

Het digitaliseren van kranten

Wie betrokken is bij het digitaliseren van kranten, moest op 1 februari 2007 in Ieper zijn, op de studiedag 'Oud nieuws, nieuw leven'. Oud nieuws verwijst naar kranten als historische bron. Dit papieren erfgoed is kwetsbaar. Het heeft nood aan een nieuw leven om beschikbaar te blijven voor een groot publiek. In dit nummer licht Eva Wuyts van Erfgoedcel Ieper het project Geheugen van Ieper toe. Alexander Stierman verslaat de hele studiedag, en Iris Steen presenteert een samenvatting van de voordracht van Marie-Christine Janssens over de auteursrechten die komen kijken bij de digitale archivering en ontsluiting van kranten.

Geheugen van Ieper de digitale ontsluiting van lokale krantenarchieven

Op 1 maart 2004 sloot de stad Ieper voor de duur van 10 maanden een experimenteel erfgoedconvenant af met de Vlaamse Gemeenschap en richtte zij een Erfgoedcel op¹. De Ieperse aanvraag om een integraal en geïntegreerd lokaal erfgoedbeleid te gaan uitbouwen, werd gehonoreerd op basis van een tiental projectvoorstellen. Het belangrijkste, duurste en meest arbeidsintensieve project uit de lijst behelsde de creatie van een erfgoedbank, getiteld Geheugen van Ieper, naar analogie van Geheugen van Nederland. Dat illustreert meteen dat de ambities erg hoog lagen.

De motieven

Ieper volgde de trend die enkele jaren geleden door andere Erfgoedcellen geïnitieerd werd in Vlaanderen. Beeldbank Gent, Beeldbank Brugge en de Regionale Beeldbank Mechelen waren databanken vol historisch beeldmateriaal zoals foto's en prentbriefkaarten, doorgaans uit openbare collecties, die via een website voor een breed publiek ontsloten werden.



EVA WUYTS, Erfgoedcel Ieper
erfgoedcel@ieper.be



Ieper droomde van iets vergelijkbaars, maar wou het accent leggen op de collectie historische kranten. Hiervoor waren er drie motieven:

- vooreerst was er de onschatbare waarde van de krant voor de *Ieperse geschiedschrijving*. Want het Stadsarchief, een van de rijkste van Europa, werd tijdens de Eerste Wereldoorlog volledig verwoest. Ook het Stedelijk Museum en een groot deel van zijn verzameling gingen in vlammen op. Enkel de Stadsbibliotheek kon tijdig ontruimd worden. Jaren later, na een verblijf in Frankrijk, keerden de boeken- en de krantencollecties terug naar hun plaats van herkomst, waarna generaties van bibliothecarissen ze als schatten zijn gaan koesteren. En terecht: voor de onderzoekers van het 19de-eeuwse Ieper vormen zij een unieke bron. Hun berichtgeving – gemeenteraadsverslagen, marktprijzen en advertenties, sportuitslagen, gerechtelijk nieuws, overlijdensberichten... – beslaat bovendien alle facetten van het maatschappelijke leven;



De Openbare Bibliotheek op 17 december 1914. © Antony d'Ypres, Oostende.

De partners en hun wensen

Met het erfgoedconvenant beschikte leper in 2004 over zowel de middelen als een Erfgoedcel om het project te coördineren. De nieuwe stadsdienst vormde samen met de ICT-dienst en de drie aanwezige erfgoedinstellingen – de Stedelijke Openbare Bibliotheek, het Stadsarchief en de Stedelijke Musea – een werkgroep die de krijtlijnen van het project zou bewaken. Elke partner bracht zijn specifieke expertise en zijn eigen wensen mee. Voor alle partijen gold dat digitalisering nooit een doel op zich mocht zijn. Maar over de vraag waar- toe het dan een middel kon zijn, bestonden wel uiteenlopende opvattingen.

- ten tweede zijn deze bronnen erg *bedreigd*. De veelvuldige consultatie door lokale onderzoekers, maar zeker de voortschrijdende verzuring vormen een interne tijdbom. Zonder voorzorgsmaatregelen leidt dit afbraakproces op korte termijn tot de totale vernietiging van dit erfgoed;
- ten derde wou de leperse Erfgoedcel in de geest van het convenant bewust *experimenteel en vernieuwend* werken. De zoektocht naar oplossingen voor de specifieke uitdagingen die horen bij het digitaal ontsluiten van kranten – bijvoorbeeld de combinatie van tekst met beeld, het formaat en daarmee samenhangend de grote opslagcapaciteit, de verschillende ontsluitingsniveaus... – ligt in die lijn.

De ambities met *Geheugen van leper* uitten zich niet alleen in de voorkeur voor kranten in plaats van het klassieker beeldmateriaal. Er was ook de intentie om een integraal verhaal te brengen. Ons opzet was een databank te maken met allerlei soorten (tweedimensionaal) erfgoed dat onderling gekoppeld kon worden. De kranten als eerste collectie zouden de ruggengraat vormen omdat via de artikelen heel wat inhoudelijke verbindingen konden worden gelegd met andere collecties. Een voorbeeld: een zoekopdracht naar de leperse politicus Alphons Vandenpeereboom levert zowel een overlijdensbericht, een foto als een menukaart op. Of nog: wie een artikel leest over het zwembad of de velodroom, kan meteen doorklikken naar foto's ervan. Hoewel het al mogelijk is om deze theorie om te zetten in de praktijk, is dit om allerlei redenen nog niet gebeurd. De belangrijkste is dat de Erfgoedcel ondertussen samen met de projectvereniging CO₇ gestart is met *Westhoek Verbeeldt*, een regionale beeldbank voor zwerfgoed².

Voor de erfgoedinstellingen die zelf tientallen krantentitels bewaren, waarvan sommige in precare staat, primeerde de conservering. Voor de Erfgoedcel die volgens het convenant de opdracht had om de maatschappelijke plaats van het culturele erfgoed te versterken, de ontsluitingsmogelijkheden.

Deze dualiteit bleef doorheen het hele denkproces bestaan en was bepalend voor de keuzes die gemaakt werden. Het afgelegde traject kan omschreven worden als een continue evenwichtsoefening om 'the best of the both worlds' te combineren: maximale ontsluiting met de nodige aandacht voor conservering. De digitale beschikbaarheid van deze collecties op het internet maakt alleszins dat de originelen niet langer geconsulteerd hoeven te worden. Ondanks de digitale back-ups blijft het Stadsarchief zijn microverfilmingsbeleid (vooral van naoorlogse kranten) voortzetten. Dat de huidige technologie nu ook de mogelijkheid biedt om eventueel vanuit de digitale bestanden microfilms te laten maken, opent perspectieven maar deze optie werd voorlopig nog niet genomen.

De data zelf worden op meerdere plaatsen en op verschillende dragers bewaard, namelijk op een server en op harde schijven. Het spoor van opslag op optische media werd verlaten omdat dit geen betrouwbare, duurzame archiveringswijze is. Bovendien zijn enkele lopende meters cd's of dvd's voor een klein archief een bijkomende belasting en is de jaarlijkse kwaliteitscontrole ervan niet haalbaar.

De inventarisering en selectiecriteria

Omdat de middelen niet onbeperkt waren, moest er een doordachte selectie worden gemaakt van de duizenden kranten die naar aanleiding van het project door de Stedelijke Openbare Bibliotheek overgedragen werden aan het Stadsarchief. Archiefmedewerker John Desreumaux zorgde voor de preservatie (door verpakking in zuurvrij materiaal) en een nieuwe, uiterst gedetailleerde inventaris. Het belang van dit document valt niet te onderschatten. Het Stadsarchief heeft nu niet alleen inzicht in zijn collectie – van de aard van de kranten, over de staat op paginaniveau waarin ze zich bevinden tot de lacunes in jaargangen of edities – maar ook een checklist waarmee onvolledige reeksen digitaal aangevuld kunnen worden door ontbrekende exemplaren bij andere bewaarinstellingen op te sporen. Uiteraard is de inventaris een uiterst waardevol instrument bij de selectie van de te digitaliseren kranten.

De definitieve keuze van de titels is gemaakt op grond van een aantal wegingsfactoren, de ene al wat objectiever dan de andere. De enige absolute voorwaarde was dat de kranten verschenen waren in het arrondissement leper. Kranten met een meer provinciaal of nationaal bereik rekenen we niet tot onze opdracht. Wat betreft de ouderdom werd het belangrijk geacht om een kopie te kunnen maken van de oudste exemplaren, ondanks het besef dat jongere uitgaven (jaren 1950-1980) op meer interesse zouden kunnen rekenen van de doorsnee leperlingen. De auteurswet bepaalde de keuze om voorlopig alleen titels tot aan de Eerste Wereldoorlog online te brengen. Zelfs die cesuur

houdt nog een beperkt risico in. Omdat de uitgeverijen al decennia geleden hun activiteiten stopgezet hebben en het over kranten gaat die vaak anonieme of enkel met initialen ondertekende artikelen bevatten, is de zoektocht naar rechtshabbers zeer lastig tot onmogelijk. En die zijn er nog steeds: bij een reporter die 35 jaar oud was in 1900 en op 70-jarige leeftijd overleed, kunnen de erfgenamen vandaag nog steeds rechten doen gelden.

Verder werden de volgende criteria gebruikt:

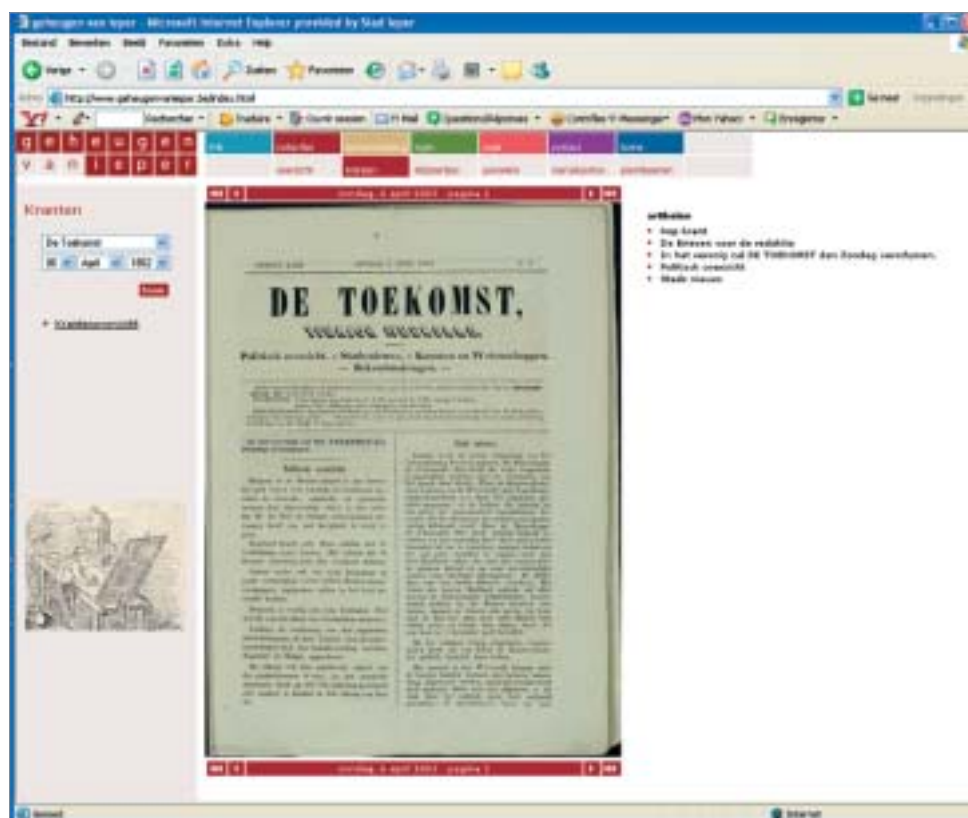
- de taal: hoewel het overgrote deel van de kranten Frans-talig is, werd voor de publieksvriendelijkheid de voorkeur gegeven aan Nederlandstalige of tweetalige kranten;
- de interesse van het publiek: vaak geraadpleegde titels kregen voorrang om verdere beschadiging door consultatie te voorkomen;
- de staat: beschadiging kon een reden zijn om prioritair te digitaliseren, zware schade een reden om het (nog) niet te doen;
- de politieke kleur: er werd aandacht besteed aan de vertegenwoordiging van alle strekkingen;
- de verschijningsperiode: er moest voldoende spreiding in tijd zijn en door de auteurswet vormde de Eerste Wereldoorlog het voorlopige eindpunt;
- de diversiteit van de inhoud: er werd een 19de-eeuws equivalent van zowel *Het Laatste Nieuws* als *De Standard* opgenomen;
- de omvang en volledigheid van reeksen: titels waarvan slechts enkele exemplaren overgebleven waren, werden niet opgenomen.

In mindere mate werd ook rekening gehouden met het bestaan van microfilms of -fiches en de zeldzaamheid.

Het ultieme criterium was uiteraard de beschikbaarheid van middelen: van de honderdduizenden pagina's konden uiteindelijk 44.000 pagina's verwerkt worden. Die worden sinds 1 december 2005 gefaseerd online gebracht³.

Nadat de voorbereidingen en het onderzoek achter de rug waren, kon het project openbaar aanbesteed worden. Hiervoor werd een omvangrijk lastenboek opgemaakt waarop

De Toekomst, een krant in webweergave.



Een beschadigde krant in de zoomfunctie.

gespecialiseerde bedrijven konden inschrijven voor één tot vier afzonderlijke delen: het digitaliseren van de collecties; de bouw van de database; de bouw van de website; en de hosting. Na een lange onderhandelingsfase werd de opdracht integraal toevertrouwd aan X-Cago uit Roermond (Nederland). Dit bedrijf verzorgt de online edities van heel wat binnen- en buitenlandse kranten en heeft ervaring met het digitaal conserveren, beschrijven en ontsluiten van allerlei soorten (beeld)collecties.



De digitalisering

In tegenstelling tot de meeste andere digitaliseringsprojecten van kranten, werden de leperse exemplaren niet louter beschouwd als tekstdragers, maar als volwaardig erfgoed. Daarom was, behalve de optimale representatie van de tekstuele informatie, evenzeer het behoud van de originele 'look and feel' van belang. De kleur en de textuur van het papier mochten letterlijk gezien worden. We geloofden immers rotsvast in de visuele kracht en de ontwapenende charme van een oude krant. Een lezer die ook de verkleuringen, de vouwen en scheurtjes of in sommige gevallen zelfs de kogelgaten kan zien, zal er meer respect voor opbrengen. Deze ervaring, het 'archieffevoel' genoemd, was de toetssteen bij allerlei beslissingen.

Vooreerst impliceerde dit dat de 19de-eeuwse bindingen niet werden losgemaakt omdat zij integraal deel uitmaken van de collectie. Dit ging in bepaalde gevallen weliswaar ten koste van de leesbaarheid omdat soms tekst verdween door een te strakke band. Voorts werd beslist om niet te kiezen voor de goedkopere digitalisering van de microfilms, maar voor scanning vanuit de originele druk én in kleur. Elke pagina werd op locatie apart en handmatig gedigitaliseerd met een videoscanner. De moederbestanden in tiff-formaat werden gearchiveerd maar voor de webweergave ook omgezet in jpg. Tot slot werden de kranten altijd in hun huidige staat gescand, wat in het geval van vouwen betekende dat soms een regel tekst verdween.

De digitale kopieën werden eveneens zo min mogelijk gemanipuleerd. Omdat het merendeel van de kranten ingebonden was, stond het beeld lichtjes al eens bol of scheef.

Dit probleem werd softwarematig verholpen. Maar rafelige randen werden niet recht afgesneden. Kortom: de digitale weergave moest de originele krant zoveel mogelijk benaderen. Voor de conservering garandeerde dit de best mogelijke kopie en voor de ontsluiting hielp dit de archiefbeleving tot in de huiskamer te brengen.

Tekstherkenning

Hoewel het opzet was de raadpleging van kranten zoals die in een echte archiefleeszaal gebeurt zoveel mogelijk te evenaren, werd toch de beschikbare technologie optimaal ingezet om onderzoek te vergemakkelijken. Daartoe krijgt de gebruiker een zo diep mogelijk ontsluitingsniveau aangeboden. Bij kranten is dit behoorlijk complex: er zijn titels (die in tijd al eens evolueren of verdwijnen en enkele jaren later opnieuw verschijnen), jaargangen, edities, pagina's, rubrieken, kolommen, artikelen, alinea's en woorden.

Het handmatig toekennen van trefwoorden aan artikelen is onbegonnen werk: elk woord is een mogelijk trefwoord, meta-informatie impliceert dat elke tekst moet worden gelezen en de relevantie zal verschillen van onderzoeker tot onderzoeker. Daarom werd gekozen voor 'Optical Character Recognition' (OCR), meer bepaald voor een door X-Cago getrainde versie van Abby's Fine Reader. De voordelen van tekstherkenningssoftware zijn dat alle geëxtraheerde woorden automatisch in XML worden toegevoegd aan de metadata, dat de artikelen woordelijk doorzoekbaar zijn en dat zoekresultaten kunnen worden gehighlight in de pdf.

versie. Hiermee is er ontsluiting op het hoogste niveau. Hoewel er dankzij de kleurnuances geen verlies van informatie was, verliep de tekstherkenning niet altijd probleemloos. Doordruk van de artikelen op de andere bladzijde, vlekken, beschadigingen, slijtage van de drukletters, de diverse typografie en verschillende schrijfrichtingen op een pagina verlaagden de herkenbaarheid. Dit werd in de mate van het mogelijke geminimaliseerd door het contrast van de beelden te optimaliseren. Daarnaast zijn er enkele vrijwilligers die de artikeltitels manueel opschonen.

De taalproblematiek

Wat vaak vergeten wordt is dat de terugvindbaarheid van de informatie ook bepaald wordt door de taal. Eén blik in een oude krant is genoeg om te beseffen dat de spelling en het woordgebruik vroeger anders was. 'Ieper' bijvoorbeeld werd gespeld als 'Yper' en de 19de-eeuwse spoorweg was gekend als de 'ijzeren weg'. Om het resultaat van zoekopdrachten te verbeteren heeft X-Cago de zogenaamde 'stemmer' geïmplementeerd. Via deze applicatie worden Oud-Nederlandse varianten en spellingsvarianten aan de hedendaagse woordenlijst gekoppeld. Zo wordt er naast de ingevoerde zoekterm 'pasen' ook gezocht naar de variant 'paschen'. Op diezelfde manier zorgt de stemmer ervoor dat typische OCR-foutpatronen worden gecorrigeerd.

Om de Franstalige kranten te doorzoeken moeten voorlopig nog Franse zoektermen worden ingegeven. Maar koppeling van de stemmer aan (vertaal-)woordenboeken zou ook hier een oplossing kunnen bieden. Voorlopig worden enkel zoektips weergegeven, waaronder truncatie.

De segmentering

Een tweede instrument om onderzoek te optimaliseren was de weergave van kranten op artikelniveau. Met behulp van ClipWorX kon X-Cago de artikelen en advertenties digitaal uit de krantenpagina's extraheren. Net als bij de OCR is deze software niet onfeilbaar, maar is het een uitstekende toepassing die talloze voordelen heeft. Vooreerst kan er binnen de contouren van individuele artikelen worden gezocht, wat leidt tot een veel beter zoekresultaat. Het is bijvoorbeeld mogelijk om in één keer alle relevante artikelen over een bepaald onderwerp te zoeken en te vinden.

Ten tweede is elk artikel in zwart-witversie afzonderlijk opvraagbaar. Zo hoeven er geen grote bestanden gedownload te worden. In de pdf-versie is elk artikel trouwens afdrukbaar, vergezeld van een disclaimer, de volledige bronvermelding (titel, datum, pagina) en een verkleind beeld van de pagina waaruit het artikel afkomstig is. Gebruikers die toch de oorspronkelijke kleurweergave van het hele artikel willen bekijken, kunnen dat via de zoomfunctie.

Een derde voordeel is dat met Editor Assist, een applicatie van X-Cago om de metadata aan te passen, manueel linken gelegd kunnen worden tussen verschillende artikelonderdelen. Dat is bijvoorbeeld bij een tekst die doorloopt op een andere pagina, of bij een roman die in verschillende afleveringen gepubliceerd is, of tussen een artikel en een ander bestand binnen én buiten de eigen databank. Ditzelfde programma biedt de mogelijkheid om bepaalde paginaonderdelen – foto's of artikelen bijvoorbeeld waarvoor van de auteur geen toelating verkregen is – niet online te tonen. Dankzij de segmentering hoeft dus niet de volledige

pagina van het net gehaald te worden. Dit brengt ons bij de laatste fase van het project. Want de selectie, de digitalisering, de tekstherkenning en de segmentering van de kranten waren stuk voor stuk afgestemd op hun publicatie op de speciaal ontwikkelde website www.geheugenvanieper.be.



Een artikel in artikelweergave

De website en zijn doelgroepen

In tegenstelling tot de instelling waar de papieren kranten bewaard worden, is *Geheugen van Ieper* altijd en wereldwijd toegankelijk. Deze erfgoedbank mikt op twee belangrijke doelgroepen. Enerzijds de streekbewoners, van wie de meeste nooit eerder met een archief in aanraking kwamen, anderzijds onderzoekers in brede zin: genealogen, heemkundigen, wetenschappers, studenten, leerkrachten..., tot ver buiten Ieper. Voor elke doelgroep zijn er specifieke toepassingen en mogelijkheden.

Voor de lokale bevolking werd een aantrekkelijke en gebruiksvriendelijke website ontwikkeld. De pagina's zijn overzichtelijk opgebouwd en de navigatiebalk met frisse kleuren is eenvoudig te bedienen. Het 'artikel van de dag' – een dagelijks veranderende selectie van ludieke berichten – prikkelt tot een herhaalbezoek. De kranten worden dankzij het programma Archive Ex Press in hun authentieke 'look and feel' gepresenteerd. Via een navigatiebalk blader je van pagina naar pagina en van editie na editie. Interessante artikelen zijn met een klik opvraagbaar, maar enkel voor geregistreerde gebruikers. Het invullen van een naam en een e-mailadres volstaat om toegang te verkrijgen tot alle functies van deze erfgoedbank.

De zoomfunctie op de originele bestanden is niet alleen een hulpmiddel bij slechte of te kleine druk, maar speelt ook in op het archiefgevoel omdat de details van het origineel zichtbaar worden. Dit alles is mogelijk zonder grote bestanden over het netwerk te transporteren en zonder originele beelden voor derden kopieerbaar te maken.

Onderzoekers worden bediend door een uitgebreide zoekfunctie (and, or, not) en de mogelijkheid hun zoekresultaten te sorteren (per collectiesoort, op datum of volgens relevantie). Relevante artikelen kunnen niet alleen afgedrukt worden, maar ook via bookmarks met cookies worden bewaard op de eigen computer. De digitale tentoonstelling die in de vorm van een nieuwsbrief verzonden wordt naar alle geregistreerde gebruikers, geeft op thematische wijze duiding bij de collecties. Maar de hoop is dat er op termijn ook een publicatieforum van gemaakt wordt dat de resultaten van allerlei krantenonderzoek publiek toegankelijk maakt.

De resultaten

Geheugen van Ieper werd op 1 december 2005 gelanceerd met een uitgebreide communicatiecampagne. Er werden onder meer notablokjes in de vorm van een muismatje verspreid en elke Ieperling kreeg *De Toekomst* in de bus, een historisch uitzijnde krant met een selectie van de markantste artikelen, gecombineerd met uitleg over het project. Sindsdien is de erfgoedbank een belangrijke, veel-geraadpleegde bron geworden voor historisch onderzoek. Maar het project heeft ook geleid tot een verhoogd erfgoedbewustzijn bij het beleid en het publiek. Zij herontdekten het

19de-eeuwse Ieper – een periode die nogal in de vergetelheid was geraakt door de grote aandacht voor de Eerste Wereldoorlog en het middeleeuwse Ieper – en daarmee ook hun eigen dorps-, verenigings- of familiegeschiedenis. We zien krantenartikelen opduiken op kermisaffiches, in nieuwsbrieven en publicaties allerhande, maar evengoed als inspiratiebron voor theaterstukjes. Dat het project geresulteerd heeft in een beter behoud en beheer van de kranten zelf, is de kers op de taart. Zo kreeg oud nieuws, alsnog een nieuw leven.

Noten

1. Een erfgoedconvenant is een contract met resultaatsverbintenis tussen een stad of samenwerkingsverband en de Vlaamse Gemeenschap. Een Erfgoedcel is verantwoordelijke voor de uitvoering ervan. Het Iepers convenant werd eind 2004 verlengd voor de periode 2005-2008.
2. CO₇ werd eind 2005 opgericht als een culturele projectvereniging van de volgende steden en gemeenten: Ieper, Poperinge, Zonnebeke, Heuvelland, Langemark-Poelkapelle, Vleteren en Mesen. De regionale beeldbank bestaat uit een koepsite – www.westhoekverbeeldt.be – en 7 gemeentelijke sites waarvan www.ieperverbeeldt.be er een is.
3. *Gazette van Yperen, L'Opinion, Het Weekblad van Ieperen, De Kunstbode, De Weergalm, Nieuwsblad van Yperen, De Toekomst en La Lutte/De Strijd* uit de periode 1862-1914 zijn al consulteerbaar. De komende maanden worden *Journal d'Ypres* en *Het Ypersche toegevoegd* en er zijn ook plannen om de collectie dit jaar nog uit te breiden met titels tot aan het einde van de Tweede Wereldoorlog.

Een volledig verslag van de studiedag is te lezen op:
www.erfgoedcelieper.be · www.vbad.be · www.culturelebiografie.be.
 Een gedrukt exemplaar is op aanvraag verkrijgbaar bij Erfgoedcel Ieper.

SAMENVATTING

'Geheugen van Ieper' is een erfgoedbank die historische kranten uit het arrondissement Ieper digitaal ontsluit. De krantenpagina's worden in kleur weergegeven, ze zijn woordelijk doorzoekbaar en alle artikelen zijn afzonderlijk raadpleegbaar zijn. Dat maakt dit project uniek. Sindsdien worden de originele kranten beter bewaard en maakt een nieuw publiek kennis met dit erfgoed. Dit artikel schetst de doelstellingen, de uitdagingen en knelpunten van het project.

ABSTRACT

'Geheugen van Ieper' is a heritage database indexing historical papers from the Ypres district. The newspaper articles are reproduced full-colour and can be searched word for word, and articles can be consulted separately. That makes the project unique. Now, the original newspapers enjoy a better conservation and a new public gets acquainted with this heritage. The article outlines the objectives, the challenges and the bottlenecks of 'Geheugen van Ieper'.