Digital innovation in the heritage sector
Digital innovation in the heritage sector
Contents

CATCH en CATCHPlus 4-5
Foreword 6-7
Introduction 8-11
Facts & Figures 12-13
SCRATCH4All 14-17
Workspaces 18-19
CHoralPlus 20-23
WitchcraftPlus 24-27
MuSeUMPlus 28-31
ZieOok en UPR 32-35
Multiply 36-39
STITCHPlus 40-41
OpenSKOS 42-43
DocumentChecker 44-45
Annotation Repository 46-47
Persistent Identifiers 48-49
Towards a heritage broker 50-53
Organisation 54-55
More information 56-57
Cultural heritage is everywhere, and constitutes our collective memory. CATCH carries out research and develops scientific methods and techniques at the crossroads of the humanities and information technology. The results will make heritage collections more accessible to the public, professionals and scientists. The Netherlands Organisation for Scientific Research (NWO) funds the CATCH programme in which the NWO Divisions Physical Sciences and Humanities collaborate.

Innovation, collaboration and transferability are key concepts in CATCH. A question from the heritage sector (in later projects also from the humanities) underlies each project. Research teams made up of a PhD student, a postdoc and a research programmer operate within the walls of a cultural heritage institution (‘embedded research’).

Eighteen projects have been funded since 2004.

As well as scientific output, the research projects from CATCH resulted in pilots and prototypes. CATCHPlus was established to facilitate the implementation of these pilots and prototypes in the heritage sector. CATCHPlus received funding from the interdepartmental Programme Implementation Agenda ICT policy (PRIMA), the Ministry of Education, Culture and Science and the Netherlands Organisation for Scientific Research (NWO). Besides the further development of pilots, heritage-wide services were realised and activities were deployed to develop parts of the infrastructure needed for digital heritage. All of these activities contributed to more collaboration and greater coherency in the heritage sector. This made it easier to open up the collections and make these more accessible.

CATCHPlus ran from 2009 to 2012.
Foreword

Hans Bennis
Director Meertens Institute (KNAW)
Chair Supervisory Board CATCHPlus

You are now holding an overview of the results from CATCHPlus. This booklet is a catalogue of information about new software and services developed for the heritage sector. The foundation for these applications was laid in the NWO research programme CATCH, in which scientists from computer science and the humanities collaborated with staff of various cultural heritage institutes. Within CATCHPlus the highly promising demos of CATCH were further developed into end products that can be widely used in the Dutch heritage sector.

CATCH is the acronym for Continuous Access To Cultural Heritage. Much of our cultural heritage has already been digitised or described in databases. In theory, you can therefore search through this wealth of data and find previously invisible relationships. Unfortunately, this option has proven to be rather limited in practice due to the use of different standards and the limitations of existing technologies. The software and services developed within CATCHPlus will change that for heritage institution employees as well as visitors to these institutions.

So I hope you will be surprised and inspired by the applications described in this booklet. Each application has a specific description, sometimes supplemented with the personal experience of a user from the heritage sector. You will then find a link to a website where you can obtain further information about the application concerned.

CATCHPlus has made a start on the joint opening up of cultural heritage for different target groups. We are pleased to announce that this mission will be continued, for example, within the Cultural Heritage network of the innovation platform Creative Industry. A new heritage broker will also be appointed who will act as a link between the world of cultural heritage and the development of new technology. As a result of this you will soon be able to explore part of our heritage in an innovative manner.

I hope that you will enjoy reading this booklet and that it will enable you to embark upon a memorable journey through the fascinating world of digital heritage.
Introduction

Paul Doorenbosch
National Library of the Netherlands
Chair CATCHPlus Steering Group
Vice-chair executive board CATCH

Patricia Alkhoven
Meertens Institute
Project leader CATCHPlus

The CATCH phenomenon

IT has played a key role in heritage management and access for more than 0 years. Digitisation, management databases, websites, games, et cetera, have become part and parcel of the heritage process. Initially, these activities were often realised by heritage managers who enjoyed using computers (‘in the land of the blind …’) or by outsourcing the activities to external companies. Fantastic results were achieved with both approaches, especially where the IT applications were very similar to the traditional processes.

However, as the quantity of digital data grew people started to realise that changes in the heritage chain (from analogue to digital) might require the use of new methodologies and knowledge. At the start of this millennium, plans were developed to use knowledge and methods from IT in the heritage sector. This was expected to lead to new possibilities for both parties. An initial plan submitted to the Economic Structure Enhancing Fund failed to obtain funding. However, NWO’s Divisions for the Physical Sciences and the Humanities saw the enormous potential of this collaboration. The original plan was rewritten to give it a more scientific perspective whilst ensuring that the questions from the heritage sector still formed the basis of the programme.

The programme CATCH (Continuous Access to Cultural Heritage) was launched in 2004. It started with six projects that were defined in the programme and a further four projects granted funding in the open competition were subsequently added. Since then two more funding rounds have been held and each of these led to four more projects.

CATCH was different not only because the projects were based on a question from the heritage sector, or later from the humanities, but also because use was made of embedded research teams. The presence of researchers in the heritage institutions facilitated an informal exchange of knowledge and also led to an increased awareness of the possibilities provided by IT.

The first ten CATCH projects have been completed and were a major scientific success. They resulted in PhD theses and
a long list of top scientific articles. But they also delivered pilot applications, algorithms, databases, standards and techniques that were new for the heritage sector. Despite this newness, it was clear to all involved that they could help the heritage sector to develop further. NWO and the government fund PRIMA saw possibilities to valorise the societal relevance of the research (NWO) and to further build the infrastructure needed for digital heritage (the government). These two aspects added to the wish of the heritage institutions to convert the pilot applications into practical tools resulted in the follow-up project CATCHPlus. The start capital was the heritage of the scientific projects, several infrastructure wishes of the Ministry of Education, Culture and Science, money from the PRIMA fund, money and support from NWO and the in-kind efforts of staff from the heritage institutions. It was then 2009.

Two organisations played a major role in the organisation of the CATCHPlus project: during the first year that was the Netherlands Institute of Sound and Vision and during the rest of the project the Meertens Institute. They accommodated the project office and assumed responsibility for the project on behalf of the steering group. And the results achieved were excellent.

With one small exception all of the projects were completed. Sometimes the objectives were adjusted and in one case a project received a different ‘owner’. Along the way several new heritage institutions became involved as well. Each project had its own hurdles to face. The initial problems caused by the complex nature of the funding were quickly overcome and the projects then enjoyed two glorious years. The results are a range of applications, sometimes small and limited to a single institution and sometimes large and generic and appropriate to the entire heritage sector. Significant advances have been made with the introduction of the persistent identifier infrastructure. A central repository for vocabularies has been created and there is an annotation room. A start has been made on appointing a broker who must facilitate the sustainability of applications. Yet most of all a wide range of interesting tools has been developed for the heritage sector.

Fortunately, however, more has been achieved than just concrete project results. The marriage between the heritage sector and IT has proved to be a golden opportunity from both a business and a people perspective. The heritage sector has gained knowledge about the possibilities offered by IT and IT has gained knowledge of the unfathomable richness of data and semantics in the heritage sector. Equipped with their new knowledge people have expanded their professional networks and have often done so outside of their own disciplines. Sometimes institutions even modified their organisation, once they became aware of the new opportunities.

The synergy of heritage, IT and, to an increasing extent, the humanities has also contributed to the development of the digital humanities. There is an emerging awareness that combining the strengths of people and disciplines puts you in a far better position to realise innovations and innovative research. Moreover, collaboration between apparently very different disciplines is often very fruitful.

Yet people continue to be important too. All of the people involved in the CATCHplus projects are named in this booklet. We must not forget, however, that all of this has been achieved thanks to a small group of people who met in a motel near Den Bosch in 2002 and managed to break through the impasse between cultural heritage and the different parties in IT. They wrote the initiative for the CATCH programme: Alice Dijkstra, Paul Doorenbosch, An-nemarie Bos, Elco Bruinsma, Jaap van den Herik, Frank van Harmelen, Mark Kas, Martin Kersten, Peter Sigmund and Jos Taekema. We are indebted to them.

CATCH has a past, but more importantly CATCH has a future. Not necessarily as a programme but certainly through the people who have participated and through the material and immaterial results that will find their way into the collaboration between the heritage sector, IT and the humanities to realise an increasingly more effective and worthwhile use of our cultural heritage.

October 2012
Facts & Figures

CATCHPlus is the acronym for Continuous Access To Cultural Heritage Plus.

CATCHPlus is an elaboration of the current NWO research programme CATCH.

CATCHPlus ran from 2009 to 2012.

In CATCHPlus various CATCH pilots and prototypes were further developed into more widely usable applications.

CATCHPlus has resulted in nine different applications for heritage institutions and five services for the entire heritage sector.

From November 2010, the project office (led by Patricia Alkhoven) was located at the Meertens Institute. Hennie Brugman was the technical coordinator and monitored the development of the applications.

The project office worked under the supervision of a Steering Group (led by Paul Doorenbosch) and a Supervisory Board (led by Hans Bennis).

A Supervisory Committee (led by Marco de Niet) supervised progress within the projects. They made site visits to the various heritage institutions involved.

Eleven heritage institutions were involved in CATCHPlus: Amsterdam Museum, Rijksmuseum, Meertens Institute, Naturalis Biodiversity Center, Gemeentemuseum Den Haag, National Library of the Netherlands, National Archives of the Netherlands, Netherlands Institute for Sound and Vision, Cultural Heritage Agency of the Netherlands, Netherlands Theatre Institute, Municipal Archive of Rotterdam.

Six knowledge institutions were also involved: DEN Foundation, University of Groningen, University of Amsterdam, VU University Amsterdam, Tilburg University, University of Twente.

The total project costs were 3.1 million euros.

The financiers of CATCHPlus were:
- Programme Implementation ICT Agenda (PRIMA), Ministry of Economic Affairs, Agriculture and Innovation
- Ministry of Education, Culture and Science
- Netherlands Organisation for Scientific Research (NWO)
- The heritage institutions involved
- Knowledge institutions
SCRATCH4ALL
SCRept Analysis Tools for the Cultural Heritage

Digital Monk deciphers digitised manuscripts

Over the past few years heritage institutions have invested a lot of time and money in digitising their archives. Printed texts can be searched relatively easily with the help of optical character recognition, but for handwritten texts that is almost impossible. SCRATCH4All has changed that.

A computer cannot search through digitised texts as it sees an image but fails to realise that this contains letters. Optical character recognition (OCR) can be used to convert the image into letters but this technique only works for printed texts. The results of OCR are certainly not perfect but they are usually good enough to allow a search application to search through the texts.

Handwritten texts contain too much variation to be recognised using OCR. No two people write in exactly the same way and even within the texts of a single author, there can be variation in how any given letter is written. Scientists from the University of Groningen have therefore developed Monk, software that can make handwritten texts searchable. In the CATCH project SCRATCH, a user interface for Monk was developed which allows the National Archive of the Netherlands to search through digitised manuscripts.

The National Archive of the Netherlands is using Monk to make the archive of the Queen’s Office searchable. The indexes of this archive were written by the same person for a long period of time. That is very important because Monk is trained for one person’s handwriting. The system must see the same word five times before it can recognise it the sixth time. A large database of the same person’s handwriting is therefore vital for a good result. Complete automatic recognition of manuscripts is not possible with the software developed in SCRATCH4All. However, the archives will be made searchable enough for a person to use the results obtained to search further in a targeted and efficient manner.

Both the professional and amateur user of SCRATCH4All can further process the results from Monk via Workspaces (see following pages). Monk’s results can be improved by manually annotating the digitised manuscripts.
Henny van Schie, National Archive of the Netherlands: “Within the National Archive of the Netherlands, Monk is being used to open up the handwritten archive of the Queen’s Office. The indexes of this archive were written by the same person for many years and it is therefore a highly suitable candidate for Monk. SCRATCH4All has not directly changed my work as an archivist. An archivist is expected to provide certainty and completeness, and so if you cannot find something it definitely should not be there. You cannot say: ‘I’m sorry but I could not find it because my tools were unable to trace it.’ So I remain bound to the 19th century manual way of working with alphabetical files and indexes.

But for researchers, the use of Monk within the working environment of SCRATCH4All is definitely interesting. Once Monk has learned a person’s handwriting – the necessary work is required here – the researcher can access the archives far more directly. Then archives can be consulted that otherwise could not be inspected because the metadata provides too little certainty. Monk will not give 100% reliable results as it cannot guarantee to have recognised all of the words correctly. The users will still have to confirm or correct the results in the SCRATCH4All interface in order to help Monk learn the handwriting better. Nevertheless SCRATCH4All is a useful application within the National Archive of the Netherlands as it helps to open up chronological archives that are difficult to search. It gives a good indication of where you can find relevant information. In the longer term Monk will certainly make research quicker and easier.”

Project team SCRATCH4All
Marika Ceunen (Municipal Archive Leuven, Belgium)
Rolf Fokkens (Target Holding)
Yvette Hoitink (National Archive of the Netherlands)
Edzer Lawerman (Nipype)
Jean-Paul van Oosten MSC (University of Groningen)
Minne Oostra (de Ontwikkelfabriek)
Harry Romijn (Groninger Archieven)
Henny van Schie (National Archive of the Netherlands)
Lambert Schomaker (University of Groningen)
Petra Schoen (National Archive of the Netherlands)
Anco Westra (Nipype)
Michiel van Wijngaarden (Gelders Archief)
Rees Williams (University of Groningen)

Interested?
Are you interested in SCRATCH4All?
Visit www.catchplus.nl/scratch4all

Project team SCRATCH
Jacques Bogaarts (National Archive of the Netherlands)
Cathy Jager (National Archive of the Netherlands)
Fons van der Laan (University of Groningen)
John Nerbonne (University of Groningen)
Henny van Schie (National Archive of the Netherlands)
Lambert Schomaker (University of Groningen)
Tijn van der Zant (University of Groningen)
Sveta Zinger (University of Groningen)
Workspaces
Collaborating in a digital working environment

In a web-based world where different systems and users collaborate on a single task, data often needs to be stored temporarily. At one end the user must upload or enter data and at the other end check or further process the data. Workspaces offers a digital working environment in which different colleagues can carry out these activities together.

SCRATCH4All, presented on the previous pages, is a good example of this. The user has digital scans of a manuscript and uploads these to the application to analyse the manuscript. Subsequently the user can further process the result and if necessary correct it. Workspaces is an indispensable tool for the user in this case.

Workspaces offers even more advantages for heritage institutions. The digital working environment can provide access to different applications all conveniently arranged in the same system. Moreover, Workspaces can be flexibly arranged according to the wishes of the institution. Via the platform, different colleagues (from within and outside of the institution) can collaborate on collections that are still under development. The institution can control exactly which user or system can gain access to which information from the collection.

Workspaces is also interesting for the developers of applications for the digital heritage world. They can simply reuse the in-built modules for logging in, authorisation or uploading and connect these to the new applications they have developed themselves.

Project team
Hennie Brugman (Meertens Institute)
Rolf Fokkens (Target Holding)
Minne Oostra (Target Holding)

Interested?
Are you interested in the Workspaces? Visit www.catchplus.nl/workspaces
CHoralPlus
Access to oral history

Searching in speech

Digital text can be searched quite simply using ctrl+F. Searching for a fragment in a sound or video recording is a bit more difficult, however. Complete transcripts are not usually present and the information from the metadata is limited. Thanks to automatic speech recognition these files can now be searched as well.

Speech recognition can convert spoken Dutch into written text. By linking the transcribed text to the time coding on the recording, fragments at word level can be traced in the files. This technique has already proven itself for recordings made under ideal circumstances: no background noise and a single speaker who speaks clearly.

In the CATCH project CHoral researchers from the University of Twente investigated how audio files recorded under less ideal circumstances could be transcribed. Many archives have such recordings in their collections from non-professional speakers, several speakers per recording and/or recordings with background noise. Interviews or programmes from regional broadcasters are a case in point.

The software developed in CHoral works step by step. First of all it separates the speech from the background noise. Next it clusters fragments from the different speakers and tries to automatically recognise these. Finally the transcribed text is aligned with the recording so that it is clear exactly who says what and when.

Within CATCHPlus a web interface was developed to enable heritage institutions to use the speech recognition software developed. They can log into the application via their own account and upload files. The results of the speech recognition are subsequently returned to the user via e-mail.
Mies Langelaar, Municipal Archive of Rotterdam:
"Within the Municipal Archive the speech recognition tool was used for the disclosure of interviews with survivors of the bombing of Rotterdam. These interviews can be listened to on the Brandgrens website. In collaboration with RTV Rijnmond we also worked on opening up radio material from the period 1980 – 1986. This has a two-fold use: interested researchers can search through the material, but current programme makers at RTV Rijnmond can use the indexed material for reference purposes.

ChoralPlus has added value for the archive because video and sound material that could previously only be opened up by watching or listening to a lot of material, can now be indexed and made available in a relatively simple way. This allows the material to be searched through in a more targeted manner. The metadata are too limited to give a complete picture of the recording and so in the past relevant data for researchers and other interested parties was often left untouched.

The results of the speech recognition are not perfect, as there is too much variation between the different speakers and recordings. However, the results are good enough to make automatic searches in the recording possible. Furthermore, the transcription of ChoralPlus will become increasingly better as users confirm the outcomes or correct these where necessary.

The Municipal Archive of Rotterdam is very pleased with the functionality of ChoralPlus, as it helps to open up video and sound collections that were previously hardly accessible. The automatic transcriptions give a good indication about where you can find relevant information."

Interested?
Are you interested in ChoralPlus?
Visit www.catchplus.nl/choralplus

Project team
Marijn Huibregts (XMI Cross Media Interaction)
Franciska de Jong (University of Twente)
Mies Langelaar (Municipal Archive of Rotterdam)
Roeland Ordelman (XMI Cross Media Interaction)
Jantje Steenhuis (Municipal Archive of Rotterdam)

Project team Choral
Wilma van Giesbergen (Municipal Archive of Rotterdam)
Mies Langelaar (Municipal Archive of Rotterdam)
Willempijn Heeren (University of Twente)
Franciska de Jong (University of Twente)
Roeland Ordelman (University of Twente)
Jantje Steenhuis (Municipal Archive of Rotterdam)
Thijs Verschoor (University of Twente)
Laurens van der Werff (University of Twente)
Online searching in melodies

Folk songs are passed on from one generation to the next. Some songs are already more than five centuries old. Each new generation subtly changes the melody of these songs giving rise to different variants of the same melodies. A special melodies search engine can help to trace these variations.

Songs are part of the ‘immaterial heritage’. Around the world folk songs have been collected: in large databases with sound fragments as well as in textual descriptions of the songs. The famous Dutch song collection Onder de Groene Linde is housed at the Meertens Institute and is part of the Dutch Song Bank. This databank opens up the metadata of several song collections but had a handicap: the musical content of the songs could not be searched.

In the CATCH project WITCHCRAFT a search engine was developed that can determine the similarity of melodies. This makes it possible to identify an unknown melody or to identify different variants of the same song among thousands of other melodies. Moreover, the software developed helps researchers to gain a better understanding of how melodies might be related to each other.

In WITCHCRAFTplus this search engine was further developed and made suitable for use online. Researchers, musicians and other users can search for variations of melodies in the Dutch Song Bank, for example by uploading a MIDI file or by typing in a melody. In addition to this, an online music editor was built with which the collection specialists from the Meertens Institute and other users can enter played music or music visible on scans in searchable musical notation. This musical notation can be uploaded and downloaded in various formats so that as many users as possible, from researchers to musicians, can use and reuse the material.
Martine de Bruin, Meertens Institute:

“I suspect that most people have experienced at some time or other how difficult it is to name a song if you only remember the music and not the text. For researchers of song culture this was a daily recurring problem. New texts were often written for existing melodies, making it a challenge to find which melody was reused! For more than two centuries efforts were made to make melodies retrievable, for example by organising them in card index boxes. And if that worked – this is far more difficult than alphabetical indexing – there was still the major limitation that only the start of a melody could be used for searches! So ultimately you always had to depend on your own memory.

Now we have an important addition to the human memory. With the brand new melodies search engine, music can be searched through in many different ways and the first interesting results have already been obtained. Now at last we are no longer dependent on the text or context data.

Yet not all of our wishes have been fulfilled: the search engine has whetted our appetite for more in terms of both data and technology. Search engines are only successful if there is sufficient data to search in. For WITCHCRAFT we opted for the manual transcription of melodies; that provides the best results but it is very labour intensive. We transcribed more than 7500 melodies but we would like to significantly increase our corpus.

Also as a result of the WITCHCRAFT project we are thinking about other algorithms that could improve the search function. There are also plans to start a European collaborative project – melodies often occur in several countries and thanks to melody search engines we can now trace the migrations of melodies.”

WITCHCRAFT has also helped us in another way. In order to develop as good an algorithm as possible, researchers were asked which techniques they use to establish a similarity between melodies. These insights were incorporated and tested while developing the search engine. This not only led to an optimised algorithm but also an increased awareness among researchers as to what such a comparison entails.”
MuSeUMPlus
MULTiple-collecTion SEarching Using Metadata Plus

Improved searching with MuS and Geméén

A complete database is a utopia. Some data are always located in the wrong field, are inconsistent or contain textual or intrinsic errors. With MuSeUMPlus these ‘contaminated’ databases can be searched and the search results subsequently processed in an online working environment.

Most heritage institutions manage their collection in a structured, specialised database. These systems often have their own search forms with which specific terms can be sought within certain search fields (author, date, title, etc.). If the relevant information is not in the correct field, the search form will not find it. There is then a high chance that the user will miss out on relevant items in the collection.

In the CATCH project MuSeUM it was investigated how ‘contaminated’ databases could best be searched. Via the traditional, structured search methods or via a Google-like unstructured approach that includes all of the text present? The final conclusion was that both specialised users and typical users obtain the best results from a combination of both approaches.

In MuSeUM-Plus, the results from MuSeUM were developed further to usable applications. The search engine MuS makes it possible to search different databases simultaneously, both within the indicated fields as well as in the other text present. MuS also considers the interrelatedness of databases to a certain extent in its approach.

The heritage institution retains control over the presentation of the items found in the collection to ensure that external users cannot gain access to sensitive information. In the most conservative form only the PID (Persistent Identifier) codes are shown. Then the user only knows that the source probably contains relevant information and for further details he or she will have to approach the institution in person.

Besides MuS the project team has also developed Geméén. Geméén is an online working environment in which users in a workgroup context can manage and process the results from MuS, for example. The so-called O_og workgroups are examples of a MuS application in which only object descriptions are shown. With this relevant objects for a specific theme or a certain exhibition can be collected, for example. The search queries performed by the workgroup are continuously updated so that changes in the collection are directly visible in the search results.
Vincent de Keijzer, Gemeentemuseum Den Haag: “Imagine you are a lover of samplers, bookplates, model boats or Delft pottery. Or that you collect the early Hague School, works of artists from Arnhem or images of shoes. How do you know if something that suits your taste can be found within Dutch museum collections? You might think that searching with Google will deliver quite a bit of information, at least for a first impression. However, the opposite is true. Apparently our museum treasure-houses are strongly protected from the global finder Google. Then all you can do is write to the various museums and ask if you can see the collection.

Passionate collectors are not easily deterred by such high thresholds, but an increasing number of users refuse to put up with this situation. They consider the museum collections to be public property to which they want to have unhindered access. Museums refer to intrinsic, technical and organisational problems that hinder the large scale disclosure of information about the collection. However unjust it might be, many users see this as old-fashioned protectionism and want to judge for themselves whether the data in the museum systems are relevant for them or not. MuSeUMPlus lays the foundation for a solution that meets the interests of cautious museums and impatient users. Museums can provide rapid access to the information and retain control over what they do and do not want to present. Enthusiasts and collectors can search through all of the information, collect relevant objects and then process these as they wish. At the very least they can gain an answer to that question: where do I find samplers, bookplates, model boats, Delft pottery, the early Hague School, works of artists from Arnhem or images of shoes?”

Project team MuSeUMPlus
Hanno Lans (Datascape)
Frodo Schering (platform Geméén)
Jur de Vries (Triquanta)
Volkan Florchinger (Rhizom)
Vincent de Keijzer (Gemeentemuseum Den Haag), Marijn Koolen (University of Amsterdam)

Interested?
Are you interested in MuSeUMPlus?
Visit www.catchplus.nl/museumplus

Project team MuSeUM
Avi Arampatzis (University of Amsterdam)
Jaap Kamps (University of Amsterdam)
Marijn Koolen (University of Amsterdam)
Vincent de Keijzer (Gemeentemuseum Den Haag)
Nir Nussbaum (University of Amsterdam)
Maarten de Rijke (University of Amsterdam)
UPR en ZieOok

User Profile Repository and ZieOok

Recommendations based on a personal profile

A visitor who appreciates a painting of the Mint Tower, might also be interested in paintings of other townscapes of Amsterdam. The application ZieOok can give the visitor suggestions for other items in the collection based on his personal profile.

In the CATCH project CHIP (Cultural Heritage Information Presentation) scientists from Eindhoven University of Technology and Novay investigated how visitors to the Rijksmuseum could receive personal recommendations using semantic techniques. The demos from this project formed the basis within CATCHPlus for the development of new tools that can be used by the Amsterdam Museum and Netherlands Theatre Institute.

The User Profile Repository (UPR) stores personal profiles of users. Besides personal details, such profiles can also contain descriptions of and references to preferred objects. These preferred objects have previously been rated by the user (by adding them to the favourites). In addition to this the UPR records statistics about how often a user visits a given object.

Of course the user has access to his own profile. He can register for this and log in via his account on Facebook, Google or Twitter, for example. Then he can give his personal ratings for various items in the collection on the website of the heritage institution concerned. The visitor determines which heritage institutions are granted access to his profile so that he can receive personalised services. This profile always remains anonymous for these services.

The UPR is linked to ZieOok, the recommendation platform developed by the Netherlands Institute for Sound and Vision. Website visitors of heritage institutions that make use of ZieOok can grant access to their personal profile. Based on that profile the institution can make recommendations from its own collection. The heritage institution can also see how items in the collection are visited and rated. In principle, it is even possible to make recommendations that cross the boundaries of collections and institutions.
Judith van Gent, Amsterdam Museum:

“Two years ago the Amsterdam Museum put its collection online as a digital repository. Although the data are sometimes incomplete, we have chosen to make the entire collection accessible via the Internet so as to encourage the use and reuse of our data. We did this because we believe public collections should indeed be accessible to the public. Furthermore, the collection is internationally important from the perspectives of history and the history of art. Everybody must be able to find and use the objects. As only 20% of the collection is physically accessible, opening up the collection online is a good alternative for making our objects visible.

Now that the collection is available online, we are searching for ways to make its use more attractive. For example, we are looking for applications to make the collection even more accessible. And as encouraging innovation and knowledge dissemination is part of our policy, we are also participating in various innovative projects that make use of our data. ZieOok therefore fits perfectly within our objectives. On our narrative websites Geheugen van Oost [Memory of East-Amsterdam], Buurtwinkels [Local shops] and Hart van Amsterdam Museum [Heart of Amsterdam Museum] we encourage the reading of stories by recommending other stories on these sites. We are under the impression that this functionality really boosts use of the sites. That is why we are so interested in ZieOok, as this application provides similar functionality for our collection online. Recommendations from ZieOok invite the visitor to explore other objects in the collection. When the digital repositories of other heritage institutions are also added to ZieOok, the functionality will become even more interesting. Then the interrelated use of different collections will be encouraged and the user experience will be further enriched.”

Project team

Patricia Alkhoven (CATCHPlus Project Office)
Hennie Brugman (CATCHPlus Project Office)
Mart Trautwein (Gridline)
Job Tiel Groenestege (Gridline)
Sien Vaessen (Netherlands Institute for Sound and Vision)
Judith van Gent (Amsterdam Museum)
Marijke Oosterbroek (Amsterdam Museum)
Pim Luiten (Netherlands Theatre Institute)
Roelof Brandsma (Netherlands Theatre Institute)

Project team CHIP

Lora Areyo (Eindhoven University of Technology)
Paul de Bra (Eindhoven University of Technology)
Rogier Brussee (Novay, formerly Telematica Instituut)
Peter Gorgels (Rijksmuseum Amsterdam)
Lloyd Rutledge (Novay)
Peter Sigmond (Rijksmuseum Amsterdam)
Natalia Stash (Eindhoven University of Technology)
Mettina Veenstra (Novay)
Yiwen Wang (Eindhoven University of Technology)

Interested?

Are you interested in UPR en ZieOok? Visit www.catchplus.nl/upr
Multiply

Keyword suggestions based on automatic text analysis

Archivists and documentalists use keywords to describe documents from their collection. If the computer can automatically index archive material then it can also make suggestions for relevant keywords and save the documentalist a lot of work. That is exactly what Multiply does.

Multiply builds further upon the CATCH project CHOICE, acronym for CHarting the informatiOn landscape employ- ing ContExt information. In this project an application was worked on that supports documentalists of the Netherlands Institute for Sound and Vision in the indexing and finding of audiovisual documents.

In Multiply this application was expanded into software that enables documentalists to analyse automatic keyword suggestions and to use these in the descriptive process. The keywords suggested originate from the Netherlands Institute for Sound and Vision thesaurus, which contains more than 150,000 terms. The manual selection of keywords is a very time-consuming and subjective process and so the automated system can save documentalists a lot of work.

Within the framework of Multiply, the Netherlands Institute for Sound and Vision thesaurus (the GTAA) was connected to OpenSKOS (see p. 42-43), so that external users can also access GTAA.

The application bases its keyword suggestions on the results of an automatic text analysis of contextual text documents. These documents are stored in a so-called context database and for Netherlands Institute for Sound and Vision these consist, for example, of broadcaster websites, wikis and programme details. The application can also make suggestions on the basis of time-based metadata, such as subtitling files or the outcome of automatic speech recognition technology.
Johan Oomen, Netherlands Institute for Sound and Vision: “The Netherlands Institute for Sound and Vision spends a lot of man hours describing audiovisual material such as radio and television programmes. Making these descriptions is very time-consuming and therefore expensive. At the same time more and more content is entering the digital archive each year and users expect far more refined access. We therefore need to use technology that can improve the efficiency of the description process and provide better access.

Within the Netherlands Institute for Sound and Vision we worked on the annotation environment, which is called MASS: the Media Archivist Support System. On the basis of text input the system automatically offers suggestions from the GTAA thesaurus. Documentalists use the MASS interface to analyse suggestions and to award the correct keywords. This semi-automatic form of description considerably speeds up the annotation process.

In the current version of MASS (October 2012) use is made of subtitling files. In the future a link will also be made to the context database of the Netherlands Institute for Sound and Vision. Furthermore the technique developed within Multiply will play an important role in an initiative started in 2012 to realise a Dutch cultural thesaurus. The Netherlands Institute for Sound and Vision, Netherlands Agency for Cultural Heritage and Naturalis have agreed to collaborate over the next few years to develop the technology to manage thesauri at a national scale.”

Project team Multiply
Jaap Blom (Institute for Sound and Vision)
Hennie Brugman (CATCHPlus Project Office)
Vincent Huis in’t Veld (Institute for Sound and Vision)
Willem Melder (Institute for Sound and Vision)
Johan Oomen (Institute for Sound and Vision)
Roelolland Ordelman (Institute for Sound and Vision)
Alma Wolthuis (Institute for Sound and Vision)
Picturae
RCE (Cultural Heritage Agency of the Netherlands)

Project team CHOICE
Hennie Brugman (Max Planck Institute for Psycholinguistics)
Luit Gazendam (Institute for Sound and Vision)
Veronique Malaise (VU University Amsterdam)
Johan Oomen (Institute for Sound and Vision)
Mettina Veenstra (Novay)

Interested?
Are you interested in Multiply?
Surf dan naar www.catchplus.nl/multiply
STITCHPlus
Semantic Interoperability To access Cultural Heritage
Meaningful searching in existing collections

For years many heritage institutions have often worked with several thesauri to open up as much of their collection as possible. The applications of STITCHPlus automatically link different thesauri allowing existing collections to be searched in a more meaningful manner.

For a computer, text is nothing more than a long string of zeros and ones. A computer has no real understanding of the content of the text or databases in which heritage institutions manage their collection. That has gradually changed over the past few years, however: scientists have developed algorithms with which a computer can independently derive relationships and meaning from text. This semantic technology can be used to automatically analyse text files.

In the CATCH project STITCH, scientists from the VU University Amsterdam investigated how semantic techniques could automatically link the various thesauri and terminology lists of the National Library of the Netherlands. This project lead, for example, to the development of the vocabulary bank OpenSKOS, described on the following pages.

The team from STITCHPlus continued where STITCH had left off. They developed two suites. The first contains a workflow in which existing thesauri are copied from their original location and converted to SKOS, a universal format that the vocabulary bank OpenSKOS also makes use of. This new thesaurus is stored locally at the heritage institution and is also added to OpenSKOS. That makes it easier to link different collections with each other.

The second suite consists of a series of applications that can recognise names (of people, locations, companies, etc.) in text and can link these to terms in SKOS thesauri or to external sources such as Wikipedia. The user can check these automatically made links and, if necessary, correct these as well.

Project team STITCHPlus
Rene van der Ark (National Library)
Paul Doorenbosch (National Library)
Willem-Jan Faber (National Library)
Dirk Kramer (National Library)
Henk Matthezing (National Library)
Theo van Veen (National Library)

Project team STITCH
Paul Doorenbosch (National Library)
Marjolein van Gendt (VU University Amsterdam)
Frank van Harmelen (VU University Amsterdam),
Antoine Isaac (VU University Amsterdam, National Library)
Henk Matthezing (National Library)
Lourens van der Meij (VU University Amsterdam, Nat. Library)
Stefan Schlobach (VU University Amsterdam)
Shenghui Wang (VU University Amsterdam, National Library)
Claus Zinn (Max Planck Institute for Psycholinguistics)
Dirk Kramer (National Library)

Interested?
Are you interested in STITCHPlus?
Surf dan naar www.catchplus.nl/stitchplus
OpenSKOS
An online platform for thesauri

The development and management of thesauri to describe collections is a time-consuming task. And once these thesauri exist they are not usually suitable for different software tools and small updates require a relatively large number of technical modifications. The online platform OpenSKOS helps to resolve these problems.

In recent years an increasing number of thesauri from the humanities and the heritage sector have been converted into SKOS (Simple Knowledge Organization System). This standard for vocabularies makes use of semantic Web technologies and Linked Open Data, as a result of which such knowledge sources can be more easily shared and linked over the Web.

OpenSKOS was developed within CATCHPlus as an online platform for thesauri in the SKOS format. OpenSKOS is the result of a collaboration between the VU University Amsterdam, the Netherlands Agency for Cultural Heritage and companies such as Adlib, Picturae and Trezorix. Institutions such as the Netherlands Institute for Sound and Vision and the National Archive of the Netherlands are already using this platform. OpenSKOS offers many advantages for both the developers and users of thesauri.

The developers of thesauri can easily upload the vocabularies they have designed and share these with other users. The developer can implement updates himself without the need to use third parties. Concepts from the developer’s own vocabularies can be simply linked to existing vocabularies.

In OpenSKOS the user can search for vocabularies that suit his requirements and wishes. Existing thesauri can be simply reused and that can save an institution a lot in development costs. The thesauri from OpenSKOS can be utilised by many different applications and are therefore not limited in how they can be used. The user always works with the latest version of the thesauri, as updates can be implemented straightaway.

Project team
version 1
Hennie Brugman (Meertens Institute)
Lourens van der Meij (VU University Amsterdam)

specification phase
Patricia Alkhoven (Meertens Institute)
Hennie Brugman (Meertens Institute)
Bert Degenhart-Drenth (Adlib)
Mark Lindeman (Picturae)
Sander van der Meulen (Trezorix)
Hans Nederbragt (Trezorix)
Supported and facilitated by RCE

version 2
Hennie Brugman (Meertens Institute)
Mark Lindeman (Picturae)

Interested?
Are you interested in OpenSKOS?
www.catchplus.nl/openskos
Concepts from thesauri of heritage institutions are often found in the archived text documents as well. Archivists often spend a lot of time linking the correct thesauri terms to these documents. The DocChecker can do part of this task for them.

DocChecker was developed by the Netherlands Agency for Cultural Heritage for text files in various formats such as Word, PDF, HTML and Excel. By uploading new thesauri or new versions of thesauri in SKOS format, the user can determine which thesauri are used. In addition to this, the thesauri in the DocChecker can be further enriched with data from other databases, such as synonyms, diminutives and conjugations.

DocChecker can be used in a web service where the user can make profiles containing the thesauri that must be used in the analysis. DocChecker then automatically links concepts from the thesauri to the text fragments in the imported documents. It suggests relevant thesaurus terms to the user that he can then validate. Matched data are stored in a metadata file that includes the links between the item concerned and the concepts selected.

**DocChecker**  
*Automatic keyword suggestion*

Interested?  
Are you interested in DocChecker?  
www.catchplus.nl/docchecker
Annotatie Repository Service
Sharing and reusing standardised annotations online

You possibly associate the word annotations with textual descriptions. Yet in a growing number of cases it can also concern web pages, photos or multimedia recordings. In this project, the participants worked on the uniformity of these annotations so that optimal use of them can be made in internal and external collaborations.

The CATCHPlus applications in this booklet demonstrate that the concept of annotations covers far more than just textual descriptions. In SCRATCH4All parts of scanned manuscripts are transcription. In CHoralPlus the outcomes of automatic speech recognition are linked to audio or video fragments. And the DocChecker links the outcomes of language analysis algorithms to words in the text. So in practice, annotations can be very diverse.

With the CATCHPlus Annotation Repository and Service all these different types of annotations can be brought together in a uniform manner and subsequently searched. This enables users to build upon each other’s results by exchanging these as annotations.

The algorithms of the DocChecker can, for example, be used on the results of the speech recognition from CHoralPlus or the handwriting recognition of SCRATCH4All.

Furthermore, searching in annotations yields a results list with references to parts of online heritage objects that might be found at different institutions.

For the standard requirements of the format of annotations CATCHPlus joined the Open Annotation Collaboration, an international collaboration that tries to realise a generic model for annotations. The annotation model of this consortium is web-based and is in keeping with the Open Data movement. CATCHPlus is collaborating with them in developing a future standard model.

Together with Seecr, CATCHPlus worked on the Open Annotation Server. This is a web service in which annotations can be uploaded, searched and exchanged. Furthermore, the annotations are available for use online and different institutions can exchange annotations using the so-called data harvester.

Interested?
Are you interested in Annotatie Repository Service? www.catchplus.nl/annotation

Project team
Hennie Brugman (CATCHPlus Project Office)
Erik Groeneveld (Seecr)
Johan Jonkers (Seecr)
Open Annotation Collaboration
PIDs
Persistent Identifiers

Confidence in unique and persistent identifiers

In the archive systems of heritage institutions unique names or numbers (identifiers) are used to refer to physical or digital objects in the collection. Thesaurus terms and metadata descriptions also have their own identifier with which they can be simply found.

Heritage institutions mostly use different identifier systems for different information systems. All objects in the collection have their own system, just like thesaurus terms, metadata descriptions and, if present, the annotations and user profiles as well. Within any given institution these identifiers are unique, but what happens if the institution concerned collaborates with other parties? Or what if the institution starts to work with a different database management system or another identifier system?

These practical examples demonstrate the need for a guarantee that identifiers really are unique and that they will not become obsolete over the course of time. If an identifier refers to an object in the collection of another institution then the user must always be able to find this object even if the other institution changes its database system or web server. The increasing importance of Internet and collaborations such as CATCHPlus necessitates solutions for Persistent Identifiers (PIDs).

The PID solution developed within CATCHPlus is based on existing and well-tested Handle technology. Handle offers a worldwide 'resolver' service: if you send a persistent identifier to the service, you are automatically led to the current URL for the identifier. CATCHPlus has developed a web service in which the user can simply manage the link between PIDs and URLs.

The resolver must always be available, of course, as otherwise users will no longer be able to access the collection. The Handle guarantees this access with several redundant copies of the resolver service. Computer centre SARA has a collaborative agreement with other European computer centres so that PIDs can be sustainably hosted and resolved even if the object or the owner no longer exist. CATCHPlus has just started a two-year project with eight large institutions and organisations to introduce the PID system as a tailor-made service.

Project team
Hennie Brugman (Meertens Institute)
Pieter van Beek (SARA)
Tibor Kalman (GWDG)
Daan Broeder (Max Planck Institute for Psycholinguistics, on behalf of CLARIN)
Eric Auer (Max Planck Institute for Psycholinguistics)
Jan Pieter Kunst (Meertens Institute)
and many more were/are involved

Interested?
Are you interested in PIDs?
www.catchplus.nl/pids
Towards a heritage broker

CATCHPlus, in contrast to CATCH, focused on valorisation and the capitalisation of research results. In other words, the construction and implementation of tools and services to improve and maintain access to heritage collections. To realise this valorisation, the heritage institutions involved were asked to submit a business plan for the tool or service they had developed. In this plan they had to detail how they thought the heritage institution would manage and exploit the software once the project has been completed.

Analyses of these business plans and the market revealed that CATCHPlus – just like the heritage sector in general – faces an important issue. Many heritage institutions are experimenting with and investing in the development of (prototypes of) digital products and services that will strengthen the digital infrastructure for heritage. However, that does not provide sufficient safeguards for sustainable management, supply and ongoing development of the tools. The parties involved in CATCHPlus were often found to lack the knowledge and capacity needed to realise sustainable safeguarding – keeping products and services available on a large scale for the long-term. Parties who can and want to take on the responsibility do not offer their services as a matter of course. Other heritage institutions and market players are often not aware of the existence and value of the tools and spin-offs. CATCHPlus lacks a party who on its behalf, and in the interest of the heritage institutions, can find a supplier for the tools and services developed by the heritage institutions. In other words a broker is needed who can make the link between the development of knowledge, prototypes and demos, and the further development and exploitation of these in the longer term. Although there are companies active in CATCHPlus, to date no organisation has come forward that actively engages in the valorisation of project results managed or retained by a heritage institution.

Therefore, in close collaboration with the CATCHPlus Project Office, Kennisland and DEN wrote a proposal for a heritage-sector-wide solution to safeguard the valorisation of project results after the completion of the project. This has laid a foundation for a broad ICT-supporting institution, a shared service organisation for the heritage sector akin to those in the education (Kennisnet) and science (Surfnets) sectors. Within this institution joint tasks on the basis of a new, cost effective and, if successful, profitable business model can be brought together. Any profits made can subsequently flow back into the sector.

Together we itemised various scenarios:
- A temporary broker who is only concerned with CATCHPlus is probably the fastest and simplest to realise. However this construction does not fulfil the vision and mission of the broker. It is not very likely that such a temporary broker
will be realised as a separate entity in the form of a separate legal entity or department with an existing organisation. • A broker who works more broadly than for CATCHPlus alone requires a good coordination. If this broker were to be managed by a consortium of collaborating heritage institutions then the work pressure for this consortium would quickly become too great. In this scenario the need would soon arise for an independent broker instead of a partnership that divides the task between its members. As an independent entity this broker can be guided by a number of heritage institutions.

All in all it seems that a heritage broker who operates proactively as an independent entity with a broad mandate will ultimately be in the best position to promote the interests of the heritage sector and to make an actual contribution to a strong infrastructure for digital heritage products and services.

In the proposal an embedded model is put forward that enables the broker to start with the CATCHPlus results and subsequently to grow towards a broader approach. In this approach the broker’s tasks will in part be realised by the collaborating heritage institutions and other parties.

The concept of a heritage broker is unique within the heritage world: it anticipates the entrepreneurial heritage director who thinks beyond institutions. The idea has already gained a foothold in the Top Sector Creative Industry and has been included in the work plan of the innovation network CLICK Cultural Heritage. The broker is expected to make an important contribution to maintaining the availability of the results that have formed the basis of countless national and international spin-offs.

This article is based on the report by Kennisland en DEN “Businessplan op hoofdlijnen. De Erfgoed Makelaar” [Outline business plan: the Heritage Broker].

Project team Businessplans
Patricia Alkhoven (CATCHPlus Project Office)
Thomas van Andel (Kennisland)
Martijn Arnoldus (Kennisland)
Hennie Brugman (CATCHPlus Project Office)
Marco Streefkerk (DEN)
In close collaboration with the CATCHPlus Steering Group

Interested?
Are you interested in the heritage broker? Visit www.catchplus.nl
Organisation

Project Office
Patricia Alkhoven (since November 2010)
Hennie Brugman
Enica Rencikers
Marius Snyders (until November 2010)
Julia Vytopil (until November 2010)

Supervisory Board
Hans Bennis, voorzitter (Meertens Institute)
Martin Berendse (Nationaal Archief van Nederland)
Riad Dekker (Naturalis)
Taco Dilberts (Rijksmuseum)
Paul Doorenbosch (National Library)
Jaap van den Herik (CATCH board)
Bas Savenije (National Library)
Guus Schreiber (VU University Amsterdam)
Pieter van Beek (SARA)
Bert Degenhart Drenth (ADLIB)
Gert-Jan van Dijk (Target Holding)
Bas Goedhart (RDE)
Erik Groeneveld (Seecr)
Thijs Janssen (Seecr)
Hanno Lams (Gemeentemuseum, external)
Marc Lindeman (Picture)
Hans Nederbragt (Trezorix)
Roeland Ordelman (X-MI)

Steering Group
Hans Bennis (Meertens Institute)
Paul Doorenbosch, voorzitter (National Library of the Netherlands)
Marco de Nett (DEN Foundation)
Eric Postma (Tilburg University)
Rosemarie van de Veen-Oei (Netherlands Organisation for Scientific Research)
Hans Westerhof (Netherlands Institute for Sound and Vision)

Supervisory Committee
Marc van der Berg (Tilburg University)
Frank Bosmans (Tresorat (Frisian Historical and Literary Centre))
Renze Brandsma (University Library, University of Amsterdam)
Nikola Elkin (Van Gogh Museum)
Ellen Fleurbaay (Amsterdam Municipal Archive of Amsterdam)
Paul Gompers (Music Center the Netherlands)
Jaap van den Herik (CATCH board)
Pim Luiten (Netherlands Theatre Institute)
Marco de Nett, voorzitter (DEN Foundation)
Marijke Oosterbroek (Amsterdam Museum)
Boudewijn Ridders (Netherlands Fotomuseum (Netherlands Museum of Photography))
Eddy Tulp (Stadsarchief & Athenaeum Bibliothek Deventer)
Henk Vanstappen (PACKED Brussels)
Rosemarie van de Veen-Oei (Netherlands Organisation for Scientific Research)
Olby Veenstra (Fries Film Archief)
Reimer van ’t Zelfde (Netherlands Institute for Art History)
More information

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

Editor: Erica Renckens, Tatataal
With thanks to everyone involved at CATCHPlus

Design and Lay out: Jeannette van Bommel
Photography: Bart van Vliet
Images: Shutterstock
Printed by Zalsman, Zwolle
Translated by Native Speaker Translations

November 2012
Cultural heritage is everywhere, and constitutes our collective memory. CATCH carries out research and develops scientific methods and techniques at the crossroads of the humanities and information technology. The results will make heritage collections more accessible to the public, professionals and scientists. Within CATCHPlus the highly promising demos of CATCH were further developed into end products that can be widely used in the Dutch heritage sector.