatica-onderzoeksprogramma CATCH. Het onderzoeksprogramma CATCH is opgezet door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) en financiert tien projecten die worden uitgevoerd binnen de muren van negen cultureel erfgoedinstellingen. De tien projecten onderzoeken nieuwe technieken en methoden om digitale collecties van de erfgoedinstellingen op een adequatere en snellere manier te ontsluiten en te analyseren met inzet van state-of-the-art informatietechnologie. Tijdens het onderzoek ontwikkelen de CATCH-projecten demonstrators en pilots. De demonstrators dienen om de gemeenschappelijke ideeën en concepten die tijdens het onderzoek ontstaan zijn, te demonstreren en de waarde daarvan aan te tonen voor de praktijk van de erfgoedinstellingen. De pilots zijn een eerste stap op weg naar de realisatie van bruikbare specifieke softwaretoepassingen. Gebrek aan financiële middelen vormt vaak een hindernis voor de culturele erfgoedwereld om bruikbare softwaretoepassingen te (laten) ontwikkelen.

Het bedrijfsleven verkeert in een vergelijkbare situatie. In de erfgoedsector werken kleine en middelgrote bedrijven die veelal onvoldoende mogelijkheden hebben om met eigen middelen te investeren in onderzoek en toepassing van geavanceerde innovatieve oplossingen. Werken vanuit het open-sourceprincipe, waarbij winst kan worden behaald uit een bundeling van krachten en kennis, komt slechts aarzelend tot stand. In CATCH Plus krijgen de ondernemers een stimulans om open source met een beperkt bedrijfsrisico verder te exploreren.

De doelgroep voor de software die in CATCH Plus wordt ontwikkeld is de erfgoedexpert in de instelling. Slechts in enkele gevallen worden directe eindgebruikersapplicaties ontwikkeld. De erfgoedexpert is echter de intermediair tussen de erfgoedkennis en de gebruiker. Door de ontwikkeling van de CATCH Plus-software kan de erfgoedexpert ervoor

¹⁰ Onder 'pilot' wordt verstaan de softwareapplicaties die in de projecten van CATCH zijn ontwikkeld om de mogelijkheden en de resultaten van het onderzoek in dat project tastbaar en zichtbaar te maken. De 'demonstrators' in CATCH zijn applicaties waarin de gemeenschappelijke kennis van alle CATCH-projecten is samengebracht om te laten zien wat de integratie van die kennis kan opleveren.

zorgdragen dat objecten en kennis effectiever hun rol kunnen spelen in de interactie tussen gebruiker en erfgoed in de digitale omgeving.

De verdere ontwikkeling van de kennis en software uit het CATCH-onderzoeksprogramma is het uitgangspunt van het implementatieproject CATCH Plus. Daarnaast wordt in CATCH Plus de kennisrelatie met de IT-onderzoeksinstellingen bestendig door onderzoekers en programmeurs voor langere duur te betrekken bij de sector. De innovatieve kracht van de erfgoedsector krijgt hiermee een blijvende impuls. Tevens wordt hierdoor ruimte gecreëerd voor de wetenschappelijke onderzoekers om over te stappen naar de erfgoedsector en daar hun kennis te valoriseren.

2.1 Achtergrond van CATCH Plus: het onderzoeksprogramma CATCH

CATCH Plus beoogt de resultaten van het NWO-onderzoeksprogramma CATCH in de erfgoedwereld te implementeren.

CATCH is een informatica-onderzoeksprogramma, ontstaan vanuit de behoefte van de cultureel-erfgoedinstellingen.

Behalve aan financiële middelen, ontbreekt het deze instellingen vaak ook aan technische kennis om digitale collecties adequaat te ontsluiten, om efficiënte methoden te ontwikkelen om data te verrijken, of om informatie aan te passen aan de behoeften van de gebruikers via personalisatiemethoden. Voor de informaticaonderzoekers is het erfgoeddomein een uitdagend onderzoeksterrein door de enorme hoeveelheden heterogene data die zonder uniforme standaarden zijn geannoteerd en zijn opgeslagen in verschillende systemen.

Bij het inrichten van het programma zijn vragen vanuit het erfgoed centraal gesteld: hoe kunnen we de verschillende collecties aan elkaar verbinden, is het mogelijk om nieuwe informatie op een (semi-)automatische wijze toe te voegen aan bestaande data en op welke manier kan de informatie op maat aan de gebruiker worden aangeboden? Die vragen hebben geleid tot de volgende drie centrale onderzoeksthema's:

- 1 Semantische interoperabiliteit via metadata;
- 2 Kennisverrijking met behulp van automatische analyse;
- 3 Personalisatie in presentatie.

De tien projecten die CATCH sinds 2005 financiert, werken volgens de formule *laboratorium extra muros*: de CATCH-onderzoekers hebben hun werkplek voor een deel van hun tijd in de erfgoedinstellingen, die functioneren als hun laboratorium buiten de muren van de universiteit. Er is directe interactie tussen de onderzoekers en erfgoedprofessionals. De wetenschappelijke kwaliteit van het onderzoek staat voorop, en tegelijkertijd wordt de toepasbaarheid van de onderzoeksresultaten bewaakt. De aanwezigheid van informaticaonderzoekers, met hun andere methoden en technieken en hun wetenschappelijke benadering van erfgoedkwesties, leidt ook buiten de directe projectcontacten tot herbezinning op het primaire proces in erfgoedinstellingen.

2.2 Resultaten van CATCH Plus

De onderzoeksresultaten uit het CATCH-programma van de afgelopen jaren zijn opzienbarend en worden nationaal en internationaal als zodanig erkend. Het is tijd om te oogsten, om te werken aan de valorisatie van de onderzoeksprojecten en de implementatie van deze innovatieve technieken en resultaten. Het implementatietraject CATCH Plus zal de prototypen (demonstrators en pilots) en de kennis uit CATCH omzetten in diensten en tools die in het primaire proces van erfgoedinstellingen en in de interactie met het publiek worden ingezet. De essentie van CATCH Plus is het vertalen van wetenschappelijke resultaten naar betrouwbare productievergies.

Primaire projectresultaten

1. Twee centrale diensten - een repository of vocabularies en een annotation- & recommendation-dienst zoals omschreven in Werkpakket 1;
2. Minimaal vier en maximaal acht softwaretools, zoals omschreven in Werkpakket 2¹¹, geënt op de praktijksituatie van een deelnemende erfgoedinstelling. De tools moeten bij minimaal één collectie van een andere instelling succesvol getest zijn;
3. Implementatie van een persistente identifier-resolverstructuur zoals omschreven in Werkpakket 3, om te beginnen binnen die collecties die ook worden gebruikt bij de softwareontwikkeling;
4. Businessplan voor minimaal twee van de ontwikkelde CATCH Plus-tools als basis voor een publiek-private samenwerking (Werkpakket 4);
5. Onderbrenging, met in achtname van het auteursrecht, van de software bij een *Archive* voor open-sourcesoftware, in overeenstemming met Werkpakket 4;
6. Het beleggen van het eigenaarschap van alle software-onderdelen, bij voorkeur in de vorm van kenniscentra, met een verplichting tot onderhoud en beheer na afloop van CATCH Plus gedurende minimaal 3 jaar, in overeenstemming met Werkpakket 4.

Secundaire projectresultaten

1. **Ontsluiting** van data uit gedistribueerde heterogene collecties (videobestanden, foto's, tekst, muziek, etc.) binnen de deelnemende projectpartners, en de beschikbaarstelling van de verkregen kennis over gedistribueerd ontsluiten. De ontwikkelde methoden en technieken dragen bij aan een gezamenlijke aanpak om cultureel erfgoed integraal aan te bieden op het web;
2. Verspreiding en **verankering** van binnen CATCH opgebouwde kennis en expertise van de onderzoekers binnen de muren van erfgoedinstellingen, mede door het scheppen van arbeidsmogelijkheden in het cultureel erfgoed voor informatici en gespecialiseerde programmeurs. Met hun erfgoedkennis en ICT/informatica-expertise helpen de onderzoekers de ICT-bedrijvigheid en –toepassingen binnen het cultureel erfgoed te versterken. Tijdens dit proces wordt de wetenschappelijke kennis gedeeld met het bedrijfsleven;
3. Optimalisatie van de **toegankelijkheid** van het nationale cultureel erfgoed. Uiteindelijk gaat het om de gebruiker in dit implementatietraject. De erfgoedinstellingen krijgen door CATCH Plus de kans om de applicaties volledig aan te passen aan de behoeften en vragen die er leven. Het is de eindgebruiker van digitaal cultureel erfgoed die via internet profijt heeft van het materiaal dat beschikbaar komt door de CATCH Plus-tools en -methoden.

De ambities van CATCH Plus

- De expertise die in het onderzoeksprogramma is opgebouwd, wordt met het bedrijfsleven gedeeld. Hierdoor kan die kennis geïntegreerd worden in commerciële producten. Door gezamenlijk de softwareontwikkeling ter hand te nemen worden standaardisatie en interoperabiliteit verder verbeterd. CATCH Plus levert een substantiële bijdrage aan een verdere oplossing van de zogenoemde innovatieparadox: in Nederland wordt bovengemiddeld gepresteerd op het gebied van fundamenteel onderzoek, maar in sommige sectoren minder als het gaat om de doorontwikkeling van onderzoek naar praktische toepassingen.
- De samenwerking binnen de erfgoedsector wordt naar een hoger plan gebracht en de interoperabiliteit van collecties wordt vergroot. Dit zijn belangrijke stappen op weg naar een integrale Digitale Erfgoedcollectie

¹¹ Hoewel uiterste inspanning wordt geleverd om maximale resultaten te halen, moet er bij softwareontwikkeling en -implementatie altijd rekening gehouden worden met een zekere faalkans.

Nederland. Tevens geeft het Nederland een goede positie binnen het Europeana-netwerk, waardoor collecties in Europa beter zichtbaar worden en er meer mogelijkheden ontstaan aan Europese projecten mee te doen;

- De toegankelijkheid van het cultureel erfgoed wordt sterk verbeterd. Toegankelijkheid van het erfgoed versterkt de kenniseconomie. Dat levert zowel een belangrijke culturele en maatschappelijke meerwaarde als ook een economische. De creatieve industrie kan in haar productontwikkeling moeiteloos gebruik maken van het culturele werkkapitaal. Gezien de omvang van de educatieve sector en de wens om die op digitale leest te schoeien, ligt daar een omvangrijk toepassingsgebied en dus economisch potentieel;
- De open innovatie van CATCH Plus uit zich in de ontwikkeling van open-sourcesoftware, gemeenschappelijk gebruikte applicaties en interoperabele data. Zowel de erfgoedsector, de wetenschappelijke informatica als profit- en non-profitpartijen krijgen kansen om een blijvende synergetische relatie met elkaar aan te gaan en deel te gaan uitmaken van een nieuw gezamenlijk netwerk.

2.3 Werkpakketten

CATCH Plus kent vijf werkpakketten die door hun samenhangende activiteiten elkaar versterken. Ieder werkpakket levert concrete eindproducten. Alle partijen binnen CATCH Plus krijgen de beschikking over de eindresultaten. De samenhangende activiteiten dragen bij aan een gezamenlijk doel: de realisatie van bruikbare softwareapplicaties, inzetbaar voor het gezamenlijke cultureel erfgoed in Nederland en daarbuiten. Bij de bouw van software zal het bedrijfsleven uitdrukkelijk een rol spelen. Dit kan in de vorm van uitbesteding van werk of in een gezamenlijk partnership. Bij de gedetailleerde uitwerking van CATCH Plus worden uitdrukkelijk de mogelijkheden, kennis en resultaten van het MultimediaN-eCulture project (BSIK) betrokken. Hiervoor is toestemming verkregen. eCulture loopt nog tot in 2009, en heeft de afgelopen jaren al interessant onderzoek gedaan en prototypen ontworpen op het gebied van de interoperabiliteit van en het semantisch zoeken in erfgoedcollecties.

De inrichting in werkpakketten biedt de mogelijkheid tot een realistische (tussentijdse) controle op voortgang en resultaten en maken een heldere terugkoppeling mogelijk over de besteding van het subsidiegeld dat beschikbaar is voor CATCH Plus.

CATCH Plus omvat de volgende werkpakketten:

- Werkpakket 1: Centrale diensten
- Werkpakket 2: Toolontwikkeling
- Werkpakket 3: Framework en interoperabiliteit
- Werkpakket 4: Onderhoud en duurzaamheid
- Werkpakket 5: Projectmanagement

Bij de verschillende werkpakketten wordt hierna de aanpak globaal aangegeven. Essentieel daarbij is dat afstemming tussen de verschillende onderdelen en brede inzetbaarheid van de resultaten gegarandeerd zijn.

2.3.1 Werkpakket 1 Centrale diensten

CATCH Plus levert twee centrale softwarediensten op, gebaseerd op de demonstrators waarin de gemeenschappelijke aspecten van de onderzoeksprojecten bij elkaar zijn gebracht. De output van deze centrale diensten wordt gebruikt in de ontwikkelde tools uit werkpakket 2. Deze twee diensten liggen aan de basis van veel erfgoedactiviteiten; het ontsluiten van informatie in onderlinge samenhang en het toevoegen van kennis aan objecten voor beter begrip en gebruik daarvan.

Dit werkpakket bestaat uit twee diensten (deelpakketten):

1. Repository of vocabularies

In de repository of vocabularies worden thesauri en andere typen vocabulaires (gecontroleerde termen- en conceptenverzamelingen) op uniforme wijze en onderling gerelateerd beschikbaar gesteld. De auteursrechten zijn, voor zover mogelijk, door de erfgoedinstellingen vrijgegeven voor gebruik en voor projectdoeleinden. Bij de repository worden vrij beschikbare tools gebouwd om thesauri simpel om te zetten naar gestandaardiseerde semantische structuren. Bekend is dat niet alle wenselijke vocabulaires gratis en vrij te gebruiken zijn. Het projectbureau zal met de erfgoedinstellingen samen de licentieproblematiek voor deze gevallen onderzoeken en binnen de mogelijkheden aanpakken;

2. Annotation- & recommendation-dienst

De annotation- & recommendation-dienst maakt het mogelijk (semi-)geautomatiseerd inhoudelijke metadata toe te kennen aan digitale objecten (audio, video, teksten, beeldmateriaal, muziek, et cetera) en aan segmenten van dergelijke objecten (woorden, elementen binnen beelden, scènes). Hierbij wordt onder meer gebruik gemaakt van de vocabulaires uit de repository.

De centrale diensten zijn nader uitgewerkt in bijlage 6.

NB. In Werkpakket 3 zijn de activiteiten opgenomen om tot een persistente identifier-objectinfrastructuur in de erfgoedwereld te komen. Het is belangrijk om die stappen te nemen om de resultaten voortkomend uit het gebruik van de diensten in Werkpakket 1 en de hieronder genoemde tools, duurzaam beschikbaar te houden.

Globale werkwijze

De aansturing van de bouw van de centrale diensten gebeurt vanuit het projectbureau. Voorafgaand aan de bouw van de centrale diensten wordt door het projectbureau de SOA-architectuur van het gehele project uitgewerkt. De architectuur omvat specificaties, voorschriften en aanbevelingen. De deelnemende instellingen worden nauw betrokken bij het opstellen van de *requirements*. De keuze voor een softwareontwikkelaar geschiedt door open inschrijving. Het projectbureau voert tijdens de bouw directe regie over de inhoud van de software.

Voor de bouw en het gebruik wordt een adequate hardware- en softwareomgeving ingericht (ontwikkel-test/acceptatie-productie).

2.3.2 Werkpakket 2 Toolontwikkeling

Het werkpakket voor toolontwikkeling bestaat uit de volgende acht CATCH Plus tools (deelpakketten), ontwikkeld door de erfgoedinstellingen. Hoewel bij de ontwikkeling van de tools vaak gebruik gemaakt wordt van een specifieke lokale collectie, is een eis dat de tools werken op meerdere gedistribueerde collecties. Alle tools maken zoveel mogelijk gebruik van de centrale diensten. Bovendien worden, waar mogelijk, onderdelen zo ontwikkeld dat ze door anderen ook weer als aparte diensten kunnen worden hergebruikt. Voor een nadere uitwerking van ieder van de deelpakketten, zie bijlage 7.

De softwareontwikkeling vindt plaats bij de erfgoedinstellingen, waar mogelijk in samenwerking met commerciële softwarebedrijven. In de voorbereiding van CATCH Plus zijn verkennende gesprekken gevoerd met zeven¹² bedrijven, waaruit blijkt dat de interesse voor samenwerking bestaat. De deelpakketten bestaan uit de volgende tools:

1. Een tool voor het indexeren van gesproken tekst met behulp van spraaktechnologie. Bij de ontwikkeling van deze tool worden audio-opnamen van raadsvergaderingen in Rotterdam gebruikt. De notulen daarvan dienen om opnames te structureren en de kwaliteit van de index te verhogen. Deze tool maakt gebruik van bovengenoemde

¹² Het betreft de volgende bedrijven: CIT, Deventit, Furore, Q42, Seek You Too, Trezorix en Zylab, waarmee in de voorbereidingsfase is gesproken.

- centrale annotation-diensten. [De ontwikkeling vindt plaats in samenwerking met het Gemeentearchief Rotterdam.]
2. Een webgebaseerde tool voor het geavanceerd, geünificeerd zoeken in heterogene, gedistribueerd opgeslagen erfgoedcollecties. Deze kan als geavanceerde zoekcomponent tevens in andere tools worden hergebruikt. Bij het presenteren gaat de eigen identiteit van de bijdragende erfgoedcollecties zo min mogelijk verloren. [De ontwikkeling vindt plaats in samenwerking met het Gemeentemuseum Den Haag.]
 3. Een database-organizer. Deze tool omvat een module voor het semi-automatisch detecteren en corrigeren van foute waarden in database-velden, voor het invullen van lege velden met behulp van referentiestructuren (vocabulaires), en voor het automatisch koppelen van ongestructureerde teksten aan thesauri (annotatie). [De ontwikkeling vindt plaats in samenwerking met Naturalis en RACM].
 4. Een documentalist supportsysteem dat binnen een geïntegreerde zoek- en browse-omgeving suggesties voor trefwoorden levert aan documentalisten, op basis van automatisch aan thesauri gekoppelde teksten. Het maakt daarbij gebruik van centrale vocabulaire- en annotationdiensten. Diverse modules waaronder het op relevantie rangschikken van trefwoordsuggesties, worden als aparte herbruikbare diensten aangeboden. [De ontwikkeling vindt plaats in samenwerking met Beeld en Geluid]
 5. Een melodieënzoekmachine, bestaande uit een suite van componenten waaruit een robuuste en effectieve webgebaseerde zoekmachine voor melodieën wordt samengesteld. Daarnaast worden voor dataproductie en –management een muziek-editor en een audio-transcriptietool opgeleverd. [De ontwikkeling vindt plaats in samenwerking met het Meertens Instituut]
 6. Een webgebaseerde softwareomgeving voor beheerders van gedigitaliseerde, gestructureerde handgeschreven documenten. De betreffende documentcollecties worden in eerste instantie ge-upload en van relaties voorzien. Voor een deel moeten ze vervolgens met behulp van een annotationtool handmatig worden getranscribeerd. Vervolgens kan de gehele collectie worden doorzocht op grond van handschriftgelijkenis. [De ontwikkeling vindt plaats in samenwerking met het Nationaal Archief]
 7. Een toolset om thesauri en classificatiesystemen semantisch aan elkaar te koppelen, op te slaan, te beheren, te verrijken en te benutten voor geïntegreerde ontsluiting van digitale objecten. De toolset levert een directe bijdrage aan de repository of vocabularies. [De ontwikkeling vindt plaats in samenwerking met de Koninklijke Bibliotheek]
 8. Een gepersonaliseerde eindgebruikeromgeving waarin semantische annotaties worden toegevoegd aan de itembeschrijvingen, die vervolgens op grond van gebruikersprofielen worden aangeboden aan eindgebruikers. De gebruikersprofielen komen voor de gehele erfgoedgemeenschap beschikbaar. Om de semantisch verrijkte collecties te ontsluiten wordt gebruikt gemaakt van de semantische browse- en zoekfaciliteiten die zijn ontwikkeld in eCulture (onderdeel van het BSIK-programma MultimediaN). [De ontwikkeling vindt plaats in samenwerking met het Rijksmuseum]

Globale werkwijze

Het projectbureau stelt een set voorwaarden, voorschriften en aanbevelingen op waaraan de bouw van deze tools moet voldoen. Het gaat om overall architectuur, input-outputformaten, standaarden, voorwaarden voor interoperabiliteit en dergelijke. Bij de bouw van de software committeren de ontwikkelaars zich aan deze voorwaarden en specificaties. Via een open inschrijving kunnen softwareontwikkelaars hun belangstelling kenbaar maken. Uit de gegadigden wordt een partij gekozen met wie een contract wordt aangegaan. De deelprojectleider op de instelling is verantwoordelijk voor de goede uitvoering van het project. De instelling die de verantwoordelijkheid draagt voor de uitvoering van het project kan de

voorkeur uitspreken voor ofwel samenwerking met een externe partij ofwel de aanstelling van een eigen programmeur. Dit gebeurt met instemming van het centrale projectbureau.

Het deelproject start met het opstellen van een gedetailleerd plan voor de werkzaamheden, waarin onder andere de deliverables expliciet worden omschreven. Voor de bouw van de software worden een functioneel en een technisch ontwerp gemaakt. Deze ontwerpen behoeven een check van het programmabureau op voornoemde aanbevelingen en voorwaarden. Bij het ontwerpen van de software kan eventueel een beroep worden gedaan op de technische coördinator.

2.3.3 Werkpakket 3 Framework en interoperabiliteit

Om de diensten en tools in onderling verband te kunnen gebruiken, wordt een geformaliseerde workflow-managementmodule ontwikkeld (framework). Om webservices efficiënte input te geven en om de output van webservices eenduidig te kunnen verwerken, is het noodzakelijk dat de metadata van objecten uniform (gestandaardiseerd) herbruikbaar en uitwisselbaar zijn. Door dit zo dicht mogelijk bij de bron te organiseren wordt voorkomen dat bij conversie de consistentie van de data verdwijnt. Standaardisatie door middel van bestaande en gemeenschappelijk ondersteunde (nationale en internationale) standaarden en protocollen van data en metadata maakt dit mogelijk. Door deze standaardisatie kunnen data ook eenduidig door collega-instellingen worden gecombineerd met (delen van) hun eigen collectie.

Om te kunnen verwijzen naar gedistribueerd opgeslagen annotaties zonder te weten hoe de data in die instelling georganiseerd zijn, zijn standaardisatie en persistente identifiers een *must*. Persistente identifiers zijn identifiers die niet meer veranderen in de loop van de tijd. Bij afwezigheid daarvan raakt de toegevoegde kennis, die via een identifier aan het object of de metadata is gekoppeld, verweesd en dus waardeloos.

Er is een internationaal redelijk geaccepteerde set primaire standaarden en protocollen voor vindbaarheid van en verwijzing naar erfgoedobjecten (in Nederland vastgelegd in het ICT-register van Digitaal Erfgoed Nederland (DEN)¹³, in Europees verband in gebruik bij Europeana¹⁴). Verder zijn er (althans in de bibliotheekwereld) Europese afspraken over het opzetten van een persistente-identificerstructuur. Voor beide onderdelen moet nog een kennisverspreiding- en applicatieslag gemaakt worden om deze standaarden, protocollen en afspraken ook daadwerkelijk in de instellingen geïmplementeerd te krijgen. Zonder dat is er geen gemeenschappelijke infrastructuur die uitwisseling en hergebruik van data mogelijk maakt, en gemeenschappelijk gebruik van diensten, tools en webservices faciliteert. CATCH Plus neemt met de implementatie hiervan het initiatief tot de realisatie van een gemeenschappelijke infrastructuur voor de Digitale Collectie Nederland. Het gebruik van protocollen en standaarden maakt het mogelijk de objecten decentraal te blijven opslaan, waarmee de rechtendiscussie rond gebruik van objecten weliswaar niet wordt opgelost, maar wel wordt vereenvoudigd. In het OAI-metadataharvestprotocol zijn afspraken t.a.v. 'rights' en 'trust' gemaakt¹⁵. 'Security' een ander belangrijk issue voor erfgoedinstellingen is niet in dit protocol afgevangen, maar kan eenvoudig geregeld worden via standaard http-mechanismen.¹⁶ De deelnemende instellingen in CATCH Plus committeren zich eraan dat zij gebruik maken van persistente identifiers en van de door DEN geadviseerde interoperabiliteitsformaten, en ze garanderen de toegang tot hun data volgens één van de door DEN geadviseerde protocollen. De CATCH Plus-activiteiten zijn hierbij ondersteunend.

¹³ De BASIS: <http://wiki.den.nl/Vindbaarheid>; ICT-register voor het Cultureel erfgoed: <http://www.den.nl/register/>

¹⁴ Europeana: <http://www.europeana.eu/home.php>

¹⁵ Rights statements: some repositories may find it desirable to attach terms of use to the metadata they make available through the OAI-PMH. No specific set of XML tags for rights expression is defined by OAI-PMH, but the about container is provided to allow for encapsulating community-defined rights tags;

Provenance statements: One suggested use of the about container is to indicate the provenance of a metadata record, e.g. whether it has been harvested itself and if so from which repository, and when. (<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>)

¹⁶ Arc - An OAI Service Provider for Digital Library Federation (Xiaoming Liu, Kurt Maly, Mohammad Zubair, Old Dominion University, Norfolk, Virginia USA; Michael L. Nelson, NASA Langley Research Center, Hampton, Virginia USA) (<http://www.dlib.org/dlib/april01/liu/04liu.html>)

In het werkpakket 'Framework en interoperabiliteit' zijn drie deelpakketten gedefinieerd:

1. **Workflowmanager**

De workflowmanager is een geformaliseerde en gestandaardiseerde workflowmanagementmodule. Doel is om te komen tot een eenvoudige, gestandaardiseerde en expliciete methode (bijvoorbeeld door een taal als BPEL¹⁷ te gebruiken) om tools en componenten te verbinden met webservices de het werk doen of data leveren. Het inpassen van nieuwe bouwstenen van buiten CATCH Plus wordt hiermee vergemakkelijkt, en instellingen kunnen bestaande bouwstenen (tools, webservices, eigen applicaties) op eenvoudige wijze aan elkaar koppelen tot nieuwe softwareonderdelen met een specifieke functie.

2. **Standaardisatie**

Enkele van de belangrijkste systemen in de erfgoedwereld (zoals AdLIB, TMS, GGC) hebben mogelijkheden om data gestandaardiseerd te kunnen harvesten of gedistribueerd te zoeken en te presenteren. In de praktijk wordt hier nog te weinig gebruik van gemaakt of blijken de resultaten nog niet voldoende. In dit deelpakket worden de mogelijkheden uit die systemen benut ten behoeve van CATCH Plus, en wordt tevens gewerkt aan een grotere nationale en internationale interoperabiliteit. Daar waar (meta)data nog niet in een gemeenschappelijk standaardformaat beschikbaar zijn, worden API's¹⁸ gerealiseerd om de (meta)data van de deelnemende partijen via een standaardformaat en met standaardprotocollen af te kunnen zoeken, te kunnen harvesten en te kunnen presenteren. Deze API's worden zodanig ontwikkeld dat ze als generieke tool ook voor andere erfgoedpartijen buiten het consortium bruikbaar zijn. Ook worden de voorwaarden gecreëerd voor standaardisatie (DIDL, METS, OAI-ORE, e.d.) van structurele metadata (de machineleesbare vorm van een object dat uit meerdere onderdelen bestaat). Dit is nodig om meervoudige objecten te kunnen presenteren in een interoperabele omgeving.

3. **Persistente identifiërs**

De ontwikkeling van een resolverstructuur voor persistente identifiërs van data en metadata bij de deelnemende instellingen, waarbij andere erfgoedinstellingen eenvoudig kunnen aansluiten. De resolverstructuur sluit aan bij bestaande nationale initiatieven binnen en buiten de erfgoedsector en de (inter)nationale bibliotheekafspraken op dit gebied. In deze structuur is iedere verwijzing naar een digitaal object een basis-URL naar de instelling die het object beheert. Ieder object krijgt een simpele en binnen de instelling unieke identifier. Een resolver verwijst vervolgens binnen de instelling door naar de fysieke plaats van het digitale object. Bij wijziging in die plaats (bijvoorbeeld vanwege vervanging van de server) hoeft alleen groepsgewijs de verwijzing in de resolver te worden aangepast en heeft dit geen gevolgen voor de buitenwereld. Groepen instellingen kunnen ook gezamenlijk een resolver inrichten. Invoering van een persistente identifier is van belang om de resultaten van gebruik van de centrale diensten en de tools duurzaam beschikbaar te houden. Dit onderdeel sluit aan bij de expliciete wens van OCW (Directie Cultureel Erfgoed) om een persistente-identifierstructuur in te voeren. De resolverstructuur van CATCH Plus fungeert als pilot hiervoor.

¹⁷ BPEL4WS is an XML based language for describing interaction between services on a network. The interactions take place via web services. BPEL4WS has facilities for sequencing commands, looping, conditional jumping, launching asynchronous and synchronous operations, message correlation and more.

¹⁸ Een API (application programming interface) is een klein stukje software dat een programmeur de gelegenheid geeft in zijn applicatie op eenvoudige wijze gebruik te maken van een andere applicatie of een database, zonder de interne code van die applicatie te hoeven kennen.

Globale werkwijze

De workflowmanager wordt vanuit het projectbureau ontwikkeld onder regie van de technische coördinator. Al naar gelang onderdelen van CATCH Plus gereedkomen wordt, indien wenselijk, gezorgd voor inpassing in dit framework. Om de mate van standaardisatie in kaart te brengen start het projectbureau een inventarisatie onder de deelnemende instellingen. Daar waar nodig worden vanuit de technische coördinator, in samenwerking met de betreffende instelling, de specificaties opgesteld voor een API om de betreffende collectie te kunnen gebruiken.

Het invoeren van de persistente-identificerinfrastructuur wordt eveneens vanuit het projectbureau aangestuurd. Na een ontwerpfase, waarin ook de andere partijen die in Nederland actief zijn op dit gebied worden betrokken, zullen de mogelijkheden en problemen voor de invoering van persistente-identificerinfrastructuur besproken worden met de deelnemende instellingen. Daarna wordt een plan van aanpak opgesteld en de uitvoering ter hand genomen. Dit onderdeel wordt aan het begin van het project uitgevoerd, omdat het belangrijk is dat alle partijen hier in hun ontwikkeling gebruik van gaan maken.

2.3.4 Werkpakket 4 Onderhoud en duurzaamheid

Het hoofddoel van CATCH Plus is software ontwikkelen. Het beheer en onderhoud van deze software na afloop van CATCH Plus moet geregeld worden om te voorkomen dat de investering verloren raakt. In dit werkpakket wordt hiervoor het kader met randvoorwaarden geschapen. De erfgoedsector kent wel een aantal infrastructurele partners en projecten, maar daarbinnen bestaat nog geen organisatie van erfgoedbrede software. De bijdrage van CATCH Plus kan niet anders zijn dan partieel. Inbedding in de beoogde nationale infrastructuur voor het digitale cultureel erfgoed is echter randvoorwaarde. Onderbrengen van beheer bij een of meerdere erfgoedinstellingen met een R&D-faciliteit behoort tot de mogelijkheden. Een andere optie is om het beheer onder te brengen bij een erfgoedbrede instelling als Stichting Digitaal Erfgoed Nederland¹⁹, parallel aan haar taak om het kwaliteitszorgsysteem van het digitale erfgoed op te zetten en te beheren. Ook beheer door SURFnet is een mogelijkheid die verder onderzocht kan worden. Randvoorwaarde voor structurele inbedding in een organisatie zijn onder andere duidelijke financiële kaders en garanties voor het beschikbaar stellen van de resultaten aan andere geïnteresseerde (erfgoed)instellingen. Ook het bedrijfsleven moet een rol spelen in het duurzame onderhoud van de CATCH Plus-resultaten. Gezamenlijk met Nederland Kennisland, die op het gebied van erfgoed en creatieve industrie ruime ervaring heeft opgebouwd, wordt nader onderzocht welke van de diverse CATCH Plus-resultaten interessant zijn om succesvol door het bedrijfsleven te worden opgepakt.

Het werkpakket bestaat uit twee onderdelen:

1. Onderhoud

Het bundelen van gezamenlijke belangen en behoeften in het cultureel erfgoed via CATCH Plus draagt bij aan de duurzaamheid van de ontwikkelde softwareapplicaties. Dit proces vraagt een gezamenlijke aanpak en regie. Door te investeren in kennis over een gezamenlijke aanpak stimuleert CATCH Plus het vermogen van erfgoedinstellingen om zich aan te passen aan verandering in hun primaire proces en permanent te investeren in ICT en innovatie. De kennis en kosten die met innovatie in de informatiesector gepaard gaan, zijn niet meer door afzonderlijke instellingen op te brengen. Een gezamenlijke aanpak vraagt om nieuwe businessmodellen, met name voor publiek-private samenwerking, ten behoeve van beheer en innovatie van software in het erfgoed. OCW, Nederland Kennisland en DEN gaan buiten dit project onderzoek doen naar businessmodellen in de erfgoedsector. CATCH Plus levert exemplarische businessplannen op en zal dat in nauwe samenspraak met genoemd initiatief doen. In dit onderdeel van CATCH Plus wordt tevens een model ontwikkeld en geïmplementeerd voor duurzaam beheer en doorontwikkeling van de CATCH

¹⁹ In een voorbereidend gesprek heeft DEN zich bereid verklaard deze optie te onderzoeken.

Plus software, dat rekening houdt met de bestaande nationale infrastructuur (waaronder DEN, Het Geheugen van Nederland, sectorinstellingen, kennisinstellingen), het auteursrecht, de financiële mogelijkheden van de sector en de generieke toepasbaarheid van de software voor het erfgoed. Nederland Kennisland zal hierbij een adviserende rol spelen.²⁰ Eén van de tot nu toe verkende modellen is het creëren van een vrije maar wel juridische vastgelegde omgeving, waarin erfgoedinstellingen en de creatieve industrie met enige regelmaat bij elkaar komen om over innovatie te praten en waarin op ad-hocbasis onderzoek en ontwikkeling kan worden gedaan. Deze samenwerkingsvorm kan zonnig de juridische basis zijn om op te treden als de formele partner bij het aanvragen van subsidies.²¹

Alle software komt in open-source-repositories beschikbaar voor gebruik en doorontwikkeling door erfgoed en bedrijfsleven, uiteraard indien het auteursrecht zich hier niet tegen verzet.

2. Disseminatie

Een duurzaam onderhoud van software in een gedistribueerde organisatievorm heeft alleen kans van slagen als de deelnemers voldoende op de hoogte zijn van elkaars mogelijkheden en onmogelijkheden en als er onderling vertrouwen heerst. Daarvoor is het nodig dat kennis niet binnen organisaties blijft, maar wordt uitgewisseld met andere partijen. Dat geldt voor de kennis die binnen het gebied van de softwareontwikkeling wordt gegenereerd én voor de kennis uit CATCH, het informaticaonderzoeksprogramma dat voorafgaat aan CATCH Plus (en waarvan de eerste vier projecten aflopen in 2009, de volgende zes in 2010 en waarin er nog vier projecten bijkomen die tot 2013 doorlopen). Het is van belang dat de informatica-kennisinstellingen actief betrokken blijven bij CATCH Plus. De huidige CATCH-gemeenschap is een succesvol voorbeeld hoe inspirerend en vruchtbaar het werkt als verschillende disciplines elkaar met grote regelmaat ontmoeten en gemeenschappelijke doelen benoemen. Uiteraard moet rekening gehouden worden met bedrijfsbelangen van de afzonderlijke instellingen en van de sector als geheel. Door middel van publicaties, workshopsessies, het opzetten van een gemeenschappelijk communicatiepunt en het organiseren van bijeenkomsten wordt de synergie tussen alle betrokken partijen bevorderd en de kennisuitwisseling bewaakt. De code van de ontwikkelde software wordt waar dat auteursrechtelijk is toegestaan als open-sourcesoftware onder een GNU General Public Licence versie 2.0 (GPL 2.0) bij een organisatie voor open source (SourceForge, OSOSS) geplaatst.

Er wordt een website ingericht waarop alle diensten en tools toegankelijk worden gemaakt.

Globale werkwijze

In nauwe samenwerking met OCW, Nederland Kennisland en DEN worden in enkele workshops de mogelijke samenwerkingsvormen na afloop van het project onderzocht. Het gaat om samenwerkingsvormen tussen erfgoedpartijen onderling en met private partners. Vervolgens wordt een keuze gemaakt voor een klein aantal (2-3) succesvol geachte combinaties, waarvoor een businessplan wordt ontwikkeld. Op de overige onderdelen is hierboven voldoende ingegaan.

2.3.5 Werkpakket 5 Projectmanagement

Werkpakket 5 'Projectmanagement' omvat alle organisatorische werkzaamheden die noodzakelijk zijn voor het bewaken van de voortgang en samenhang (waaronder de technische infrastructuur) van het project en het bereiken van de einddoelen. Het werkpakket omvat tevens de publicitaire activiteiten van het project. In de initiatiefase wordt in samenwerking met Nederland Kennisland onderzocht welke marktpartijen een rol kunnen spelen in de technische ontwikkeling tijdens en na afloop van het project. Op te leveren producten zijn onder andere projectinitiatieocument,

²⁰ Nederland Kennisland heeft ruime ervaring in het ontwikkelen van businessplannen en heeft ook het matchingssysteem Digitale Pioniers opgezet, waarin met innovatieve kleine software gewerkt wordt om hun producten economisch levensvatbaar te maken.

²¹ Geïnspireerd op het initiatief DevLab: <http://www.devlab.nl/index.html>

faseplannen, fase-eindverslagen, begrotingen, overzichten van inkomsten en uitgaven, hoofdpuntenrapportages per maand, afwijkingenrapportages en eindrapport.

2.4 Organisatie van CATCH Plus

CATCH Plus wordt uitgevoerd volgens de Prince2 projectmanagementmethode.

2.4.1 De organisatiestructuur

Alle deelnemende partijen in CATCH Plus committeren zich aan het projectvoorstel CATCH Plus. Om de uitvoering en een succesvolle oplevering te realiseren wordt gewerkt met de volgende organisatiestructuur:

1. **Raad van Toezicht.** De Raad van Toezicht bestaat uit

- de directeuren van alle deelnemende erfgoedinstellingen;
- de voorzitter en de vice-voorzitter van het dagelijks bestuur van de CATCH-programmacommissie;
- de directeuren van de in CATCH Plus deelnemende profit-bedrijven die uit eigen middelen in het project investeren;
- een vertegenwoordiger van NWO.

De Raad van Toezicht is de opdrachtgever van het CATCH Plus-project. Zij draagt de uitvoering van het project op aan een stuurgroep. De voorzitter van de stuurgroep legt verantwoording af aan de Raad van Toezicht. De Raad van Toezicht beslist ad hoc over het eventueel opnemen van nieuwe partners in het project.

De Raad van Toezicht vergadert minimaal 1 maal per jaar.

2. **Stuurgroep.** De stuurgroep is verantwoordelijk voor de uitvoering van het project. De dagelijkse uitvoering wordt opgedragen aan een projectmanager, bijgestaan door een projectbureau. Het leidende principe voor de stuurgroep is '*management bij exception*'. De stuurgroep bestaat uit:

- (Opdrachtgever) de voorzitter van de Raad van Toezicht (tevens voorzitter van de stuurgroep), bijgestaan door de vice-voorzitter van de CATCH-programmacommissie en een vertegenwoordiger van NWO;
- (Senior-gebruiker) de voorzitter van de begeleidingscommissie en een vertegenwoordiger namens de deelnemende erfgoedinstellingen;
- (Senior-leverancier) een vertegenwoordiger van de kennisinstellingen en een vertegenwoordiger namens de in CATCH Plus deelnemende bedrijven.

De projectmanager is secretaris van de stuurgroep. De voornaamste taken van de stuurgroep zijn:

- Goedkeuring projectinitiatiedocument en andere managementproducten opgeleverd door de projectmanager;
- Beslissing in die gevallen waarin het Projectplan niet voorziet, of afwijkingen daarvan. Financiële beslissingen die consequenties hebben voor de partners in het project behoeven de instemming van de verantwoordelijke in de instelling en worden gemeld aan de Raad van Toezicht;
- Acceptatie van de deliverables;
- Beslissing bij overschrijding van de vastgestelde project-toleranties.

Voorzitter: Prof. dr. Hans Bennis (directeur Meertens Instituut).

De stuurgroep vergadert minimaal 5 maal per jaar.

3. **Projectmanager.** De projectmanager voert het projectmanagement en wordt daarin ondersteund door een projectbureau dat bestaat uit een technisch coördinator en secretariële/financiële ondersteuning. De projectmanager

is hoofd van het projectbureau. Het projectbureau wordt bij het Nederlands Instituut voor Beeld en Geluid ondergebracht.

De voornaamste taken van de projectmanager zijn:

- Voeren van de dagelijkse projectleiding binnen de toleranties die zijn afgesproken met de stuurgroep;
- Bewaken van de voortgang van het project en verzorgen van de projectadministratie;
- Verzorgen van rapportages, publiciteit, onderlinge communicatie en disseminatie van de resultaten naar buiten het project;
- Bewaken van de oplevering van de afgesproken producten binnen de gestelde kwaliteitscriteria;
- Bewaken van de projectdoelstelling en de onderlinge samenhang van de projecten;
- Organiseren van vergaderingen en verzorgen van de administratieve werkzaamheden voor de stuurgroep en de begeleidingscommissie.

De projectmanager legt verantwoording af aan de stuurgroep.

4. **Begeleidingscommissie.** De begeleidingscommissie vervult de rol van kwaliteitsbewakende instantie voor de inhoudelijke doelen van het project. De begeleidingscommissie bestaat uit:

- Vertegenwoordigers uit de erfgoedwereld (waaronder DEN en Erfgoed Nederland);
- Een vertegenwoordiging vanuit de in CATCH participerende kennisinstellingen (informatici);
- De voorzitter van de CATCH-programmacommissie;
- Een vertegenwoordiger vanuit het CATCH-programmabureau.

De voornaamste taken zijn:

- Adviseren over de vast te stellen kwaliteitscriteria en beoordelen van de opgeleverde producten aan de hand van de kwaliteitscriteria;
- Geven van (on)gevraagd adviezen over de inhoudelijke uitvoering van het project.

Een vertegenwoordiging van de commissie brengt jaarlijks een site visit aan de projectuitvoerders, maar kan dit op eigen gezag ook vaker doen. De commissie rapporteert aan de voorzitter van de stuurgroep. De stuurgroep beslist over de adviezen, rapporteert daarover terug en legt daar verantwoording over af naar de Raad van Toezicht.

5. **Werkpakketleider.** De werkpakketleiders zijn verantwoordelijk voor de uitvoering van de werkzaamheden binnen de diverse werkpakketten. De voornaamste taken van de werkpakketleider zijn:

- Bewaken van de planning van de uitvoerende werkzaamheden;
- Verdelen van het werk onder de betrokken partijen;
- Zorgen dat de beoogde resultaten van het werkpakket opgeleverd worden en voldoen aan de gestelde kwaliteitscriteria.

De werkpakketleider is verantwoording schuldig aan de projectmanager over de uitvoering van het werkpakket.

6. **(Deel)projectleiders.** Met name in Werkpakket 2 Toolontwikkeling, waarbij onder verantwoordelijkheid van de deelnemende erfgoedinstellingen een aantal tools worden ontwikkeld en geïmplementeerd, worden projectleiders aangesteld. De voornaamste taken van de (deel)projectleider zijn:

- Bewaken van de planning van de uitvoerende werkzaamheden binnen zijn deelproject;
- Zorgdragen dat de beoogde resultaten van het deelproject opgeleverd worden en voldoen aan de kwaliteitscriteria van het deelproject en van CATCH Plus als geheel;

- Zorgdragen voor de instellingscontacten, het draagvlak in de instelling en de verspreiding van de resultaten binnen de instelling.

De (deel)projectleider wordt aangestuurd door het management van de erfgoedinstelling, is verantwoording schuldig aan de werkpakketleider en de projectmanager over de uitvoering van het deelproject

Rol van de erfgoedinstellingen: De erfgoedinstellingen leveren de praktijkkennis en de data. Zij sturen het deel toolontwikkeling dat zij tot hun verantwoordelijkheid nemen, geven hun akkoord aan de oplevering van dat onderdeel en de oplevering van het totaal aan diensten, software en infrastructuur.

Rol van de kennisinstellingen: De prototypen (demonstrators en pilots) en de kennis uit het onderzoeksprogramma zijn het startpunt voor CATCH Plus. De onderzoekers en programmeurs van CATCH dragen hun kennis en ervaring over op de uitvoerders van CATCH Plus in de vorm van wetenschappelijke advisering. Om tot effectieve valorisatie te komen is de communicatie tussen onderzoekers en uitvoerende partijen in alle stadia noodzakelijk. Op die manier wordt de continuïteit tussen CATCH en CATCH Plus gewaarborgd.

Rol van de softwarebedrijven: Via een open inschrijving kunnen softwarebedrijven hun belangstelling voor de uitvoering van bepaalde onderdelen duidelijk maken. De keuze wordt daarna door het projectbureau en de betreffende erfgoedinstelling gemaakt. Daarbij spelen kwaliteit, kosten en bestaande relaties een rol. Een erfgoedinstelling kan ook gebruik maken van een eigen programmeur. De softwarebedrijven moeten bereid zijn volgens het open-sourceprincipe te werken.

Rol van NWO: NWO als initiatiefnemer ondersteunt CATCH Plus bij de kennisbenutting (communicatie en kennisoverdracht) waarbij onderzoeksresultaten worden omgezet naar praktijktoepassingen. NWO beheert de financiën in het project, investeert er zelf middelen in en houdt zich op hoofdlijnen bezig met de bewaking van uitvoering en kwaliteit.

2.4.2 Fasering

CATCH Plus start in december 2008 en loopt tot en met het derde kwartaal van 2011.

In 2008-2009 zal de nadruk liggen op het ontwikkelen van de centrale diensten, de interoperabiliteit en de persistent identifier. In 2010 en 2011 vindt formeel de toolontwikkeling plaats - waarvoor de voorbereiding reeds in 2009 start -, komen de businessplannen tot stand en worden het onderhoud en de duurzaamheid van de projectresultaten geregeld. Een overzicht is uitgewerkt in Bijlage 3.

2.5 Technische architectuur

Het internet is het platform waarop de instellingen informatie verwerken en uitwisselen. Het is ook hun communicatiekanaal met de eindgebruiker. De architectuur in CATCH Plus voor het ontwikkelen van software is dan ook webgebaseerd. Een servicegeoriënteerde architectuur (SOA) is de aangewezen vorm hiervoor. Deze zal dan ook de leidende architectuur zijn voor de softwarebouw in CATCH Plus.

- **Webservices**

De gebruiker op internet ziet software en data in de vorm van webpagina's die worden aangeboden via websites. Vaak worden deze webpagina's real time samengesteld door webapplicaties op basis van data uit verschillende databronnen. Webpagina's zijn bedoeld voor menselijke communicatie. Ze bevatten layout, interactieve elementen en andere functionaliteiten. Nederlandse erfgoedinstellingen bieden momenteel de data uit hun digitale collecties op het internet bijna uitsluitend verpakt in zulke webpagina's aan. Dit maakt het in technische zin lastig voor machines om zonder tussenstappen direct collectiedata of -software te gebruiken. Daarnaast is het ingewikkeld om gezamenlijk te werken aan instellingoverstijgende webapplicaties.

Een architectuur van webservices is de aangewezen oplossing daarvoor. Hiermee wordt de toegang tot data en functionaliteit verpakt in webservices die zich op verschillende locaties op het web kunnen bevinden. Een dergelijk webservice is een klein stukje relatief zelfstandige software dat op een gestandaardiseerde manier wordt aangeroepen door een computerprogramma, niet door mensen. De service geeft eveneens op een gestandaardiseerde wijze de opgevraagde data terug, of een bewerking van de inkomende data. Het aanroepende programma maakt met deze resultaten vervolgens webpagina's voor presentatie aan en interactie met eindgebruikers of geeft ze door aan een andere webservice voor verdere processing.

Deze modulaire benadering biedt voor de samenwerking tussen erfgoedinstellingen binnen CATCH Plus grote voordelen: ieder van de partijen houdt de controle over toegang, inhoud en beheer van de aangeboden data of software-modules, terwijl die toch voor de andere partijen open en op een gestandaardiseerde manier beschikbaar zijn. Het maakt een heldere taakverdeling bij samenwerking aan softwareontwikkeling of het delen van datacollecties mogelijk, en het maakt het voor nieuwe partijen eenvoudiger met nieuwe services, data of applicaties aansluiting te vinden. CATCH Plus deliverables worden daarom opgeleverd als webservices, dan wel als webapplicaties die gebruik maken van de CATCH Plus 'pool' van webservices.

CATCH Plus levert twee centrale softwarediensten op (de repository of vocabularies en de annotation and recommendation) waarin de gemeenschappelijke aspecten van de onderzoeksprojecten bij elkaar zijn gebracht. Ook deze diensten worden als pakket van samenhangende webservices gebouwd. De output van deze centrale diensten wordt gebruikt in de tools die in werkpakket 2 worden ontwikkeld. Dat is de reden dat ze in de eerste fase van het project tot stand worden gebracht. De twee centrale diensten liggen aan de basis van veel erfgoedactiviteiten, het ontsluiten van informatie in onderlinge samenhang en het toevoegen van kennis aan objecten voor beter begrip en gebruik daarvan.

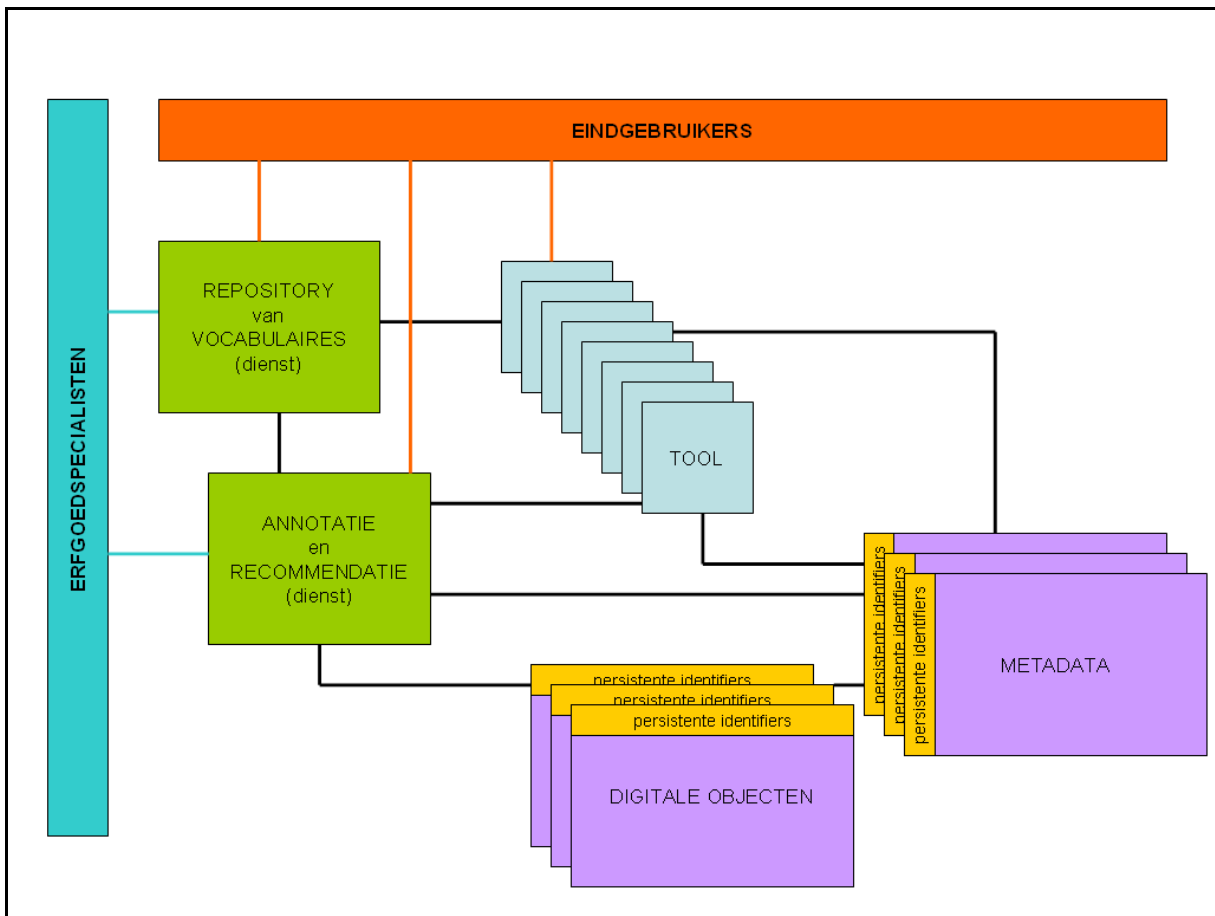
- **Standaarden en protocollen**

In de onderzoeksomgeving is het geen probleem om databestanden te kopiëren en naar eigen behoefte te structureren en te selecteren. In een bedrijfsomgeving is het noodzakelijk dat softwareapplicaties gebruik maken van echte, actuele data in plaats van demo-datasets of statische kopieën van catalogi. Hiervoor is het nodig om op eenduidige wijze online toegang tot collecties en vocabulaires te hebben. Gangbare methoden zijn harvesting (via bijvoorbeeld OAI-PMH) of federated search (bijvoorbeeld SRU). Binnen CATCH Plus wordt voor de deelnemende partijen deze online toegang tot digitale collecties en eventueel vocabulaires gerealiseerd. Er is toestemming om hierbij ook gebruik te maken van de hulpmiddelen en de kennis die zijn ontwikkeld in eCulture, onderdeel van het BSIK-project MultimediaN.

- **Persistente verwijzingen**

Zowel de vocabulairdienst als de annotation-dienst maken op grote schaal gebruik van referenties naar digitale objecten (dat geldt overigens ook voor de catalogi van de erfgoedinstellingen zelf, evenals voor vele andere services, webpagina's etc.) Als een dergelijk referentie verwijst naar een 'fysieke' opslaglocatie (bijvoorbeeld via een URL op het web) ontstaan er problemen bij verplaatsen, hernoemen of kopiëren van objecten. Referenties worden dan ongeldig. Ook hiervoor bestaan oplossingen: objecten worden dan eenmalig geïdentificeerd door middel van een unieke naam, een persistente identifier. Een resolutieservice (resolver) beheert de koppeling tussen deze identifier en de fysieke locatie(s) waar dat object gevonden kan worden. Bij verplaatsen hoeft alleen de resolver aangepast te worden. Een vergelijkbaar resolversysteem wordt ook voor de instellingen zelf gerealiseerd. Referenties gebruiken een open URL en de persistente identifiers. Op deze manier blijven de

referenties geldig. Binnen CATCH Plus worden voor de betreffende data van de deelnemende erfgoedinstellingen persistente identifiers met de bijbehorende resolver ingevoerd. Uiteraard wordt aansluiting gezocht bij lopende initiatieven in Nederland (DANS²²) en in Europa.



Figuur 1 CATCH Plus architectuur

De bovenstaande figuur toont de relatie tussen de verschillende onderdelen in CATCH Plus:

- De beide centrale diensten maken gebruik van elkaar. Ze kunnen ook samenwerken met één of meerdere tools en kunnen op hun beurt weer worden gevoed door de resultaten van de tools;
- Zowel de centrale diensten als de specifieke tools werken op basis van metadata, digitale objecten of informatie uit de centrale diensten;
- Metadata en digitale objecten blijven op de instellingen. Door het gebruik van persistente identifiers blijven de relaties op lange termijn bestaan;
- De tools hebben als resultaat ook verrijking van de bestaande metadata op de instellingen;
- De meeste tools en de centrale diensten hebben als directe doelgroep de erfgoedprofessionals, maar onderdelen kunnen ook door eindgebruikers worden gebruikt. Dat is bijvoorbeeld het geval bij de vocabulaires om semantisch te zoeken door verschillende collecties. Uiteindelijk moet CATCH Plus, door betere toegankelijkheid, de interactie tussen gebruiker en erfgoeddata verbeteren en culturele rijkdom binnen het bereik van iedereen brengen.

²² DANS: Data Archiving and Networked Services, instituut van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) en de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) die een dergelijke service voor de wetenschappelijke wereld beheert.

2.6 Positionering van Catch Plus in Nederland en daarbuiten

Nederland kent een landelijke infrastructuur voor het Digitale Cultureel Erfgoed. Hier is door de overheid de afgelopen acht jaar in geïnvesteerd. Onderdelen zijn Digitaal Erfgoed Nederland (standaarden, kennis), Het Geheugen van Nederland (digitaliseren, opslag, beschikbaarstelling), WatWasWaar (geografische toegang), Collectiewijzer (wegwijzer op collectieniveau), Digitaliseren met Beleid (subsidiëring van informatieplannen, ontwikkelingsprojecten), Beelden voor de Toekomst (digitalisering AV-materiaal), taskforces (professionalisering) en sectorinstituten. Leveranciers van enkele van de meest gebruikte systemen hebben hun systemen aangepast om uitwisseling en interoperabiliteit van data in principe mogelijk te maken. Uit eigen budgetten zijn door instellingen digitaliseringactiviteiten bekostigd. Ook de provinciale en gemeentelijke overheid draagt hieraan bij. Zeker moet ook genoemd worden het uit BSIK gefinancierde deelproject van MultimediaN eCulture (semantische integratie). Desondanks zijn er nog forse inspanningen nodig in stabiliteit en de daadwerkelijke inbedding en integratie van systemen, data en kennis in de erfgoedsector om tot een volwaardige infrastructuur te komen en de Digitale Collectie Nederland te realiseren.

De Digitale Collectie Nederland bestaat uit de verzameling van gedigitaliseerde deelcollecties. Deze digitale objecten zijn (deels) gedistribueerd opgeslagen, maar moeten zich naar de eindgebruiker voordoen als een ongedeelde collectie. Er is een duidelijke scheiding tussen data en de retrieval daarvan, en het verlenen van diensten met die data door cultureel erfgoedinstellingen en commerciële partijen. CATCH Plus sluit hier goed bij aan. Het is een innovatief onderdeel van deze infrastructuur. CATCH Plus draagt bij aan de interoperabiliteit van gedistribueerde en heterogene collecties. Het vindt oplossingen voor de semantische verschillen in de wijzen waarop collecties ontsloten zijn (verschillende trefwoordensystemen en dergelijke). Het draagt hulpmiddelen aan voor beter zoeken en vinden in gedistribueerde collecties, en maakt tools voor lastig te ontsluiten materiaal als handgeschreven tekst, audio en video. Bovenal zorgt het ervoor dat innovatieve informatica ingebed raakt in de dagelijkse praktijk van de erfgoedinstellingen.

Erfgoedinstellingen zijn nog sterk gericht op behoud en beheer en op beschikbaar stellen van analoge objecten. De omslag naar de digitale omgeving is gaande, maar verloopt in verschillend tempo. Bedrijven die applicaties maken voor erfgoedinstellingen volgen de internetontwikkelingen, maar hebben vanuit die situatie nog weinig stimulans om hun applicaties te vernieuwen voor innovatieve beschikbaarstelling van digitaal materiaal. Zij hebben belangstelling voor de ontwikkeling van CATCH, maar aarzelen om zelf te investeren in dergelijke innovatie. Een extra impuls draagt eraan bij om ze over de streep te trekken. In CATCH Plus zijn de voorlopers binnen het erfgoed verenigd die de mogelijkheden van de digitale wereld durven te verkennen en inzien dat digitale innovatie noodzakelijk is om mee te blijven spelen in een veranderende wereld. De behoefte om daar ook marktpartijen in mee te nemen is volop aanwezig. De scheiding tussen erfgoedinstellingen op internet en commerciële en particuliere initiatieven wordt steeds geringer en de expertise van de commerciële partijen kan goed gebruikt worden. Uit eigen middelen is het veelal niet mogelijk om die stap goed en consequent te zetten. CATCH Plus heeft niet de pretentie om het veranderingsprobleem van de erfgoedinstellingen op te lossen, maar levert wel een belangrijke bijdrage om stappen voorwaarts te zetten.

Samenwerking binnen de erfgoedsector gebeurt op vele fronten, maar vaak op ad-hocbasis. In CATCH zijn allerlei samenwerkingsverbanden ontstaan, met wederzijdse waardering. De CATCH-community komt op regelmatige basis bij elkaar over de vloer, leert van elkaar en werkt met elkaar. Vanuit de samenwerkingsgedachte die de overheid voorstaat en de netwerksamenleving waarin alle partijen functioneren, kan de gemeenschap in CATCH Plus uitgroeien tot een community die gezamenlijk de softwareontwikkeling kan sturen en aansturen, en die de data-interoperabiliteit ondersteunt. Dat houdt niet in dat zij het bedrijfsleven willen uitsluiten, integendeel, maar de gezamenlijke aanpak door

erfgoedinstellingen wordt wel de stabiele basis waarop het bedrijfsleven zijn commerciële producten kan bouwen, terwijl tegelijkertijd de vrije toegang tot de informatie zelf door de erfgoedinstellingen gewaarborgd blijft.

Europeana is een van de belangrijkste Europese convergerende activiteiten op het gebied van het internationale erfgoed. De overheden in de lidstaten hebben zich eraan gecommitteerd om hun nationale digitaliseringactiviteiten af te stemmen op de realisatie van Europeana. Europeana biedt een centraal punt waar de nationale digitale collecties aan elkaar worden verbonden. De gerichtheid op Europeana schept op nationaal niveau toenemende eenheid in het denken over beleid, standaarden en infrastructuur. Deelnemers aan CATCH Plus maken ook deel uit van de diverse commissies die inhoud en ontwikkeling van Europeana bepalen. De Koninklijke Bibliotheek biedt Europeana onderdak en OCW ondersteunt Europeana beleidsmatig en financieel. De uitvoering van CATCH Plus ligt in lijn met de Europeana-activiteiten en de resultaten van CATCH Plus dragen bij aan het succes van Europeana als portal en als initiatief met grote aanjagende kracht op het gebied van publieksbereik, interoperabiliteit en Europese samenwerking. CATCH Plus zal het contact met Europeana versterken om de resultaten ook daar te laten toepassen. Hierdoor worden voor Nederland ook nieuwe mogelijkheden geschapen om te participeren in Europese erfgoed- en informaticaprojecten.

Het informatica-onderzoeksprogramma CATCH kent internationaal erkenning. In wetenschappelijke publicaties en op congressen zijn de resultaten uitgedragen. Door deze contacten is duidelijk dat in andere landen de 'uitvindingen' van CATCH nog niet bestaan. De Internationale Wetenschappelijke Advies Commissie (ISAB), die voor het onderzoeksprogramma de projectvoorstellen beoordeelt, en de gebruikerscommissie, die bij de uitvoering betrokken is, hebben steeds geoordeeld dat wat in Nederland gebeurt uniek is. Conclusie is dat het uitbouwen en implementeren van de CATCH-resultaten niet overlappend is met elders ontwikkelde producten en diensten. De deelnemende erfgoedinstellingen zijn vrijwel alle internationale spelers die goed op de hoogte zijn van ontwikkelingen elders. Hun commitment om te participeren in CATCH Plus is voldoende om de waarde van de deliverables van CATCH Plus te tonen.

3 Relatie tot ICT-agenda

De ICT-agenda 2008-2011, met als titel 'De gebruiker centraal in de digitale dienstenmaatschappij', noemt cultureel erfgoed expliciet als gebied waar ICT-innovatie van groot belang is en dat uitstekende mogelijkheden biedt voor innovatieve ICT-toepassing ('cultuur, digitalisering en innovatie'). Het domein Digitaal Cultureel Erfgoed (de Digitale Collectie Nederland) is bij uitstek gericht op gebruikers. Het hoofddoel van de digitalisering van het cultureel erfgoed is om grotere publieksparticipatie en betere interactie tussen erfgoed en gebruiker te realiseren. Het implementatieproject CATCH Plus levert in vervolg op het informatica-onderzoeksprogramma CATCH een belangrijke en innovatieve bijdrage aan het verwezenlijken van die doelen. Doordat in CATCH Plus gemeenschappelijke applicaties worden gebouwd en de interoperabiliteit in de sector verder wordt ontwikkeld, wordt een belangrijke bijdrage geleverd aan de infrastructuur van de sector. De erfgoedpartners investeren zelf ook aanzienlijk in de ontwikkeling van deze infrastructurele voorziening, waardoor duurzaam gebruik gewaarborgd is.

CATCH Plus sluit specifiek aan bij de volgende doelen, genoemd in de ICT-agenda 2008-2011:

- **Faciliteren van hoogwaardige dienstverlening voor burgers en bedrijven en van productiviteitsgroei door mogelijk maken van interoperabiliteit waar marktpartijen zelf niet in kunnen voorzien:**

CATCH Plus ontwikkelt applicaties met gestandaardiseerde interfaces die door alle gebruikers in het erfgoed kunnen worden gebruikt, mits hun collectie via standaarden interoperabel is gemaakt. Om dat mogelijk te maken ontwikkelt CATCH

Plus waar nodig generiek bruikbare software om de collecties van afzonderlijke instellingen voor de gebruiker transparant als één collectie bruikbaar te maken, terwijl de data zelf bij de afzonderlijke erfgoedinstellingen blijven staan.

- **Ondersteunen van gebruik van open standaarden en het bevorderen van interoperabiliteit binnen de overheid en tussen overheid en bedrijven:**

Alle resultaten van CATCH Plus worden (mits bestaande rechten daarmee niet worden geschonden) als open source en met open standaarden ontwikkeld. Doordat de grote erfgoedinstellingen participeren in CATCH Plus wordt hiermee de facto een belangrijke stap voorwaarts gezet in de interoperabiliteit van collecties en diensten. Door het gebruik van standaarden en door de collecties open te maken kan het bedrijfsleven hierop zonder technische hindernissen creatieve applicaties bouwen. Software en documentatie worden bij SourceForge of een vergelijkbare organisatie gedeponeerd. Door het bedrijfsleven bij ontwikkeling en bouw te betrekken wordt de benodigde kennis overgedragen. Informatici en programmeurs worden voor langere tijd aan de erfgoedsector gebonden, zodat hun kennis kan worden ingezet.

- **Duurzame toegankelijkheid van digitale informatie waarborgen:**

De invoering van een systeem van persistente identifiers om de interoperabiliteit en de duurzaamheid van de kennisverrijking mogelijk te maken, draagt in belangrijke mate bij aan de duurzame toegankelijkheid van de data zelf. In de internationale bibliotheekwereld is een dergelijke persistente identifier-resolverstructuur reeds in ontwikkeling. Hierbij wordt aangesloten. Omdat de leveranciers van 'erfgoedapplicaties' uitdrukkelijk bij CATCH Plus worden betrokken, kan bij hen commitment worden bereikt om hun applicaties aan te passen aan deze persistente structuur. Verdere stimulering vanuit de overheid is wenselijk om te komen tot een werkelijk interoperabele en duurzame toegankelijkheid van digitale erfgoed informatie.

- **Ontwikkelen Digitale Collectie Nederland (versterken en uitbouwen infrastructuur ten behoeve van Digitale Collectie Nederland):**

CATCH Plus maakt zelf geen materiaal digitaal, maar door de interoperabiliteit in Nederlandse erfgoedinstellingen naar een hoger plan te tillen, draagt het in belangrijke mate bij aan de integratie van de afzonderlijke collecties en dus aan de vorming van de Digitale Collectie Nederland. De nadruk van CATCH Plus ligt op het ontwikkelen van diensten en tools als onderdeel van de infrastructuur in Nederland, die ook buiten de kring van primaire deelnemers gebruikt kunnen worden.

- **Verdere ontwikkeling European Digital Libraries (EDL) en toegankelijkheid Europees erfgoed, mede ter versterking van de Nederlandse positie in de kenniseconomie:**

Door haar bijdrage aan interoperabiliteit en standaardisatie van het Nederlandse erfgoed, die in wezen geen andere is dan die op Europees niveau, draagt CATCH Plus bij aan het gestandaardiseerd beschikbaar maken van Nederlands erfgoedmateriaal dat zonder meer door Europeana (voorheen EDL) kan worden geharvest, hergebruikt en verrijkt. Nu al is Nederland een van de belangrijke toeleveranciers aan Europeana met relatief veel digitaal materiaal. De activiteiten van CATCH Plus dragen bij aan betere ontsluiting van het Nederlands erfgoedmateriaal voor gebruik in Nederlandse en Europese toepassingen. Erfgoedmateriaal is niet alleen 'ter lering ende vermaak', maar een uiterst belangrijke bron van kennis over onze maatschappij. De internationaal erkende hoogwaardige informaticakennis die in CATCH is ontwikkeld, en de valorisatie daarvan in de CATCH Plus-toepassingen, versterken de Nederlandse positie in de kenniseconomie.

4 Intellectueel eigendom

De software ontwikkeld binnen CATCH Plus wordt na afloop van het programma vrij ter beschikking gesteld voor de cultureel-erfgoedsectoren en voor het bedrijfsleven. De ontwikkelde software wordt met GNU General Public Licence versie 2.0 (GPL 2.0) of hoger ter beschikking gesteld via een breed toegankelijk platform. Gedacht wordt aan SourceForge of OSOSS. Hiervan kan om praktische redenen worden afgeweken. Uiteraard worden de auteursrechten te allen tijde geëerbiedigd, waardoor het nodig kan zijn van bovengenoemde uitgangspunten af te wijken.

De eigendomsrechten voor kennis en intellectueel eigendom binnen CATCH Plus worden vroegtijdig vastgelegd in een overeenkomst tussen de betrokken partners in het project. De projectovereenkomst dient rekening te houden met aspecten zoals (commerciële) exploitatie, publicaties, geheimhouding, toetreding nieuwe partners, nieuwe financiering, et cetera. De organisatie Nederland Kennisland is bereid haar expertise in te zetten om het evenwicht te vinden tussen een zo groot mogelijk profijt van de ontwikkelde kennis en software voor het cultureel erfgoed, de commerciële belangen van het bedrijfsleven en juridisch houdbare regelingen.

Er worden geen data geïntegreerd in de applicaties. De data blijven op de erfgoedinstellingen. Het gebruik van protocollen en standaarden maakt het mogelijk de objecten decentraal te blijven opslaan, waarmee de rechtendiscussie rond gebruik van objecten weliswaar niet wordt opgelost, maar wel wordt vereenvoudigd. In het OAI-protocol zijn afspraken t.a.v. 'rights' en 'trust' gemaakt²³. Data die door verrijking of anderszins worden toegevoegd, komen onder een creative-commons-licentie. Voor zover nu voorzien worden alleen thesaurusdata centraal opgeslagen en voor een ruimere groep beschikbaar gesteld dan nu het geval is. In die gevallen zal zoveel mogelijk getracht worden gebruikslicenties te verkrijgen die de hele erfgoedgemeenschap ten goede komen.

²³ (<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>)

Bijlage 1 Belanghebbendenanalyse

Belanghebbende ²⁴	Belang	Belangrijkheid (laag/medium/ hoog)
Musea, bibliotheken, archieven	De CATCH Plus-resultaten omvatten tools en methoden die afgestemd zijn op de huidige ICT-problematiek binnen de musea. De resultaten van CATCH Plus worden ingezet als nieuwe vormen van dienstverlening. Het innovatieve karakter ervan verbetert het imago van erfgoedinstellingen en de dienstverlening voor een breed publiek.	Hoog
Bedrijfsleven	Innoveren is noodzaak. Het bedrijfsleven, met name het MKB, kan hierin niet achterblijven. CATCH Plus betekent een nieuwe manier van ondernemen door middel van 'open innovatie'. Ondernemers werken samen met onderzoekers en klanten.	Hoog
Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)	Eén van de drie pijlers van de NWO strategie 2007-2010 is 'Wetenschap voor de Samenleving' ²⁵ . Maatschappelijke vragen en de roep om versterking van maatschappelijke en technologische innovatie zorgen voor een steeds sterkere behoefte aan snel toepasbare kennis en valorisatie. Als intermediaire organisatie brengt NWO wetenschappers, kennisvragers (o.a. professionals uit het bedrijfsleven) bij elkaar. NWO als initiatiefnemer ondersteunt CATCH Plus bij de kennisbenutting (communicatie en kennisoverdracht) waarbij onderzoeksresultaten worden omgezet naar praktijktoepassingen.	Medium
Universiteiten	Een belangrijk rendement van kennisontwikkeling is de benutting van deze kennis. Kennisvalorisatie is, naast onderwijs en onderzoek, één van de kerntaken van de universiteiten. In lijn met de roep van de overheid om kennis te verzilveren biedt CATCH Plus een gelegenheid voor kennisinstellingen om de onderzoeksresultaten geschikt en beschikbaar te maken voor benutting in de praktijk. CATCH Plus scheidt werkgelegenheid voor onderzoekers en programmeurs die de overstap naar het erfgoed willen maken.	Medium

²⁴ Hoewel de CATCH- en CATCH Plus-partners een speciaal belang hebben, zijn ze hier opgenomen als onderdelen van het geheel (musea, archieven, bibliotheken aan de ene kant; universiteiten aan de andere kant).

²⁵ NWO-strategie 2007-2010: Wetenschap gewaardeerd! (uitgave mei 2006)

Overheid	De overheid, met name EZ en OCW, onderstrepen het belang van kennisvalorisatie bij een vitale economie. CATCH Plus brengt de werelden van ondernemers, klanten en onderzoekers bij elkaar. OCW maakt zich sterk voor een erfgoedbrede infrastructuur. CATCH Plus vormt daarin een innovatieve tak en bevordert samenwerking en integratie. Daarnaast draagt CATCH Plus bij aan de speerpunt van ICT-benutting uit het 'Vervolg Rijksbrede ICT-agenda 2006-2007: Acties voor Nederland in verbinding'.	Medium
HBO	Het HBO, met name de opleidingen voor de creatieve sector, kan profiteren van de wetenschappelijke input. Het inzetten van HBO-studenten bij opdrachten vanuit CATCH Plus is een mogelijkheid. Na afloop is het culturele werkkapitaal beschikbaar.	Medium

Bijlage 2 Risicoanalyse

Risico	Waarschijnlijkheid van optreden (1-5)	Omvang van het risico (1-5)	Score (W X O)	Actie ter voorkoming/beheersing van risico
Werkpakketniveau				
1. vertrek van uitvoerders	2	3	6	Formatie opnieuw vaststellen en werkzaamheden voldoende documenteren
2. wijziging in personeel	2	3	6	Plan van overdracht en werkzaamheden voldoende documenteren
Managementniveau				
1. personele wijziging in management	1	4	4	Goede voorbereiding en inschatting van activiteiten. Bewaken van activiteitenplan en regelmatige interne afstemming
2. verlies van interesse bij consortiumpartners	2	2	4	Regelmatig relatieonderhoud en informatie-uitwisseling
Technisch niveau				
1. benodigde middelen ontbreken	2	2	4	Regelmatige afstemming en tijdige inzet van middelen
2. snelle ontwikkeling van IT maakt resultaten achterhaald	3	1	3	Informatie-uitwisseling van activiteiten met werkveld bij fora en symposia
Samenwerkingspartners				
1. samenwerking komt niet tot stand	2	2	4	Bouwen aan onderling vertrouwen en afstemming
2. bedrijfsleven haakt af tijdens de projectuitvoering	2	3	6	Communicatie en goede overeenkomst opstellen
Juridisch niveau				
1. geen overeenstemming tussen partijen	2	2	4	Goede communicatie en onderling vertrouwen. Afspraken vastleggen in een overeenkomst

Bijlage 3 Overzicht begroting en liquiditeitsbehoefte

Start: 1-12-2008

Benodigde bedragen per jaar:

Jaar	Werkpakket 1		Werkpakket 2		Werkpakket 3.3		Werkpakket 3.1 & 3.2		Werkpakket 4		Werkpakket 5		TOTALEN	
	Centrale diensten		Toolontwikkeling		Framework en Interoperabiliteit				Onderhoud en duurzaamheid		Project-management			
	Repository en A&R-tool		Software-ontwikkeling		Persistente identifiers		Workflow manager en standaardisatie		Onderhoud en disseminatie		Bewaking en uitvoering			
	Eigen bijdrage	PRIMA bijdrage	Eigen bijdrage	PRIMA bijdrage	Eigen bijdrage	PRIMA bijdrage	Eigen bijdrage	PRIMA bijdrage	Eigen bijdrage	PRIMA bijdrage	Eigen bijdrage	PRIMA bijdrage		
2008	261.959	250.588	85.912	0	80.000	0	4.700	0	17.000	0	69.900	0	519.471	250.588
2009	157.176	150.353	85.912	171.823	10.000	0	37.600	0	34.000	0	69.900	0	394.587	322.176
2010	52.392	50.118	343.646	687.293	10.000	0	4.700	0	51.000	0	69.900	0	531.638	737.410
2011	52.392	50.118	57.274	286.372		0	0	0	68.000	0	23.300	0	200.966	336.490
Subtotaal	523.919	501.175	572.744	1.145.488	100.000	0	47.000	0	170.000	0	233.000	0	1.646.663	1.646.663
Totaal	1.025.094		1.718.232		100.000		47.000		170.000		233.000		3.293.326	

Inkomsten	Aard
OCW	400.000 in geld
NWO	400.000 in geld
Cultureel erfgoed	572.744 in natura
Kennisinstellingen	107.919 in natura
Beeld en Geluid	116.000 in natura ²⁶
Bedrijven	50.000 PM ²⁷
PRIMA	1.646.663 in geld
Totaal	3.293.326

Uitgaven	
Werkpakket 1	1.025.094
Werkpakket 2	1.718.232
Werkpakket 3	147.000
Werkpakket 4	170.000
Werkpakket 5	233.000
Totaal	3.293.326

²⁶ Bijdrage van Nederlands Instituut voor Beeld en Geluid ten behoeve van financiële en administratieve ondersteuning voor het projectbureau.

²⁷ PM-post: pro memorie

Toelichting begroting en liquiditeitsbehoefte CATCH Plus

Inkomsten

Bijdrage OC&W

De bijdrage OC&W bestaat uit een lumpsumvergoeding. Deze vergoeding kan worden onderverdeeld in:

- vergoeding voor de personele kosten binnen Werkpakket 1;
- vergoeding voor de kosten van de softwareontwikkeling en materiële aanschaf binnen Werkpakket 1;
- vergoeding voor de ontwikkeling van de persistente identifiers binnen Werkpakket 3.

Bijdrage NWO

De bijdrage NWO bestaat uit een lumpsumvergoeding. Deze vergoeding kan worden onderverdeeld in:

- vergoeding voor de personele kosten binnen Werkpakket 5;
- vergoeding voor de uitvoering van Werkpakket 4;
- vergoeding voor de uitvoering van de workflow manager en interoperabiliteit binnen Werkpakket 3.

Bijdrage cultureel erfgoedinstellingen

Alle deelnemende erfgoedinstellingen hebben het CATCH Plus projectvoorstel geaccepteerd. Hiermee committeren de erfgoedinstellingen zich financieel aan de uitvoering van CATCH Plus. De bijdrage van de cultureel erfgoedinstellingen bestaat uit een vergoeding in natura. De inzet in manuren wordt gedaan vanuit de erfgoedinstelling ten behoeve van de softwareontwikkeling van eigen deelprojectplan. In het geval van het Nederlands Instituut voor Beeld en Geluid bestaat de bijdrage niet alleen uit het inzetten van manuren voor het deelproject onder Werkpakket 2 maar ook de huisvesting inclusief administratieve ondersteuning van het projectbureau.

Bijdrage kennisinstellingen

De bijdrage vanuit de kennisinstellingen bestaat uit een vergoeding in natura. Hierbij gaat het om de tijd die de onderzoekers en wetenschappelijke programmeurs gaan inzetten ten behoeve van kennisoverdracht. Enkele van de hoogleraren zijn bereid om een deel van de tijd op zich te nemen, voor zover nodig, als wetenschappelijke begeleiding.

Bijdrage softwarebedrijven

De bijdrage vanuit het bedrijfsleven is een PM-post (pro memorie). Het bedrag is een indicatie die de consortiumpartners verplicht om aan het eind van CATCH Plus deze bijdrage vanuit het bedrijfsleven te kunnen opvoeren in de financiële verantwoording.

Uitgaven

Personele lasten

De personele lasten opgenomen in de begroting worden nagenoeg geheel gedekt uit de lumpsumvergoeding vanuit NWO en OCW. De personele kosten zijn gebaseerd op standaardbedragen per 1 juli 2008 volgens NWO-VSNU afspraken.

Financiële ondersteuning

- NWO voert administratie uit op werkpakketniveau op basis van begroting maximaal een jaar vooruit met verantwoording per jaar achteraf, daarna budget voor een volgende periode, eind 1-12-2011; 10% wordt pas betaald na oplevering van een inhoudelijke en financiële eindrapportage. Alle betalingen worden uitbetaald aan

het projectbureau. Behalve in Werkpakket 2 Toolontwikkeling, dan geschiedt de betaling rechtstreeks naar de uitvoerende erfgoedinstelling;

- Het projectbureau voert administratie uit op activiteitsniveau op basis van elke uitgevoerde activiteit binnen het werkpakket;
- De deelnemende erfgoedinstelling voert administratie uit op activiteiten binnen eigen projectplan;
- Bovengenoemde drie partijen voeren financiële verantwoording aan. NWO is verantwoordelijk voor de eindcontrole en legt de financiële verantwoording op projectniveau aan OCW.

De kosten voor de financiële ondersteuning vanuit het projectbureau zijn opgenomen in de begroting als bijdrage van het Nederlandse Instituut voor Beeld en Geluid. De overige kosten voor de financiële ondersteuning komen ten laste van vanuit NWO en de deelnemende erfgoedinstellingen.

Softwareontwikkeling

De keuze voor een softwareontwikkelaar geschiedt door open inschrijving. Het projectbureau voert tijdens de bouw directe regie over de inhoud van de software. De kosten voor de softwareontwikkeling in Werkpakket 1 Centrale Diensten en Werkpakket 2 Toolontwikkeling worden financieel gedekt door de PRIMA-subsidie.

Onderhoud en duurzaamheid

De lasten binnen Werkpakket 4 Onderhoud en duurzaamheid bestaan uit kosten voor de totstandkoming van de businessmodellen en de organisatie van workshops en bijeenkomsten met als doel kennisdisseminatie. Hierbij wordt gedacht aan het onderzoeken en opzetten van randvoorwaarden voor publieke-private samenwerking ten behoeve van beheer en behoud van de CATCH Plus-resultaten na afloop van het project.

Bijlage 4 Evaluatieplan

De projectmanager is bij de start van het project verantwoordelijk voor het formuleren van een evaluatieplan waarbij de bereikte resultaten worden afgewogen aan de hand van het projectinitiatiedocument (PID). Hierbij wordt de PRINCE2-methodiek gehanteerd waarbij twee momenten in de projectuitvoering belangrijk zijn: midterm-evaluatie (tussentijdse evaluatie) en eind evaluatie. Voor beide activiteiten wordt een panel ingesteld bestaande uit onafhankelijke deskundigen.

Timing	Belangrijke aspecten	Doel	Methode
Midterm:2010	<ul style="list-style-type: none"> - Projectmanagement - Technische uitvoering - Kwaliteitsbeoordeling 	<p>Het analyseren van faseovergangen;</p> <p>Het managen van risico- en succesfactoren;</p> <p>Het identificeren van vervolgacties.</p>	<p>Aan de hand van de presentaties van de tussentijdse projectresultaten worden aanbevelingen geformuleerd.</p>
Eind: 2011	<ul style="list-style-type: none"> - Projectmanagement - Technische uitvoering - Implementatie - Duurzame ontwikkeling 	<p>Het beoordelen van managementprocessen;</p> <p>Het verifiëren van (on)verwachte deliverables;</p> <p>Het formuleren van aanbevelingen.</p>	<p>Aan de hand van een post-projectbeoordelingsplan worden de voorspelde resultaten vergeleken met de werkelijke resultaten.</p>

Bijlage 5 Overzicht CATCH Plus-consortium partners

Cultureel-erfgoedinstellingen	Lopende CATCH-projecten
Rijksmuseum Amsterdam	CHIP (Cultural Heritage Information Personalization)
Nederlands Instituut voor Beeld en Geluid, Instituut Collectie Nederland	CHOICE (Charting the information landscape employing Context Information)
Gemeentearchief Rotterdam	CHoral (Access to oral history)
Naturalis	MITCH (Mining for Information in Texts from the Cultural Heritage)
Nederlands Instituut voor Beeld en Geluid	MuNCH (Multimedia aNalysis for Cultural Heritage)
Gemeentemuseum Den Haag	MuSeUM (Multiple-collection Searching Using Metadata)
Rijksdienst voor Archeologie, Cultuurlandschap en Monumenten (RACM)	RICH (Reading Images in the Cultural Heritage)
Nationaal Archief	SCRATCH (SCRipt Analysis tools for the Cultural Heritage)
Koninklijke Bibliotheek	STITCH (Semantic Interoperability To access Cultural Heritage)
Meertens Instituut	WITCHCRAFT (What is Topical in Cultural Heritage: Content-based Retrieval Among Folksong Tunes)

De deelnemende onderzoeksinstituten:

- Max Planck Instituut
- Rijksuniversiteit Groningen
- Technische Universiteit Eindhoven
- Telematica Instituut
- Universiteit van Amsterdam
- Universiteit van Tilburg
- Universiteit Twente
- Universiteit Utrecht
- Vrije Universiteit Amsterdam

Initiatiefnemer CATCH Plus:

- Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO)

Bijlage 6 Uitwerking centrale diensten (Werkpakket 1)

1. Repository of vocabularies

Binnen erfgoedinstellingen en onderzoeksprojecten nemen thesauri en andere terminologische bronnen een belangrijke plaats in. Er wordt veel tijd en kennis gestopt in het opbouwen, onderhouden, converteren, verrijken en op elkaar afstemmen van deze thesauri. Ook binnen CATCH houdt een aantal onderzoeksprojecten zich hiermee bezig. Een verkennend onderzoek binnen CATCH leverde desondanks op, dat dergelijke vocabulaires nauwelijks door de erfgoedinstellingen op het web worden aangeboden. Als dat wel gebeurt, is het bijna uitsluitend in de vorm van webpagina's, die wel door mensen, maar niet door computers kunnen worden gebruikt.

In CATCH is daarom samengewerkt aan een repository waarin vocabulaires op uniforme wijze met semantisch-webtechnologieën worden samengebracht en via webservices op het internet beschikbaar worden gemaakt. Ook relaties tussen termen in verschillende thesauri worden verzameld en via webservices opgevraagd of beheerd. Pilottoepassingen die hiermee in CATCH zijn gerealiseerd zijn het zoeken en browsen naar beschrijvende termen door documentalisten, het uploaden van thesauri naar de repository, het aanbieden van gefilterde termenlijsten in een annotation-omgeving (autocompletion) en het automatisch annoteren van teksten met termen uit een thesaurus. Daarnaast biedt de repository ook informatie aan menselijke gebruikers: informatie over specifieke thesauri en het gebruik daarvan wordt over en weer gelinkt aan de DEN overzichtssite (<http://www.den.nl/kennis/thema/terminologie>). De demo-repository kent zijn eigen browse- en zoekinterface.

Binnen CATCH Plus wordt het repository productierijp herbouwd en opgeschaald, en wordt de functionaliteit uitgebreid om er een voor de erfgoedwereld rijke en stabiele hulpbron van te maken. Er zullen meer thesauri en mappings worden toegevoegd, waarbij de performance uitstekend moet blijven. Nieuwe services worden toegevoegd zoals het automatisch aan elkaar relateren van termen, het automatisch aanbrengen van verrijkingen en thesaurusbeheerfuncties (zowel voor collectiebeheerders als voor geïnteresseerde web communities).

2. Annotation & recommendation

In nagenoeg ieder van de lopende CATCH-projecten is in één of andere vorm sprake van annotaties. Er is echter een zeer grote variatie als het erom gaat welke objecten geannoteerd worden (complete digitale objecten, delen van plaatjes, delen van audio- en video-opnamen, delen van teksten), de wijze waarop die objecten geannoteerd worden (met tekstlabels, met getallen, jaartallen, of met concepten) en de wijze waarop annotaties worden gerepresenteerd en opgeslagen. Binnen CATCH wordt samengewerkt om deze heterogene annotaties te uniformeren en te verzamelen in een via webservices toegankelijke, centrale annotation repository. Dit levert op zichzelf al een waardevolle gedeelde bron op: de repository fungeert als een centrale index, die doorzoekbaar is op tekst of concepten, en die verwijst naar fragmenten van multimediale digitale objecten in collecties van verschillende erfgoedinstellingen (een muziekinstrument op een schilderij, een regel handgeschreven tekst, een scène in een televisieprogramma of een woord in een boektekst).

Bovendien ondersteunt de repository het 'stapelen' van annotaties. Bijvoorbeeld teksten afkomstig van automatische spraakherkenning of van handmatige transcriptie van handgeschreven documenten, kunnen automatisch worden gekoppeld aan concepten uit een thesaurus (die op zijn beurt weer is ondergebracht in de

repository of vocabularies). Dit levert een rijk netwerk van referenties op, dat uiteindelijk concepten verbindt met objecten in multimediale collecties.

Binnen het lopende CATCH-onderzoeksprogramma is een eerste versie van de annotation repository ontworpen en succesvol geïmplementeerd, evenals bijbehorende webservices voor beheer en doorzoeken ervan. Daarnaast zijn een aantal applicaties en aanvullende services ontwikkeld. Voorbeelden daarvan zijn een webtool voor handmatige annotatie van beelden op het internet, en een omgeving voor het suggereren van relevante thesaurustermen aan documentalisten op basis van automatisch gegenereerde annotaties van teksten (recommendation).

Om voor erfgoedinstellingen een waardevolle hulpbron te worden moeten er grote aantallen annotaties beschikbaar zijn, die samen een interessant referentie-netwerk tussen de diverse collecties vormen. CATCH Plus zal zich dan ook richten op populatie van de repository door (1.) de binnen CATCH ontwikkelde automatische annotatie-algoritmen te verbeteren en als services in te zetten, (2.) het geïnteresseerde publiek interactief bij annotatie-taken te betrekken, en (3) bestaande annotatie-data te gebruiken. Streven naar grote aantallen annotaties impliceert ook dat er substantieel aandacht wordt gegeven aan schaalbaarheid van de repository en aan snelle doorzoekbaarheid. Ten slotte zal een multimediale browse-/zoekomgeving worden gebouwd, met viewers voor annotaties in de context van uiteenlopende digitale objecten.

Bijlage 7 Uitwerking Toolontwikkeling (Werkpakket 2)

De deelnemende erfgoedinstellingen zijn:

- | | |
|--|---------|
| 1. Gemeentearchief Rotterdam | blz. 37 |
| 2. Gemeentemuseum Den Haag | blz. 39 |
| 3. Instituut voor Beeld en Geluid | blz. 41 |
| 4. Koninklijke Bibliotheek | blz. 44 |
| 5. Meertens Instituut | blz. 46 |
| 6. Naturalis en Rijksdienst Archeologie, Cultuurlandschap en Monumenten (RACM) | blz. 49 |
| 7. Nationaal Archief | blz. 51 |
| 8. Rijksmuseum Amsterdam | blz. 53 |

CATCH Plus voor Gemeentearchief Rotterdam - contactpersoon: Jantje Steenhuis/Mies Langelaar

Huidig CATCH-project: CHORAL

Werktitel: Indexeren van audioopnames met behulp van spraaktechnologie en notulen

Achtergrond

Op dit moment worden de vergaderingen van de Rotterdamse gemeenteraad opgenomen en door middel van streaming in tijdsblokken aangeboden via www.rotterdam.nl. Daarnaast worden de vergaderingen achteraf op basis van audio genotuleerd. Omdat er geen directe koppeling is van de neergeschreven woorden in de notulen met de uitgesproken woorden in de audio, kan niet effectief gezocht worden in het materiaal. Er kan immers alleen 'op raadsvergaderingen' als geheel gezocht worden, niet op specifieke fragmenten 'binnen' vergaderingen.

Om de ontsluiting van het materiaal mogelijk te maken is er behoefte aan een gebruiksvriendelijke toegang, intuïtieve interactie met het audiomateriaal en toegang tot contextinformatie, terwijl ook ontsluiting door middel van een specifieke zoekvraag mogelijk moet zijn.

Om dit te bereiken wil het Gemeentearchief Rotterdam gebruik maken van de technologie ontwikkeld door de Universiteit Twente in het CHORAL-project, en deze technologie verder ontwikkelen en verfijnen. CHORAL houdt zich bezig met de ontwikkeling van technologie waarmee Nederlandse audioarchieven rechtstreeks doorzoekbaar en toegankelijk gemaakt kunnen worden. Het gaat hierbij om de inzet van automatische spraakherkenning met als primair doel het geautomatiseerd vervaardigen van een doorzoekbare index met tijdlabelels.

Beoogd CATCH Plus-resultaat

De realisatie van een generieke tool op basis van open source voor het indexeren van de audio-opnames van (raads)vergaderingen, met gebruikmaking van spraaktechnologie, waarbij de notulen worden gebruikt om opnames te structureren en de kwaliteit van de indexen te verhogen.

De ontsluiting kan dan ook worden verdiept door de vergaderingen te koppelen aan beschikbare metadata (contextgegevens, namen, profielen en websites van sprekers, geografische eigennamen, etc).

Naast het zoekbaar maken is het verbeteren van de verstaanbaarheid van de vergaderingen een vaak gehoorde wens. Het project komt hieraan tegemoet door de verstaanbaarheid van de audio te ondersteunen door de vergadering te "ondertitelen" op basis van de tijdgesynchroniseerde notulen of eventueel spraakherkenningstrascripts.

Doelgroepen

Doelgroepen voor het gebruik van de geïndexeerde bestanden zijn:

- Raadsleden
- College van B&W
- Ambtenaren
- Burgers
- Journalisten
- Onderzoekers
- Overige geïnteresseerden

Doelgroepen voor gebruik van de applicatie:

- Gemeenteraden en andere publieke organen in Nederland

- Andere instellingen waar gebruikt gemaakt wordt van audio voor verslaglegging, zoals de (commissies van de) Eerste en Tweede Kamer

Impact van CATCH Plus

Door de audio en de indexering daarop via internet beschikbaar te stellen, kan het beschikbare materiaal snel voor een breed publiek ontsloten worden. Wanneer bovendien een koppeling met metadata plaatsvindt, zal het materiaal in een brede context aangeboden kunnen worden.

De raadpleegbaarheid via internet vergroot de transparantie van het openbaar bestuur en de betrokkenheid van burgers bij de politiek: het maakt het mogelijk om het gesprokene woordelijk terug te luisteren (wat mogelijk ook gevolgen kan hebben voor het optreden van politici in de Raad).

Op termijn zal de benodigde capaciteit voor ontsluiting van de vergaderverslagen voor de Griffie door de automatisering afnemen. Na overbrenging naar het gemeentearchief zal bespaard kunnen worden op de kosten van ontsluiting, omdat het materiaal dan al optimaal ontsloten is (ervaring leert dat voor 1 uur audio-ontsluiting 3 werkuren benodigd zijn).

Door de collectie raadsverslagen te ontsluiten doet het Gemeentearchief Rotterdam ervaring op met de ontsluiting van eenduidige, grote audiocollecties, wat goed van pas komt bij de verdere ontwikkeling van ASR (Automatic Speech Recognition) voor het heel diverse materiaal van Radio Rijnmond, dat als niet ontsloten collectie bij het Gemeentearchief berust.

Beheer en onderhoud

Het Gemeentearchief Rotterdam maakt voor het beheer en onderhoud van haar applicaties gebruik van het Shared Service Centre ICT van de Gemeente Rotterdam. De te ontwikkelen applicatie zal eveneens bij het Shared Service Centre ICT worden ondergebracht. Daarnaast zal met de software-ontwikkelaar een contract worden afgesloten om verdere ontwikkeling van de software mogelijk te maken. De software voor spraakherkenning toegespitst op de ontsluiting van de raadsvergaderingen zal worden ondergebracht bij het bedrijf Cross-Media Interactie (XMI).

Potentiële relevantie voor andere instellingen

De tool die wordt ontwikkeld is bij voorkeur generiek, dat wil zeggen breed toepasbaar, waardoor het voor andere instellingen mogelijk is gebruik te maken van dezelfde tool. De tool zal als open-sourcesoftware beschikbaar worden gesteld zodat ook andere geïnteresseerden hiervan gebruik kunnen maken wanneer zij beschikken over goede audiobestanden en notulen om de structuur in de indexen aan te kunnen brengen.

Samenhang met centrale diensten

Het koppelen van spraaksegmenten aan tekst (uit spraakherkenning of uit notulen), en vervolgens aan concepten uit een vocabulaire is in essentie een voorbeeld van gelaagde annotatie. Tool-onderdelen voor het maken of visualiseren van deze spraak- en tekst-annotaties, en de opslag ervan, worden in samenhang met de centrale annotation-dienst ontwikkeld. De concepten worden gebruikt door te refereren naar vocabulaires in de vocabulaire-dienst.

CATCH Plus voor Gemeentemuseum - contactpersoon: Vincent de Keijzer

Huidig CATCH-project: MUSEUM

Werktitel: Ontwikkeling van collaboratieve webomgeving <http://my.gemeentemuseum.nl/>

Achtergrond

In het kader van het MuSeUM-project is recent een collaboratieve omgeving, <http://my.gemeentemuseum.nl/>, gelanceerd om te experimenteren met nieuwe diensten voor informatietoegang, gebruikersgegenereerde content en nieuwe vormen van interactie tussen gebruikers en museumexperts. Verschillende MuSeUM-resultaten (zoals CATCHUp) worden binnen deze omgeving aangeboden. Het *Gemeentemuseum* brengt door het opzetten van deze omgeving een gesprek op gang tussen de eigen museummedewerkers en ontwerpers/ontwikkelaars van software.

Het is de intentie van het *Gemeentemuseum* om succesvolle elementen uit de experimentele omgeving te integreren met de bestaande informatiearchitectuur. Het gaat dan specifiek om het ontwikkelen van additionele modules (bijvoorbeeld koppelingen tussen bestaande systemen en de experimentele website, tools voor de back-office en erfgoedgerelateerde specifieke tools).

CATCH Plus is noodzakelijk om de tools uit de experimentele omgeving verder te ontwikkelen, en om succesvolle toepassingen in de officiële *Gemeentemuseum*-website te integreren, alsmede om deze ter beschikking te stellen van andere instellingen.

Beoogde CATCH Plus-resultaat

Het beoogde resultaat is de combinatie van een experimentele omgeving en een verzameling kant-en-klare tools voor het opzetten van een collaboratieve omgeving bij een erfgoedinstelling. De software die in het kader van het CATCH Plus wordt ontwikkeld komt vrij ter beschikking voor alle geïnteresseerde (erfgoed)instellingen. Het *Gemeentemuseum* stelt het ontwerp en inhoud van de experimentele website ter beschikking.

CATCH Plus moet ertoe leiden dat er een brede experimentele omgeving voor de erfgoedsector ontstaat en dat de prototypen met een beperkt bereik worden uitgewerkt tot modules en toepassingen die eenvoudig inzetbaar zijn voor meer dan één instelling.

Doelgroepen

Doelgroepen van de ontwikkelde tools zijn erfgoed geïnteresseerden in het algemeen, en specifieke doelgroepen (zoals bij erfgoededucatie aan leerlingen) in het bijzonder. Doelgroep van de experimentele omgeving is voornamelijk de erfgoedinstellingen. Zij kunnen in deze omgeving kennis nemen van theorie en praktijk van ICT-toepassingen in de erfgoedsector en partners vinden voor eventuele experimenten. Succesvolle experimenten die zijn uitgewerkt tot universele prototypes kunnen als basis worden gebruikt voor de eigen specifieke toepassingen.

Impact van CATCH Plus

Beoogde impact van CATCH Plus is een grotere betrokkenheid van eigen medewerkers als potentiële gebruikers bij ontwerp en ontwikkeling van IT-toepassingen en meer samenwerking met collega's uit andere erfgoedinstellingen. Een van de onderwerpen die discussie zal opleveren is de wenselijkheid van samenwerking en het op grote schaal beschikbaar stellen van de eigen informatiesystemen.

Het MuSeUM-resultaat CATCHUp bijvoorbeeld zorgt voor snelle en relevante resultaten van grote zoekacties maar ook voor het verlies van identiteit van de deelnemende instellingen. De eigen informatie is immers onherkenbaar vermengd met informatie uit andere instellingen. Door het opzetten van een "eigen" collaboratieve omgeving wordt de eigen identiteit

gewaarborgd bij de vertaling daarvan in de ICT-toepassingen. De virtuele presentie van eigen collecties is een belangrijk discussie die in elke deelnemende instelling moet worden gevoerd.

Beheer en onderhoud

Het *Gemeentemuseum* zal de succesvol geïntegreerde toepassingen uit de experimentele omgeving beheren en onderhouden. Elke andere instelling die gebruik zal maken van de CATCH Plus-resultaten is zelf verantwoordelijk voor de interne introductie en de integratie met de bestaande systemen.

Voor de beoogde samenwerking met andere instellingen is centrale sturing gewenst. Een verkenning naar een geschikte partij die het experimentele platform onderhoudt en verder ontwikkelt na afloop van CATCH Plus is noodzakelijk. Met name deelname aan discussies en ontwikkeling van ideeën kan plaatsvinden zonder financiële drempels. Het uitwerken van een succesvol experiment worden financiële bijdragen gevraagd aan de deelnemende instellingen.

Potentiële relevantie voor andere instellingen

De experimentele omgeving zelf is gebouwd op een open-source CMS (drupal) dat alle basisfunctionaliteiten geeft, en dus relatief weinig onderhoud behoeft. Het systeem is uitermate modulair, zodat de instellingsspecifieke modules (cross-walks tussen systemen) en erfgoedgerelateerde specifieke modules ook zeer eenvoudig zijn te onderhouden. De grootste inspanning van de deelnemende instelling zit in het aanleveren van de content et cetera (puur de gebruikskant).

Samenhang met centrale diensten

CatchUp is zeer gebaat bij de beschikbaarheid van een centrale index voor collecties van diverse deelnemende instellingen in CATCH Plus, zoals door werkpakket 3 mogelijk wordt gemaakt. CatchUp voegt daaraan de mogelijkheid toe geavanceerd en geünificeerd te zoeken. Deze zoekmogelijkheid wordt aangeboden in vorm van herbruikbare componenten en diensten.

CATCH Plus voor Instituut voor Beeld en Geluid - contactpersoon: Johan Oomen/Annemieke de Jong

Huidige CATCH-projecten: MUNCH en CHOICE

Werktitel: Documentalist Support Systeem

Achtergrond

De 'ContextDatabase' van Beeld en Geluid zorgt ervoor dat heterogene bronnen in de productieomgeving van radio- en televisieprogramma's integraal deel gaan uitmaken van de bestaande kennisinfrastructuur. Doel is het inrichten van een opslagplaats voor allerlei (typen) gegevens die een relatie hebben met de film-, video-, audio- en fotocollectie van Beeld en Geluid, bijvoorbeeld websites van radio en televisieprogramma's, ondertiteling voor slechthorenden, teletekstuitzendingen, persinformatie, presentatieteksten, logboeken, scripts, recensies en kijk- en luistercijfers.

Binnen het CATCH/CHOICE-project is een prototype voor een Documentalist Support Systeem (DSS) ontwikkeld. Naast geïntegreerd browsen en zoeken over deze ContextDatabase, Immix catalogus, essence, thesaurus en automatisch gegenereerde semantische annotaties, biedt dit systeem relevante, automatisch uit context-informatie afgeleide trefwoorden aan documentalisten. Het DSS biedt daarmee zicht op zowel productiviteitsverhoging bij het handmatig ontsluiten van de collectie, als op nieuwe zoekingen tot de collectie. Het ondersteunt daarmee de langetermijnambitie van Beeld en Geluid om 20% van de collectie handmatig te ontsluiten, en 80% automatisch.

DSS is van meet af aan ontwikkeld met het oog op toepasbaarheid op andere tekstgebaseerde collecties, en volgens een webgebaseerde architectuur, die gebruikt maakt van algemene vocabulaire- en annotation-services, zoals ook in het CATCH Plus voorstel worden gepresenteerd. Daarom is ook vanuit een aantal andere erfgoedinstellingen belangstelling getoond voor doorontwikkeling van DSS tot een volwaardige, in het productieproces inzetbare software-omgeving.

Beoogd CATCH Plus-resultaat

Parallel aan het ontwikkelen van de ContextDatabase door Beeld en Geluid (in samenwerking met andere onderzoeksprojecten, en in het kader van de binnen BenG te realiseren wetenschapsomgeving) wordt het Documentalist Support Systeem met bijbehorende services verder ontwikkeld ten behoeve van innovatieve exploitatie van deze ContextDatabase. Daarbij wordt van meet af aan rekening gehouden met de wensen van andere geïnteresseerde erfgoedpartijen, die belang hebben bij zowel DSS functionaliteit, als bij de door DSS gebruikte web services: met name ICN, Rijksmuseum, Koninklijke Bibliotheek en Gemeentemuseum Den Haag hebben hun belangstelling uitgesproken.

In het kader van CATCH Plus worden de volgende deliverables opgeleverd:

- Een modulaire, webgebaseerde en configureerbare documentalist support omgeving, gebaseerd op automatische semantische analyse van teksten. Deze omgeving moet geschikt zijn om in samenhang met de bestaande Beeld en Geluid software-infrastructuur in het productieproces ingezet te worden. Bovendien moet de omgeving, op maat geconfigureerd, in andere erfgoedinstellingen kunnen worden ingezet ter ondersteuning van documentatieprocessen.
- De omgeving leunt sterk op gebruik van de Beeld en Geluid thesaurus (GTAA). Deze wordt via webservices toegankelijk gemaakt door hem op te nemen in de centrale CATCH Plus repository of vocabularies. Daarbij wordt er zorg voor gedragen, dat updates van de GTAA te allen tijde meteen ook via de repository beschikbaar zijn. Daarmee is de GTAA ook voor andere belangstellende partijen online als rijke resource beschikbaar.

- Toegang tot de Beeld en Geluid catalogus door implementatie van het OAI-PMH protocol. De implicaties voor DSS zijn, dat collectie-toegang gestandaardiseerd wordt geregeld, en dat transparante, gelijktijdige toegang tot verschillende collecties bij verschillende erfgoedinstellingen gerealiseerd wordt.
- Deelname aan een samenwerkingsverband ter introductie van instituut overstijgende persistente identifiers voor collectie-objecten.

Doelgroepen

Documentalisten van Beeld en Geluid, en andere bijdragende erfgoedinstellingen.

Gebruikers van de collecties van Beeld en Geluid (museumbezoekers, online gebruikers)

Impact van CATCH Plus

De ContextDatabase is een waardevolle toegevoegde bronnenbank voor zowel professionele gebruikers van de Beeld en Geluid collectie als voor het brede publiek. Het in het kader van CATCH Plus te ontwikkelen Documentalist Support Systeem verleent de ContextDatabase op een aantal manieren nog extra meerwaarde.

- DSS gebruikt de ContextDatabase als bron voor het afleiden van aanbevelingen voor beschrijvende termen aan documentalisten. Daarmee wordt het aantal alternatieven teruggebracht van 170.000 termen tot hooguit enkele tientallen. Daardoor kan het handmatige annotatieproces aanzienlijk worden versneld.
- Contextdocumenten en daarmee geassocieerde radio- en televisieprogramma's worden automatisch voorzien van semantische annotaties. Daarmee worden ze ook vindbaar zonder tussenkomst van documentalisten.
- De automatische semantische annotaties zijn veel talrijker en minder specifiek dan de annotaties die door documentalisten worden toegekend. Daarmee wordt de mogelijkheid geopend om op nieuwe manieren te zoeken en browsen in de collectie.
- De Beeld en Geluid collectie wordt via de thesaurus automatisch gekoppeld aan andere kennisbronnen op het Semantisch Web. Daarmee ontstaat voor Beeld en Geluid en anderen de mogelijkheid automatisch externe informatie te verbinden met catalogusbeschrijvingen (bijvoorbeeld links naar Wikipedia pagina's, toegang via Google Maps, etc).

Beheer en onderhoud

De binnen het project ontwikkelde software wordt geïntegreerd in het bestaande iMMix informatiesysteem van Beeld en Geluid. iMMix bestaat momenteel uit Catalogus-, Asset Management- en Rechtenbeheersmodules. De sector Media Archief Diensten (waar de afdeling applicatiebeheer onder valt) zorgt voor het beheer van iMMix. Met de softwareontwikkelaars worden afspraken gemaakt over beheer; dit zal integraal onderdeel uitmaken van het op te stellen contract.

Potentiële relevantie voor andere instellingen

Alle erfgoedinstellingen die trefwoorden willen afleiden uit tekstdocumenten hebben profijt van de ontwikkelde software. Het Document Support Systeem werkt onafhankelijk van gebruikte vocabulaire en op ieder tekstcorpus. Verschillende partijen binnen het CATCHplus consortium hebben interesse getoond deze software module te willen gebruiken. ICN heeft reeds een concrete use case geformuleerd. De Beeld en Geluid thesaurus wordt via webservices toegankelijk gemaakt door hem op te nemen in de centrale CATCH Plus repository of vocabularies. Collectiehouders kunnen de GTAA gebruiken; waardoor de interoperabiliteit tussen collecties verbetert. Het Filmmuseum gebruikt de GTAA al, ook regionale en lokale

omroepen zullen de GTAA gebruiken. De thesaurus speelt in een belangrijke rol onderzoeksprogramma's. Zo is de GTAA geconverteerd naar de webstandaard RDF en er worden koppelingen gemaakt met andere lexicons en thesauri, zoals Wordnet en Iconclass. Ook binnen initiatieven als Beelden voor de Toekomst, Nederlands Erfgoed Digitaal! en Europeana is voor de GTAA een belangrijke rol weggelegd als ingang tot het Nederlands erfgoed.

Samenhang met centrale diensten

Het Documentalist Support Systeem maakt intensief gebruik van de centrale annotation-dienst (met name voor opslag en zoeken van semantische tekst-annotaties) en van de vocabulaire-dienst: teksten kunnen worden geannoteerd in termen van een thesaurus naar keuze. Bovendien worden thesauri gebruikt voor het bepalen van de meest relevante trefwoordsuggesties. Onderdelen van de tool worden als herbruikbare diensten aangeboden: een dienst om teksten te semantisch te annoteren aan de hand van een thesaurus naar keuze en een dienst om lijsten van trefwoorden te rangschikken naar relevantie.

CATCH Plus voor Koninklijke Bibliotheek - contactpersoon: Henk Matthezing/Paul Doorenbosch

Huidig CATCH-project: STITCH

Werktitel: Integratie thesaurusontsluiting

Achtergrond

Steeds grotere hoeveelheden Cultureel Erfgoed materiaal (tekst, beeld, audiovisueel materiaal) komen digitaal beschikbaar via het Internet. Dit materiaal is zeer divers van karakter en verspreid over een scala aan instellingen, die veelal instellings- of domeinspecifieke ontsluitingstechnieken hanteren.

Gecontroleerde vocabulaires (thesauri, ontologiën, classificatiesystemen, trefwoordensystemen e.d.) en de koppelingen daartussen worden in een digitale wereld steeds belangrijker hulpmiddelen om de eindgebruiker goed te kunnen helpen bij het doorzoeken van dit diverse en verspreide aanbod om zijn informatiebehoefte snel en efficiënt te kunnen bevredigen.

Beoogd CATCH Plus-resultaat

Het vereenvoudigen van de metadata-toekenning aan nieuwe titels en de aansluiting door semantische koppeling van oude metadatasystemen op de huidige.

Dit doel wordt bereikt door het realiseren van tools om thesauri en classificatiesystemen semantisch aan elkaar te koppelen, op te slaan, te beheren en te benutten voor geïntegreerde titelbeschrijving op basis van de onderzoeksresultaten uit het STITCH-project.

Doelgroepen

Titelbeschrijvers, catalogiseerders, beheerders van thesauri, classificatiesystemen en andere gecontroleerde vocabulaires, binnen en buiten KB.

Impact van CATCH Plus

Ontwikkeling vindt plaats op basis van de situatie bij de KB. Daarnaast wordt samenwerking gezocht met partners waarvoor dezelfde toepassing wordt aangepast, met als doel een generieke tool te ontwikkelen. De KB heeft twee hoofdcollecties (wetenschappelijke collectie en depotcollectie). Deze beide collecties worden nu met twee verschillende thesauri (resp. de GTT - Gemeenschappelijke Trefwoorden Thesaurus - en de Brinkman Thesaurus) ontsloten. Een deel van de titels komt in beide collecties voor. Beide systemen worden gescheiden onderhouden en door verschillende groepen titelbeschrijvers gebruikt bij vervaardiging van de boekbeschrijvingen. Beide collecties bevatten oude titels die niet met deze thesauri zijn beschreven, maar wel moeten worden geïntegreerd. De openbare bibliotheken titels beschrijven met behulp van de NBD-Biblion-thesaurus soms dezelfde titels in hetzelfde gemeenschappelijke catalogiseersysteem (OCLC) als de KB dat doet. Daarnaast is er behoefte aan een manier om de Nederlandse ontsluiting aan te sluiten op internationale systemen als (bv) Dewey Decimal Classification (bibliotheken), CIDOC-CRM, ATT, Iconclass (kunstinstituten). Met de te ontwikkelen tools kunnen de primaire processen voor zowel bij KB als bij andere bibliotheken en erfgoedinstellingen aanzienlijk efficiënter worden, maar bovendien wordt de semantische interoperabiliteit van collecties en objecten aanzienlijk verbeterd, wat in het belang van de eindgebruiker is.

Voordelen:

- Efficiëntie in thesaurusonderhoud (slechts één ipv meerdere thesauri te onderhouden)
- Trefwoordtoekenning geschiedt met één systeem door een (samengevoegde) groep medewerkers
- Verbetering van zoekfunctionaliteit via koppeling van moderne vocabulaires met oude trefwoordensystemen.

Methode:

Het proces kent 3 fasen: pre-processing, hoofdproces (semantisch koppelen + opslag), post-processing

1. Pre-processing: Export van OCLC formats en conversie naar een uniform, machine-leesbaar formaat (SKOS of RDF:OWL)
2. Hoofdproces:
 - a. Aanmaak en kwaliteitsbeoordeling van koppelingen (alignments) tussen gecontroleerde vocabularies via semantische webtechnieken. Dit proces moet worden gefaciliteerd met een 'alignment tool', die de beheerder de mogelijkheid geeft verschillende semantische webtechnieken te combineren, en de beste(n) toe te passen.
 - b. Opslaan van vocabularies en koppelingen in een 'alignment repository'.
3. Postprocessing: Handmatig verbeteren van de koppelingen en opname daarvan in de 'repository'.
Annotatiesuggestie van termen uit het gekozen systeem voor titelbeschrijvers die gewend zijn te werken met het andere systeem. Automatische opname van de toegekende trefwoorden in gemeenschappelijk catalogiseersysteem OCLC.

Beheer en onderhoud

Beheer en onderhoud zijn afhankelijk van de keuze van de deelnemende erfgoedinstellingen om de CATCH plus tools in zijn geheel door een centrale instantie te laten beheren/onderhouden of dat men kiest voor in eigen beheer nemen van (deel)resultaten. Dit is een vraag die CATCH verband beantwoord zal moeten worden.

Gezien het aantal vocabularies in gebruik bij de KB is het wenselijk functioneel beheer heeft over een alignment tool, een eigen repository, die via internet transparant en naadloos aansluit op de centrale repository, en annotation- en vocabulary-correctietools.

Software-onderhoud cq -ontwikkeling van nieuwe versies dient belegd te worden bij een centrale instantie of een externe softwareontwikkelaar vanwege het generieke karakter en gebruik door andere instellingen.

Potentiële relevantie voor andere instellingen

Met de hier beschreven toolset kunnen cultureel erfgoedinstellingen hun vocabularies semi-automatisch koppelen, opslaan en beheren en via een annotation-tool koppelen aan de beschrijvende records van hun (digitale) collecties. Het biedt de mogelijkheid collecties cross-domain voor gebruikers toegankelijk te maken op objectniveau.

Voorwaarde is dat de toolset generiek is, aanpassingen aan de eigen werkprocessen moeten mogelijk zijn, maar koppelingsresultaten moeten eenduidig geformaliseerd zijn ten behoeve van het gedistribueerde gebruik door de overige instellingen.

Potentiële partners:

Tweede bibliotheekorganisatie, OCLC, Beeld en Geluid, (commerciële) software-ontwikkelaar met specifieke expertise op domein van semantische interoperabiliteit.

Samenhang met centrale diensten

Sommige onderdelen van de toolset zijn essentiële onderdelen van de centrale vocabulaire-dienst (bijvoorbeeld de alignment repository), andere voegen daar algemeen bruikbare diensten aan toe (concept alignment, SKOS conversie, thesaurus verrijking).

CATCH Plus voor Meertens Instituut - contactpersoon: Louis Grijp

Huidig CATCH-project: WITCHCRAFT

Werktitel: Interface-componenten ten behoeve van webgebaseerde zoekmachine voor melodieën

Achtergrond

In het WITCHCRAFT-project wordt een melodieënzoekmachine ontwikkeld die deel zal uitmaken van de Nederlandse Liederenbank (www.liederenbank.nl). Met deze webgebaseerde zoekmachine kunnen zowel onderzoekers als het grote publiek een groeiende verzameling gecodeerde melodieën doorzoeken door gebruik te maken van zuiver muzikale kenmerken. Tot nu toe is een aantal experimentele gereedschappen ontwikkeld voor de codering en het doorzoeken van de liederen, alsmede een aantal componenten voor de gebruikersinterface. Deze zijn echter nog niet robuust en gebruikersvriendelijk genoeg om ze na afloop van het WITCHCRAFT-project betrouwbaar en efficiënt te kunnen blijven inzetten. Doorontwikkeling van deze componenten is noodzakelijk, omdat de Liederenbank een essentieel gereedschap voor onderzoekers van het Meertens Instituut is, en omdat de ontsluiting van het muzikale erfgoed een taak van het Meertens Instituut vormt. Er zijn diverse andere muziekcollecties die nog niet beschikken over een 'muzikale' zoekmethode en zullen kunnen profiteren van de gereedschappen die in CATCH Plus doorontwikkeld worden.

Beoogd CATCH Plus-resultaat

Het beoogde resultaat is een suite van componenten waaruit een robuuste, effectieve en bruikbare webgebaseerde zoekmachine voor melodieën kan worden samengesteld, met bijbehorende gereedschappen voor dataproductie en –management. De zoekmachine wordt als een webservice opgezet die vanuit de Liederenbank—of een andere omgeving—wordt aangeroepen. De volgende componenten zullen worden uitontwikkeld.

1. Interface-componenten:
 - methoden om muzikale queries aan het zoekstelsel aan te bieden, zoals een piano-toetsenbord en een *Query By Humming* interface
 - methoden voor een heldere presentatie van de zoekresultaten, waarbij de gebruiker meerdere bekijk/beluister-opties heeft en gemakkelijk tussen deze presentatievormen kan switchen
 - methoden voor rank aggregation. De gebruiker moet, afhankelijk van de informatiebehoefte, uit verschillende combinaties van zoekalgoritmen kunnen kiezen; de resultaten van verschillende algoritmen moeten kunnen worden gecombineerd
2. Dataproductie en –management
 - de reeds ontwikkelde WITCHCRAFT-Editor moet verder ontwikkeld worden tot een robuuste en veelzijdige applicatie om muziekdata in te voeren en te corrigeren, die via het web in de Liederenbank of andere muziekdatabase geïntegreerd kan worden.
 - er moet een solide en betrouwbare workflow worden geïmplementeerd om de nog in te voeren (en de reeds ingevoerde data) te beheren, waarbij de koppeling met de records van de Nederlandse Liederenbank gegarandeerd is
 - om audiobestanden te ontsluiten is er software nodig voor de transcriptie van audio naar een noten-representatie. Het streven is hiervoor een aantal bestaande tools te configureren voor het materiaal in de Liederenbank.

3. Bestaande algoritmen
 - methoden voor het segmenteren en indexeren van grote hoeveelheden muzikaal materiaal
 - algoritmen om gelijkenismaten voor melodieën te berekenen

Doelgroepen

Onderzoekers van het Meertens en andere wetenschappelijke instellingen, en algemeen geïnteresseerden die gebruik maken van de Nederlandse Liederenbank. Daarnaast is belangstelling getoond door/ wordt belangstelling verwacht van de volgende instellingen.

Nederlandse instellingen die over digitale muziekcollecties beschikken:

- Beeld en Geluid
- Nederlands Muziek Instituut
- Centrale Discotheek Rotterdam
- Muziekwetenschap, Universiteit Utrecht (CMME-project)

Buitenlandse instellingen:

- Resonant (Vlaams centrum voor documentatie muzikale bronnen)
- RISM (Répertoire International des Sources Musicales)
- Volksmuziekprojecten in diverse landen

Instellingen en bestaande initiatieven die een muziekservice aanbieden als onderdeel van een breder pakket:

- Koninklijke Bibliotheek
- Geheugen van Nederland

Impact van CATCH Plus

Data invoer en data management is een essentiële taak voor verder gebruik en toegankelijk maken van verzamelde gegevens. De huidige provisorische methoden voor invoer en management van de muziekdata worden door CATCH Plus op het gewenste professionele niveau gebracht.

Voor musicologisch onderzoek naar volksliederen is tekstgebaseerde zoekfunctionaliteit niet voldoende. Met de doorontwikkeling van de interface componenten en zoekalgoritmen komt er een robuuste zoekfunctionaliteit beschikbaar ter ondersteuning van het onderzoek, met expert-opties.

De expert-opties maken het mogelijk om complexe zoekopdrachten uit te voeren, maar voor een publieksvriendelijke versie is een re-design van het user interface nodig. In het CATCH Plus project wordt een gemakkelijk toegankelijke zoekfunctionaliteit ten behoeve van het algemene publiek gebouwd.

Beheer en onderhoud

De ontwikkelde software wordt geïntegreerd in de bestaande software van de Liederenbank in het Meertens Instituut en tegelijkertijd daarmee onderhouden door de technische staf van het Meertens Instituut. Dit geldt ook voor de periode na het project; volgens het beleid geformuleerd in het Digitaliseringsplan van het Meertens Instituut.

Potentiële relevantie voor andere instellingen

Diverse instellingen en organisaties in binnen- en buitenland beschikken over aanzienlijke collecties van muzikaal materiaal (zie boven). Veelal heeft men geen mogelijkheid de muziek op muzikale kenmerken te doorzoeken, maar alleen via metadata. Wij stellen hun de resultaten van WITCHCRAFT, verzilverd via CATCH Plus, ter beschikking. De mate van digitalisering/codering verschilt nogal per collectie. Om van de CATCH Plus resultaten gebruik te kunnen maken dient het

materiaal in een symbolische codering beschikbaar te zijn, daar de gereedschappen niet op audio toegesneden zijn.

Expertise hiervoor dient voorhanden te zijn. De te doorzoeken muziek dient primair melodisch van aard te zijn. Aangezien het hier geen kant-en-klaarsysteem betreft dient er ook een software engineer beschikbaar te zijn om uit de componenten een werkend systeem samen te stellen. De benodigde investering zal aanzienlijk lager uitvallen dan wanneer een dergelijk systeem van de grond af aan opgebouwd wordt.

Belangrijke meerwaarde van het gebruik van deze tools is dat door het gebruik van een gemeenschappelijke infrastructuur en van webtechnologieën het doorzoeken van gedistribueerde collecties relatief eenvoudig te verwezenlijken valt.

Samenhang met centrale diensten

De componenten voor de melodieënzoekmachine zijn, hoewel specifiek voor muziekcollecties, herbruikbaar voor andere muziekcollecties binnen en buiten CATCH Plus. Om muziekcollecties uit te breiden of doorzoekbaar te maken zijn een muziek-editor en een audio transcriptie-tool nodig. Deze tools zijn annotation-tools en worden ontwikkeld in samenhang met andere annotation-tools en de centrale annotation-dienst. Voor deelcollecties is ook tekstmateriaal voorhanden, bijvoorbeeld liedteksten. In die gevallen is koppeling aan thesauri mogelijk door toepassen van herbruikbare diensten uit andere deelprojecten (Naturalis/RACM, Beeld en Geluid).

CATCH Plus voor Naturalis en RACM - contactpersonen: René Dekker en Guus Lange

Huidige CATCH-projecten: MITCH en RICH

Werktitel: Database organizer

Achtergrond

De RACM en Naturalis willen de CATCH-projecten RICH (RACM) en MITCH (Naturalis) verder samen laten gaan tot een zgn. "database organizer". Deze "database organizer" is een nieuw programma dat de software en ervaringen van twee CATCH-projecten (MITCH & RICH) combineert in één nieuw software programma. Het programma stelt in staat relevante informatie (wie, wat, waar) in ongestructureerde tekst in databases op te slaan. Eventuele fouten in de databases worden semi-automatisch verbeterd.

Binnen de cultureel erfgoed sector is er een grote behoefte aan dergelijke software. De goede resultaten uit de RICH en MITCH projecten tonen ook aan dat deze opties realistisch zijn – en dat de gezamenlijke ervaringen van Naturalis en RACM binnen CATCH gebruikt kunnen worden voor een oplossing.

De "database cleanup" software onder de naam Timpute is klaar om toegepast te worden. Hetzelfde geldt voor de chronologie-indexer en –retriever Open Boek van Rich.

Beoogd CATCH Plus-resultaat

De "database organizer" bestaat uit twee delen:

1. Het eerste deel, ontwikkeld door MITCH en Naturalis, behelst het "schonen" van databases m.b.v. referentiestructuren en het koppelen en inbouwen van ongestructureerde tekst. De te ontwikkelen programmatuur kan werkzaam gemaakt worden voor *open standaard* programma's. Voor bestaande (Nederlandse) collectieregistratie programma's (veelal 'gesloten' programma's) moet een extra vertaalslag gemaakt worden met de betreffende leveranciers. Naturalis zal de software inbouwen in haar collectieregistratieprogramma (in ontwikkeling).
2. Het tweede deel bestaat uit het volautomatisch herkennen en extraheren van bepaalde "Named Entities" uit vrije tekst, zoals vrije tekstvelden van databases, maar ook en vooral uit rapporten, papers en andere vrijstaande ongestructureerde teksten. Deze "Named Entities", waaronder bijv. chronologische data of geografische locaties, kunnen dan als attribuut worden toegevoegd aan bestaande databases. Ze kunnen echter ook dienen als indexen voor "free text information retrieval". Voor beide toepassingen is de software gereed onder de naam "Open Boek".

De RACM zal het programma onder andere gebruiken voor het duurzaam toegankelijk maken van de jaarlijkse stroom van archeologische rapporten (circa 4000 per jaar). Tevens zal het programma worden gebruikt om de relevante informatie uit oude handboeken, monografieën, inventarissen etc. digitaal toegankelijk te maken.

Doelgroepen

De programmatuur²⁸ wordt gebruikt door:

1. RACM en Naturalis zelf;
2. Professionele organisaties die omvangrijke verzamelingen ongestructureerde tekst beheren en/of die grote databases beheren, en waarin veel velden verwijzen naar gestandaardiseerde gegevenssets;
3. Consumenten met databases van privé verzamelingen (muziek, boeken, verenigingen).

²⁸ In de praktijk worden er waarschijnlijk verschillende versies (interfaces) ontwikkeld voor de verschillende doelgroepen.

De programmatuur wordt aangeboden aan de partners binnen de domeinen van de RACM en Naturalis en de CATCH partners (DANS, Beeld en Geluid, musea, bibliotheken, archieven en andere collectiebeheerders, zoals ICN, Meertens Instituut etc.). Wij menen dat naast de erfgoedsector de programmatuur tal van toepassingsgebieden heeft, zoals jurisprudentie, recherche, wet- en regelgeving, corporate archives, wetenschappelijke archieven, overheidsarchieven, literatuurwetenschap, etc.

Impact van CATCH Plus

De programmatuur van MITCH richt zich op databases die gebruik maken van referentielijsten en thesauri. De thesauri worden gebruikt als structuren om de **databasevelden semi-automatisch te controleren en verbeteren** (d.w.z. de expert ondersteunen met suggesties). Er is sprake van tweerichtingsverkeer; de thesauri worden ook aangevuld en verbeterd. De door RICH gebouwde programmas (Open Boek) daarentegen zijn in staat gegevens uit vrije tekstvelden te vertalen naar meer gespecificeerde velden in een record. Daarbij spelen opnieuw de thesauri en referentielijsten een hoofdrol. Een uitgebreide versie van de tweede module is in staat vrije tekstbestanden te koppelen aan thesauri en/of databaserecords op basis van combinaties van gegevens, afgeleid uit de bestaande database. In het verlengde is een derde module, na training en met behulp van thesauri, in staat uit tekstbestanden beginnende databases samen te stellen binnen nauw afgegrensde kennisdomeinen.

Beheer en onderhoud

Naturalis en RACM gaan er van uit dat de vertaling van hun resultaten in MITCH/RICH-verband in generieke programmatuur wordt uitgevoerd door een gespecialiseerd software-ontwikkelbedrijf. De ontwikkeling en het beheer van programmatuur hoort niet tot de competenties (en daarom tot de kernactiviteiten) van de erfgoedorganisaties. Projectleiderschap van IT-projecten die door derden worden uitgevoerd horen niet tot de taken van RACM/Naturalis werknemers. Wel kunnen zij helpen bij de kennisoverdracht van CATCH-expertise naar nieuw applicaties en de domeinkennis in te brengen bij verdere ontwikkelingen.

Potentiële relevantie voor andere instellingen

Het goed toegankelijk maken van de digitale bibliotheek en het archief (gedigitaliseerde en geOCRde teksten, databases) is voor iedere organisatie van groot belang, maar is voor erfgoedinstanties van vitaal belang. De software stelt in staat op betrouwbare wijze met een minimale menselijke inbreng de vragen over wie, wat en waar uit de digitale bibliotheekbestanden op te slaan of *on the fly* aan te leveren aan specialisten en andere gebruikers van de informatie. De software wordt modulair opgebouwd, op basis van open-sourcetechnieken. Voor professionele organisaties gaat het om het aanleveren van flexibele toolkits voor intern aanwezige datamanagers. Voor consumenten wordt het programma nauwer afgebakend, in een meer *'plug & play'*-format.

Samenhang met centrale diensten

De centrale onderdelen van de te ontwikkelen programmatuur (de database cleanup module en named-entity retriever en -indexer) worden aangeboden als diensten. Ze kunnen daarmee worden hergebruikt in andere tools, bijvoorbeeld als alternatieve annotation-modules in het Documentalist Support Systeem van Beeld en Geluid. Beide modules kunnen gebruik maken van de vocabulaire-dienst, met name Open Boek kan daarnaast gebruik maken van de centrale annotation-dienst voor opslag van de gevonden named-entities. Bovendien kunnen vocabulaire-verrijkingen worden aangereikt voor opslag in de centrale vocabulaires.

CATCH Plus voor Nationaal Archief - contactpersoon: Hennie van Schie

Huidig CATCH-project: SCRATCH

Werktitel: Webzoekmachine voor handgeschreven, gestructureerde teksten

Achtergrond

Het SCRATCH-project wordt uitgevoerd door de afdeling Kunstmatige Intelligentie van de Rijksuniversiteit Groningen, in samenwerking met het Nationaal Archief. In de huidige vorm is SCRATCH vooral een wetenschappelijk project. Wat betreft het doel –het kunnen googelen in gestructureerd handgeschreven materiaal—is aangetoond dat het realiseerbaar is. Daarbij is gebruik gemaakt van één collectie, namelijk de indices van het archief van het Kabinet der Koningin. Scratch4All, de beoogde voortzetting van SCRATCH-project, is bedoeld om de behaalde resultaten breder inzetbaar te maken, en om de software gebruikersvriendelijker te maken.

Een van de doelstellingen is om de bestaande procedures geschikter te maken voor het hanteren van verschillende collecties. De software zal elk handschrift namelijk afzonderlijk moeten aanleren. Daartoe dient de annotateerfunctie. Een klein deel van nieuwe handgeschreven tekst zal gewoon moeten worden overgetypt. Veel tekstcorpora maken feitelijk gebruik van een relatief beperkte set aan woorden, welke afwijkt van de tegenwoordige standaardtaal. Ten einde het aanleren te vergemakkelijken, zal van kenmerkende teksten en specifieke woordenlijsten gebruik gemaakt gaan worden. De vaste structuur van een tekst zal worden bepaald, het welk de herkenning van handgeschreven patronen zal vergemakkelijken.

De andere doelstelling van Scratch4All is het geschikter maken van de software voor de niet-deskundige eindgebruiker, en om beheersfunctionaliteit toe te voegen.

Beoogd CATCH Plus-resultaat

De realisatie van een web applicatie waarin beheerders van de documenten deze kunnen uploaden, bewerken, annoteren, leggen van relaties, beheren en doorzoeken. Ook standaardteksten en woordenlijsten, welke dienen ter referentie, moeten ingevoerd en onderhouden kunnen worden. Daarnaast dient een generieke applicatie voor het zoeken en vinden van woorden in de digitale teksten te worden gemaakt, welke beschikbaar is voor het algemene publiek.

Het project valt in de verschillende fasen uiteen, welke een of meer producten zullen opleveren.

1. In de eerste fase wordt de functionele eisen van de applicatie geformuleerd. Daarbij dient rekening gehouden te worden met het feit dat gebruik gemaakt zal moeten worden van externe, zeer krachtige supercomputers (Target cluster computer en de IBM Blue Gene van de RuG). Product: functioneel ontwerp.
2. De tweede fase bestaat uit de bouw van de applicatie op basis van het functioneel ontwerp. Product: alfa-release.
3. De derde fase is het testen en aanpassen van de applicatie, in verschillende omgevingen. Product: geaccepteerde versie 1.0 van de applicatie.
4. Het vierde is het schrijven en testen van de handleiding voor de applicatiebeheerder, de gegevensbeheerder, de annotateerder en de raadpleger (publiek). Producten: handleidingen.
5. De vijfde fase is het ontwikkelen van een standaardcontract voor de erfgoedinstelling met de beheerders van Blue Gene en - eventueel - de beheerders van dataopslagfaciliteiten. Product: een of meer standaardcontracten.

Parallel aan deze ontwikkelingsfasen zal er een traject van voorlichting en opleiding worden ontwikkeld en uitgevoerd.

Producten: presentaties en gebruikerscursussen.

Doelgroepen

De doelgroep bestaat uit (erfgoed)instellingen die in het bezit zijn van uitgebreid handgeschreven materiaal. Die teksten moeten aan drie voorwaarden voldoen:

- Het moet gestructureerd zijn qua lay-out,
- Het moet een zekere uniformiteit bevatten qua handschrift,
- Het moet in digitale vorm van hoge kwaliteit beschikbaar zijn.

Impact van CATCH Plus

Scratch4All is een extra tool voor het ontsluiten van erfgoed materiaal. Binnen het Nationaal Archief wordt het mogelijk chronologisch geordende, handgeschreven archieven te digitaliseren, en een fractie van de tijd en kosten te besteden aan het toekennen van metadata voor de ontsluiting ervan. Op lange termijn levert dat een kostenbesparing op, terwijl bovendien de kosten van het fysiek beschikbaar stellen zullen teruglopen, doordat de originele stukken minder zullen worden aangevraagd.

Als die tool niet beschikbaar komt, dan zullen de resultaten van huidige SCRATCH-project uitsluitend beschikbaar zijn voor de beperkte test set waarmee SCRATCH is ontwikkeld. De wetenschappelijk resultaten zijn behaald, maar de effecten voor het cultureel erfgoed zijn verwaarloosbaar.

Beheer en onderhoud

Voor het beheer en het onderhoud zijn er twee opties:

- 1) Een van de erfgoedinstellingen neemt de applicatie in beheer.
- 2) De applicatie wordt zoveel mogelijk als open source ontwikkeld, waardoor iedere instelling de applicatie aan eigen wensen kan aanpassen, bijv. binnen SourceFourge.net.

Voor beide opties geldt: De instellingen moeten bereid zijn in hun jaarbegroting een nu nog onbekende hoeveelheid tijd en geld te reserveren voor regulier beheer en onderhoud. Het NA doet dat door het jaarlijks afschrijven van een percentage van de ontwikkelkosten.

Potentiële relevantie voor andere instellingen

Binnen CATCH Plus is Scratch4All relevant voor alle erfgoedinstellingen met een of meer informatiebestanden die gestructureerd zijn, bijvoorbeeld een kaartstelsel, een catalogus, kadastrale registers, en die over een langere periode in één handschrift zijn bijgehouden.

Instellingen moeten bereid zijn een standaardcontract te ondertekenen met de beheerders van de Blue Gene, resp. Target computer, en bereid zijn tijd te besteden aan het aanleren van het handschrift. Tenslotte moeten de bestanden digitaal beschikbaar zijn.

Samenhang met centrale diensten

In essentie stelt de tool gebruikers in staat te zoeken naar vergelijkbare stukjes handschrift zonder tussenkomst van tekstuele beschrijvingen, en levert daarmee een dienst, die specifiek is voor handschriftcollecties. Echter, essentieel is dat voor iedere collectie eerst een substantieel deel handmatig moet worden getranscribeerd. De daarvoor benodigde webgebaseerde annotation-tool kan uitstekend in samenhang met andere CATCH Plus annotation-tools en -diensten worden ontwikkeld.

CATCH Plus voor het Rijksmuseum - contactpersoon: Peter Gorgels

Huidig project: CHIP

Werktitel: Gebruikersprofielen, aanbevelingen en annotaties in een museale publieksomgeving

Achtergrond

“*Gevangen in CATCH?*” Neen, voor het Rijksmuseum is de vertaalslag van de onderzoeksresultaten uit CATCH-onderzoeksprogramma naar bruikbare softwareapplicaties in CATCH Plus, een logische en belangrijke vervolgstap. De resultaten van het CHIP-project (VU & Rijksmuseum) lenen zich goed voor daadwerkelijke implementatie en gebruik van de gepersonaliseerde tours en onderhoud van gebruikersprofielen in de museale infrastructuur en als nieuwe dienst naar het publiek toe. Het Rijksmuseum streeft ernaar om de bezoeker meer en meer zelfontplooiër en ontdekker te laten zijn. Bezoekers hebben daarbij een gereedschapskist nodig om in het Rijksmuseum zelfstandig ‘aan de slag te kunnen’ en om zelf te leren: CATCH Plus vult die kist, met één van de eerste gereedschappen.

Het Rijksmuseum wil tevens de toepassing van een deelresultaat uit een ander CATCH-project, de webservice voor annotatie van objecten uit het CHOICE-project, uitwerken in een prototype en zo mogelijk inpassen in de huidige infrastructuur van de collectiecatalogus en de webpresentaties/xml-tussenlaag.

Kortom, het Rijksmuseum bouwt voort op de onderzoeksresultaten vanuit het CHIP-project, vanuit breder verband met het CHOICE-project en de E-Culture web service voor zoeken en browsen als verbindende factor.

Beoogd CATCH Plus-resultaat

In CATCH Plus gaat het Rijksmuseum gefaseerd werken aan de implementatie en integratie van een aantal innovatieve componenten ontwikkeld zijn binnen het CHIP- en CHOICE-project, en het MultimediaN E-culture project. Met deze componenten wordt beoogd om binnen de infrastructuur van het Rijksmuseum de onderstaande functionaliteiten in te brengen:

1. Onderhouden van gebruikersprofielen;
2. Aanbieden van webservice voor aanbevelingen/zoeken van objecten;
3. Aanbieden van webservice voor annotatie van objecten voor Rijksmuseumprofessionals.

Met het gebruik van deze componenten worden drie voorbeeldapplicaties als nieuwe diensten ontwikkeld:

1. de realisatie van een *gepersonaliseerde Rijks Widget*: de widget is een van de meest succesvolle marketingproducten voor het on-line publiek van het Rijksmuseum;
2. de realisatie van een *sociale webgebaseerde annotation-omgeving* (beta-versie) geïntegreerd met de bestaande infrastructuur (XML-tussenlaag en collectie-informatiesysteem), en waarin functionele en technische specificaties zichtbaar worden
3. de realisatie van gepersonaliseerde museumtours (als onderdeel van het Virtuele Nieuwe Rijksmuseum) zowel on-line als mobiel.

Doelgroepen

Het bereik van de nieuwe diensten richt zich vooral op kunst geïnteresseerden/liefhebbers, (potentiële) herhaalbezoekers en nieuwe doelgroepen die een ‘digital lifestyle’ hebben.

De web-annotation-omgeving is bedoeld voor Rijksmuseumprofessionals (registratoren, catalogiseerders, conservatoren, etc), maar is op termijn uitbreidbaar voor kunstliefhebbers en geïnteresseerde amateurs.

Impact van CATCH Plus

De twee hoofddoelen wat betreft de impact van CATCH Plus binnen het Rijksmuseum zijn:

1. *Toevoegen van een businessmodel voor het gebruik van bezoekersprofielen.*

Het Rijksmuseum wil het mogelijk maken dat mensen, op basis van een eigen belang (motivatie, relatie, relevantie), 'iets' kunnen doen met onderdelen van het Rijksmuseum. Men mag het 'gebruiken' naar eigen believen: de bezoeker wordt meer en meer zelf-ontplooiër en ontdekker.

Wanneer het Rijksmuseum haar rol van betekenis meer laat invullen door het publiek, zal deze betekenis veranderen, wellicht breder worden, meer uiteenlopend zijn, en misschien tegenstrijdigheden vertonen. Het driehoeksgesprek (object, bezoekers, conservator) krijgt hiermee een andere wending. De 'rollen' die het Rijksmuseum speelt kunnen verschuiven: van 'bezitter' naar 'beheerder' van 'allesweter' naar 'onderzoeker' en van 'docent' naar 'facilitator'.

2. *Ontwikkelen van een businessmodel voor het ontsluiten van de collectie.*

Implementatie en integratie van de CHOICE webservice voor annotatie van objecten betekent heel concreet een grote sprong voorwaarts in uniforme en versnelde ontsluiting van de objectinformatie door catalogiseerders van het Rijksmuseum.

Beheer en onderhoud

Rijksmuseum opteert voor:

- *Kennis verankeren in de Rijksmuseumlijnorganisatie:* collectiemanagers, IT-ers en applicatiebeheerders zijn van begin tot eind deelnemers aan het project. Het project wordt daarmee 'meer' van de organisatie zelf.
- *Ontwikkeling vindt plaats met inzet van leveranciers* waarmee het Rijksmuseum al een aantal jaren werkt en die jaarlijks bijdragen leveren aan ontwikkeling, beheer en onderhoud van de XML-tussenlaag bijvoorbeeld.
- *Kosten voor beheer en onderhoud* worden deel van de jaarlijkse, structurele begroting van het Rijksmuseum.

Potentiële relevantie voor andere instellingen

Om er voor te kunnen zorgen dat ook andere instellingen maximaal profiteren of meeliften op het ontwikkeltraject van het CATCH Plusproject van het Rijksmuseum, zullen twee consortiumpartners, te weten het Instituut voor Beeld en Geluid en het Haags Gemeentemuseum gevraagd worden om deel uit te gaan maken van de nog te ontwikkelen projectorganisatie. Beeld en Geluid heeft hierop in eerste instantie al positief gereageerd. Zij kunnen mede inzichtelijk maken welke voorwaarden gelden vanuit hun infrastructuur om applicaties voor personalisatie en aanbeveling te kunnen inpassen. Tevens zal het Rijksmuseum gebruik maken van vocabulaires van andere instellingen en ook eigen vocabulaires via derden (zoals DEN) toegankelijk maken.

Samenhang met centrale diensten

Er wordt voor de semantische annotaties gebruik gemaakt van de centrale vocabulaire- en annotation-diensten. Daarnaast biedt dit project het aanmaken, opslaan, beheren en gebruiken van gebruikersprofielen als dienst aan andere erfgoedinstellingen binnen en buiten CATCH Plus.