

ICHICH

*Increased Contextualisation of Humanist Intangible Cultural Heritage (ICHICH).
Naar een verhoogde toegankelijkheid van digitale tijdschriften.*



Collegagroep inhaalbeweging digitale collectieregistratie
27 maart 2024

Elise Dewilde en Ellen Soetens

Het project

TIJDSCHRIFTEN:

- Digitalisering
- OCR-verbetering
- Terbeschikkingstelling

IMMATRIEEL CULTUREEL ERFGOED

Citizen science onderzoeksproject

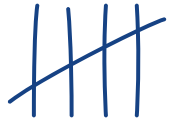


Digitalisering

tijdschriften van regionale en lokale vrijzinnig humanistische actoren



1950-



5489 tijdschriftnummers uit 128 tijdschrifttitels: 118 073 bladzijden



min. 21,5 x 13,5cm - max. 44 x 29,7cm / losbladig, gebonden / vaak
zwart-wit (ca. 20 procent kleur) / meestal in goede staat, overwegend goede druk,
maar veel materiaal is gestencild



Digitalisering



CONTEXT

OFFERTEAANVRAAG EN INVULFORMULIER

AARD VAN HET MATERIAAL

PROCEDURE

TESTFASE EN BARCODERING

VERPAKKING, TRANSPORT, BEWARING EN ZORG

OPNAME

TESTKAARTEN

NABEWERKING

KWALITEITSCONTROLE

AANLEVERING

TECHNISCHE SPECIFICATIES

RANDVOORWAARDEN

EIGENDOM EN RECHTEN

VERZEKERING

PERSONEEL

UITVOERING EN LOCATIE

BETALING

GESCHILLEN



Digitalisering

		2021	2022												2023											
			jan.	feb.	mrt.	apr.	mei	jun.	jul.	aug.	sep.	okt.	nov.	dec.	jan.	feb.	mrt.	apr.	mei	jun.	jul.	aug.	sep.	okt.	nov.	dec.
AANVRAAG		LB	gunningspr.	voorb.																						
			digitalisering (incl. controles)																							
PLANNING FIRMA	1		Kick-off + test + def. planning		barc.	digitalisering																				
	2		digitalisering																							
EFFECTIEF		LB	gunningspr.	testfase				barc.	digitalisering																	



Digitalisering



Logica digitaliseringsfirma <-> erfgoedsector



Samenwerking binnen erfgoedsector?

- Uitwisseling van ervaring en kennis
- Gezamenlijke projecten
- Gezamenlijke afspraken



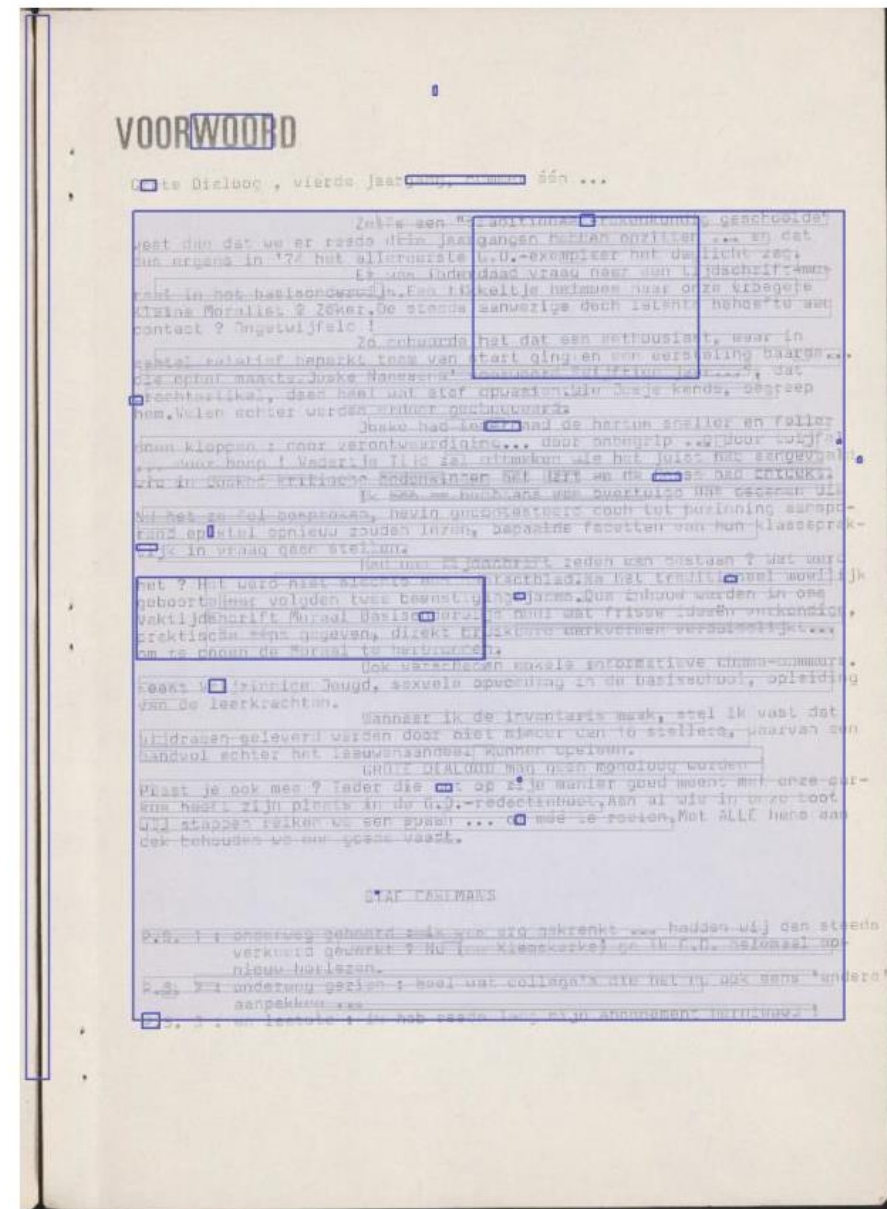
OCR

- OCR was na afronding digitalisering van mindere kwaliteit dan verwacht
 - Verschillende elementen die OCR beïnvloeden
 - Slechte/scheve print
 - Vervagende druk
 - Doordruk
 - Gestencild materiaal
 - Vastgeniete bijlages
 - Periodieken voor kinderen
- Traject OCR-verbetering met Digital Mathematics (DIMA) VUB



Lay-out analyse

- Eerste analyse alto-xml
→ lay-out niet voldoende herkend
- Samenstelling representatief sample
 - 600 pagina's
 - 11 titels
 - Periode 1960 – 2006
 - Diverse set met verschillende verbeterpunten

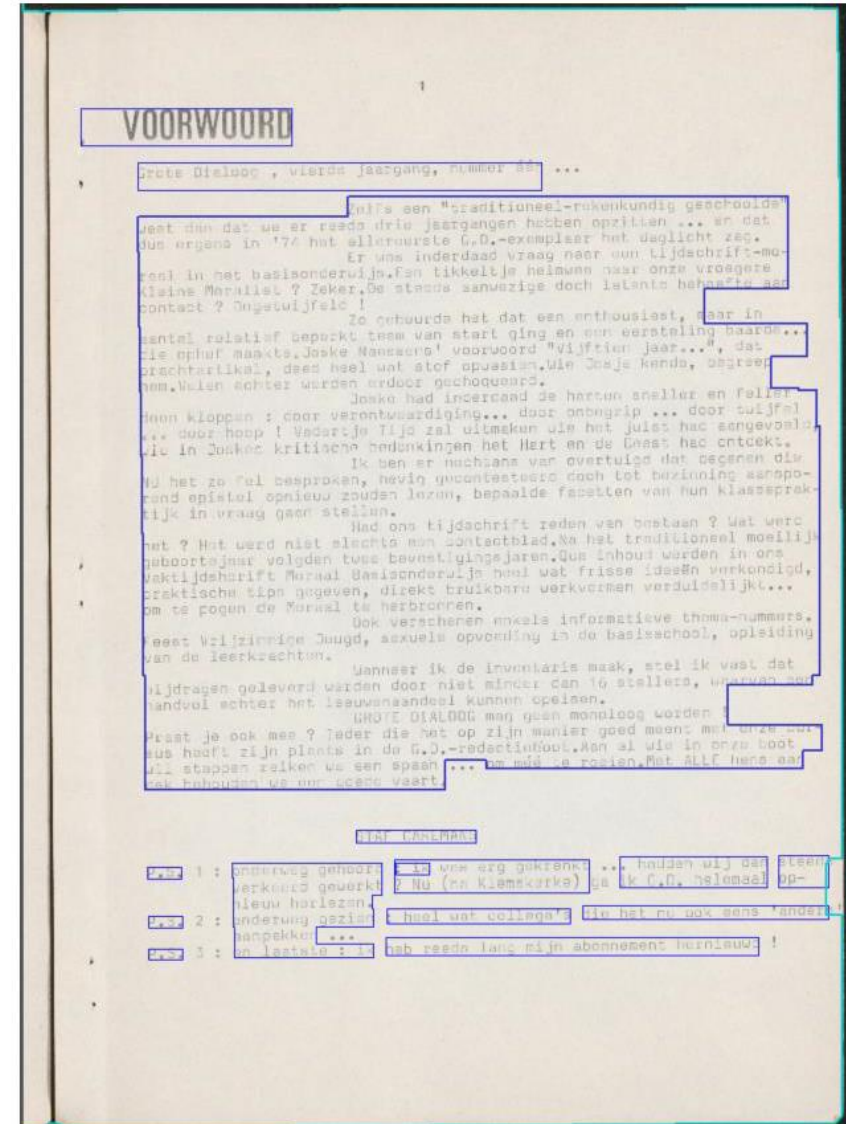
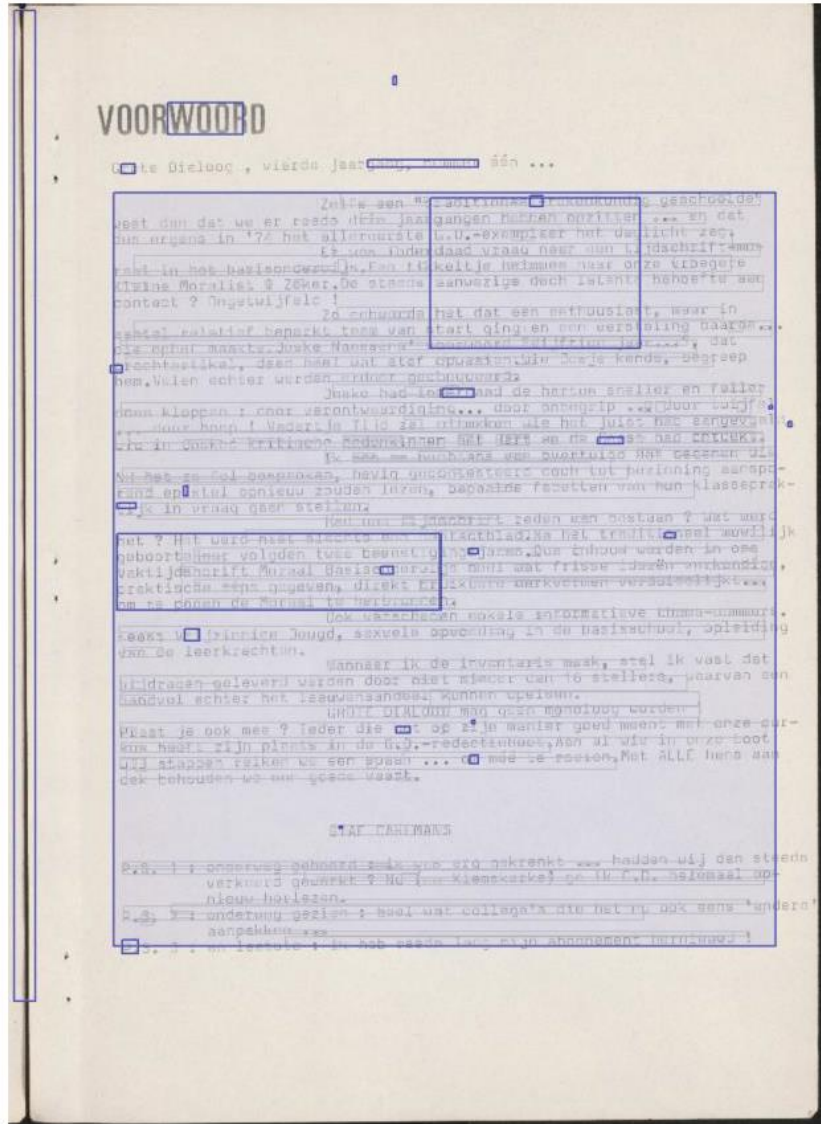


Lay-out analyse

- Lay-outanalyse van sample met DSPH
 - = algoritme dat menselijke kenmerken gebruikt om teksthomogeniteit en tekstniet-homogeniteit te onderscheiden
 - Universeel DSPH framework, niet gefinetuned per tijdschrift
- Resultaten DSPH framework vergelijken met benchmark
 - Betere herkenning tekststructuur met DSPH
 - Betere herkenning grafische componenten zoals foto's met DSPH
- Betere herkenning van lay-out met DSPH, maar verder onderzoek nodig op vlak van lay-outherkenning bij bv vervaagde print



➤ DSPH operates to extract a close description of text regions



CAVA_TS_FICT_0006_0011_005_ALTO

CAVA_TS_FICT_0006_0011_005_DSPH

Lay-out analyse

Verdere analyse met ground truth materiaal (oorspronkelijk sample + aantal extra nummers)

- representatief sample ground-truth materiaal
 - tekstcomponent door manuele transcriptie
 - Layoutcomponent door manuele annotatie
- Ground truth materiaal om
 - Objectief de resultaten van algoritmes te onderzoeken
 - Voor ontwikkeling van machine learning models
- Resultaat
 - OCR-accuraatheid sample is 72,67% (success rate measure)
 - Layout-accuraatheid sample is 74,26%



Active information exploitation

- Door middel van Intelligent Document Processing (IDP)
- METIS framework, integreert large language models (LLM) in DSPH framework
- 1 volume opgeladen
 - Layoutverwerking
 - Teksttranscriptie
 - Language model
- Resultaat: gericht vragen stellen, LLM zoekt antwoorden in tijdschrift



Terbeschikkingstelling van de tijdschriften

Inhoudelijke fiches en gestructureerde metadata over de titels:

- Voorbereiding dataset Wikidata en ODIS
 - Tiental items reeds aangemaakt in Wikidata (grote reeksen)
- Overzicht van gedigitaliseerde volumes op website
- Inhoudelijke fiches op website

Terbeschikkingstelling van gedigitaliseerd materiaal in IIF-viewer

- Samenwerking met universiteitsbibliotheek VUB
- Op basis van iGuanaDAM en Universal viewer
- Tijdschriften gefaseerd toevoegen



Terbeschikkingstelling van de tijdschriften

Afb. 1 van 16 Ga

INHOUD << Het Vrije Woord : jg. 5, nr. 8 Orgaan van het humanistisch verbond

Index Thumbnails

002 003

004 005

006 007

VIJFDE JAARGANG — Nr. 8 1 MEI 1960

HET VRIJE WOORD

ORGAAN VAN HET HUMANISTISCH VERBOND

Versijnt 20 maal per jaar — Redactie en abonnementsdienst: Alfred Cooltraat 30, Antwerpen.

Is objectieve censuur mogelijk ?

H. BUSKENS

Op 13 april werd een televisie-uitzending van « Het Vrije Woord » verboden omdat de programmaleiding van het N.I.R. meende dat er uitlatingen in voorkwamen die in hoge mate kwetsend zouden geweest zijn voor telekijkers van christelijke denkwijze. Het was een uitzending over « Het zoenoffer », welk onderwerp alleen al, vooral in de Goede Week, aanstootgevend werd geoordeeld. Het leugenen van mensensoffers om gunsten af te smeken van de goden werd erin verworpen, mede op grond van Bijbelse teksten. Nochtans werd gevreesd dat de arjelovt kijker een pa... zou trekken tussen dergelijke menseso... en het offer van Christus, alhoewel in de... het duidelijk werd gezegd dat van zulk ee... parallel geen sprake kan zijn. Christus w...rd immers helemaal niet gedood door de j...den om de gunst van God af te smeken, het was veelter een vulgair moord, sonder enig excuus.

Op de vergadering die door de Directeur-Generaal van het N.I.R., de heer Jan Boon, werd belegd met de vertegenwoordigers der gastprogramma's kwam tot uiting dat in de Beheerraad van het N.I.R. klachten zijn opgespan over de agressiviteit en het polemisch karakter van de televisie-uitzendingen van « Het Vrije Woord ». Men had de indruk dat de onderwerpen speciaal gekozen werden om polemiek uit te lokken, tervoor de uitzendingen van in de aanvang bedoeld waren om het positief propageren van een levensmoraal — dus een moraal zonder godsdienst — mogelijk te maken. De wens werd uitgedrukt dat de uitzendingen van « Het Vrije Woord » zich aan deze aanvankelijke doelstelling dienstbaar zouden maken.

Hoe staan wij tegenover de houding van het N.I.R.

Ik meen niet dat wij de heer Boon (en kunnen vervolgen). Hij draagt tegenover de Beheerraad de verantwoordelijkheid der uitzendingen te laten verlopen volgens de statuten van het N.I.R. En deze schrijven voor dat geen enkele uitzending kwetsend mag zijn voor één of andere bevolkingsgroep. Maar hier ligt de fout. Op die wijze ontstaat een censuur op dubbele subjectieve basis: wat is kwetsend voor een bepaalde groep en wie oordeelt of het kwetsend is?

Wettelijk staan wij machteloos. Alleen wil het mij voorkomen dat de statuten van het

Vender in dit nummer:

- **Tax overweging**
- **Die moet veranderen**: katechismussen in België (J. Engelen).
- **Geen politiek en sociaal** (W.B.).
- **Roos J.** (W.B.).
- **Uits over Albert Schweitzer** (H.V.T.).
- **Van en over Albert Schweitzer** (H.V.T.).
- **Vredesgezamband** (Een moralist).
- **Bleken lating of in stille**.
- **Paralysme** (W.B.).
- **Ne 150 jaar laborante** (F.R.).
- **Integratie** (F.R.).
- **Motie van het H.S.V.**.
- **Laster**, gericht door Frans De Wisme.
- **Onze vrije tribune**:
 - « **Federalisme** », door Herman J. Clarys.
 - « **Een prairie en een vaagge achterna** », door Hendrik Van Tichelen.
- **Abdeloppeven**.
- **Jean Rostaand**, door J. Scripps.
- **Over de samen**, door Jack Fauster.
- **Over haldensmoed**, interview met Louis Verbeeren.
- **Lasteren, lezen en kijkers schrijven**.
- **Bout allerlei**: Het evolutieve woord, door H. Van Tichelen.
- **Leuze godachten**.
- **Het prairie van de humanist**, door R. Lancelot.

— 113

MEER INFORMATIE

Over dit item

Titel

Het Vrije Woord : jg. 5, nr. 8 Orgaan van het humanistisch verbond

Impressum

Antwerpen : Humanistisch Verbond, 1960

Beschrijving

Het Vrije Woord : jg. 5, nr. 8 Orgaan van het humanistisch verbond

pagina 001 of 016

Zoek binnen dit item: Zoekterm

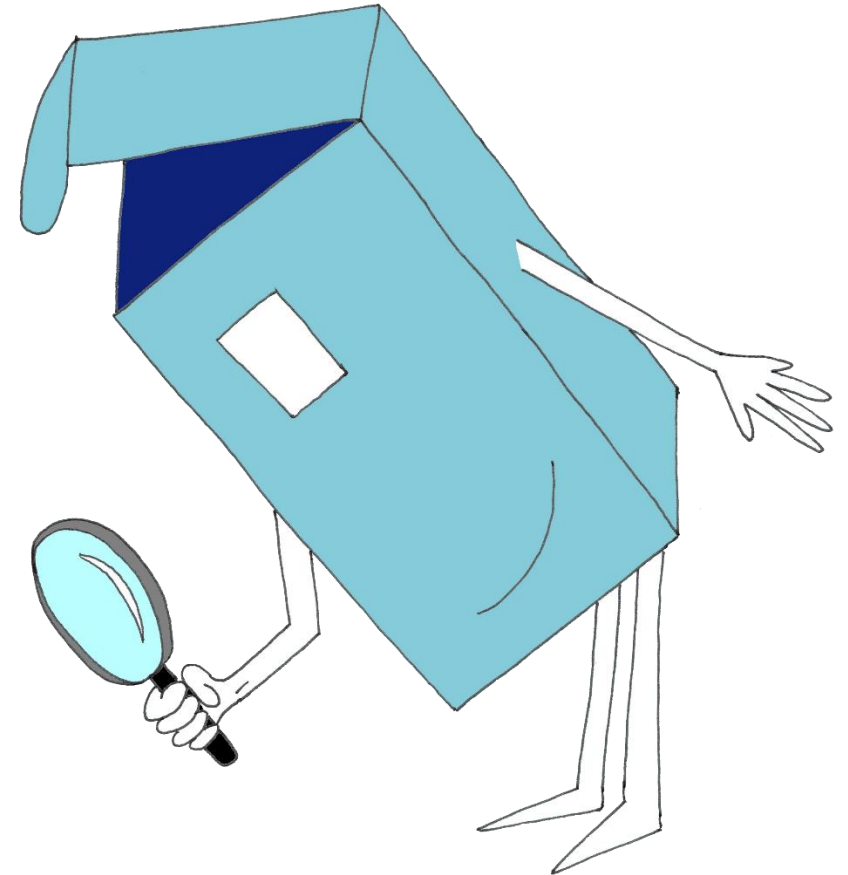


Terbeschikkingstelling van de tijdschriften

Aanpak rechtenproblematiek:

Contacteren van de verantwoordelijke uitgevers

Analyse persoonsgegevens in tijdschriften

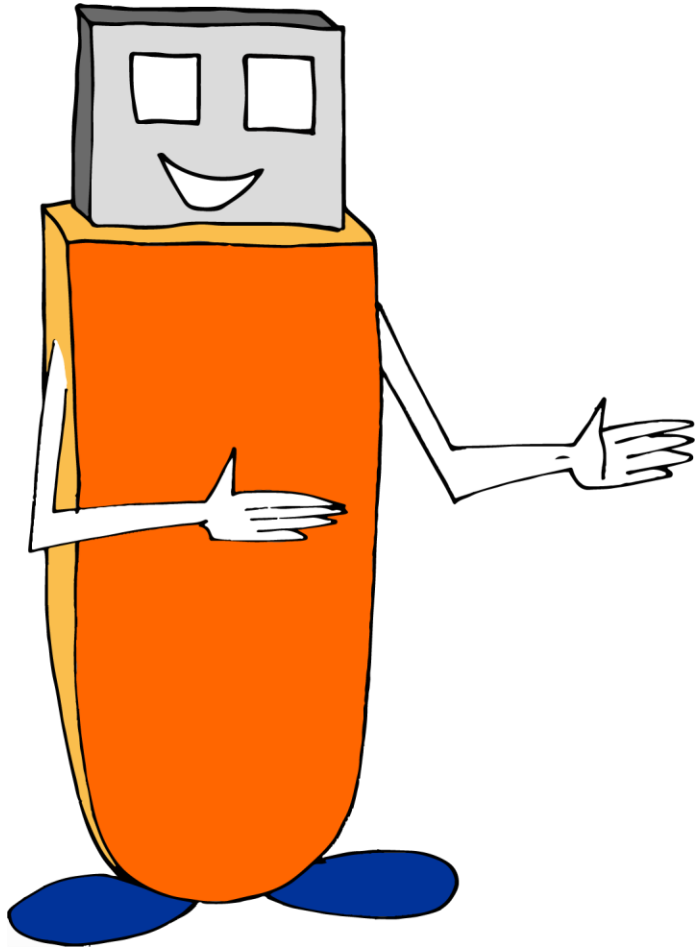


Citizen science - immaterieel erfgoed

- 1) het opstellen van een sjabloon en beschrijvingsregels om data en informatie over vrijzinnig humanistische praktijken, edities en onderdelen van edities te registreren
- 2) test in vorm van opdracht voor studenten i.h.k.v. het vak 'Archiefpraktijken'
- 3) traject met 3 vrijwilligers



Documentatie

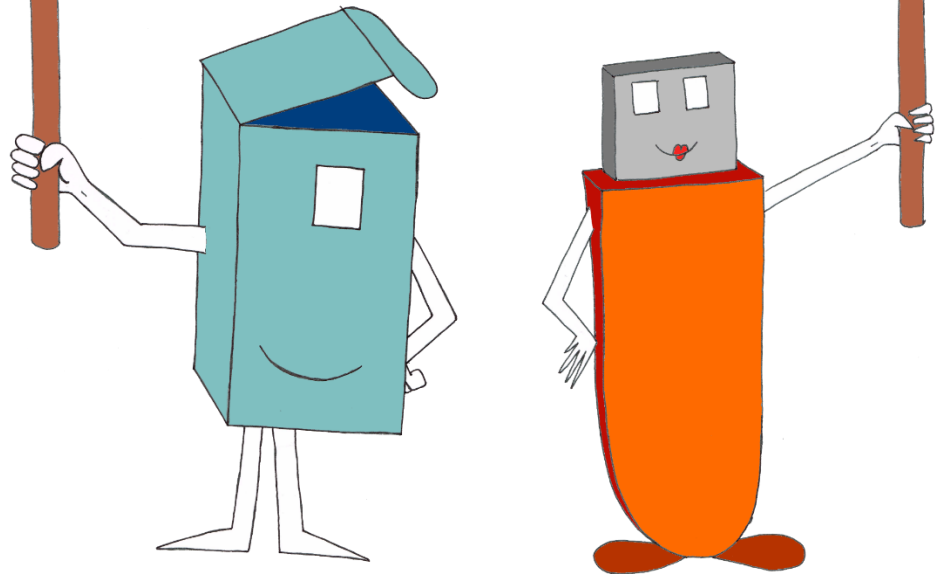


<https://cavavub.be/nl/ichich>

<https://www.projectcest.be/wiki/Publicatie:ICHICH>. Naar een verhoogde toegankelijkheid van digitale tijdschriften



Bedankt!



elise.laure.dewilde@vub.be
ellen.maria.soetens@vub.be

