

Schatgraven in de statistieken

De website met historische CBS-publicaties is een half jaar online. Tijd voor de eerste gebruikservaringen en een blik op de toekomst.
Margriet van Stiphout

De historische collectie van de bibliotheek van het CBS (Centraal Bureau voor de Statistiek) bestaat uit zo'n 5 miljoen handgeschreven en gedrukte pagina's met statistische tabellen en bijbehorende toelichtingen. Al deze documenten worden stapsgewijs gedigitaliseerd en beschikbaar gesteld via cbs.nl/historisch. De collectie wordt hiermee toegankelijk voor een groter publiek en is tegelijkertijd beter doorzoekbaar dankzij onder andere full-text zoekmogelijkheden. Tevens kunnen tabellen gemakkelijker overgenomen worden.

Positieve ervaring

Dr. J(eroen) J. Euwe, werkzaam bij de Universit  Libre de Bruxelles, maakt voor zijn onderzoeken regelmatig gebruik van de CBS-collectie. Zo komen internationale handelsstatistieken, die teruggaan tot 1846, goed van pas bij zijn onderzoek naar de Nederlandse kunstmarkt in crisistijden in de eerste helft van de twintigste eeuw. Hij schetst grote voordelen van het digitaal brengen. Zo verkort de digitalisering de voorbereidingstijd van zijn onderzoek, doordat hij niet meer stad en land hoeft af te reizen



Het projectteam Digitalisering Historische CBS-publicaties v.l.n.r.: Henk Laloli, Margriet van Stiphout (projectcoördinator) en Alfred Moest foto Sjoerd van der Hucht

om archieven te raadplegen. Ook ervaart hij een drastische vermindering van papiergebruik en enorme ruimtewinst in zijn werkkamer, nu hij geen papieren kopieën hoeft te bewaren. De mogelijkheid tot delen van informatie met collega's via links en/of afbeeldingen ervaart hij ook als positief.

Digitaal hergebruik

Ook benadrukt Jeroen het digitaal

hergebruik van statistische tabellen, ook al is de techniek hiervoor nog niet optimaal. Verder roemt Jeroen het meedenken en persoonlijke advies van de medewerkers van de bibliotheek, die de collectie kennen en hem waardevolle informatie geven voor zijn onderzoek. Dat zal hij wel gaan missen als de volledige collectie over enkele jaren digitaal is.

Er lopen verschillende onderzoek-

ken naar de mogelijkheden om de data van bijvoorbeeld de volkstellingen als zogenoemde open en linked databestanden aan te bieden. Dat zou betekenen dat geïnteresseerden niet meer alleen de tabellen kunnen raadplegen, maar ook de afzonderlijke data kunnen gebruiken. Dat biedt nog meer mogelijkheden voor onderzoekers en andere gebruikers van CBS-data.
cbs.nl/historisch

COLUMN

Mijn eigen digitale kwaliteitskrant

In de loop van de tijd ben ik kranten steeds selectiever gaan lezen. Bestede ik vroeger zeker een uur per dag aan mijn krant, tegenwoordig vlieg ik er vaak binnen twintig minuten doorheen. Het belangrijkste nieuws heeft me in de loop van de dag al bereikt. Gehoord op de radio of gelezen op internet of op mijn smartphone. Laatst moest ik een stapel vakantiekranen bijlezen – de krant onderweg downloaden was niet gelukt. Met zoveel kranten voor je neus lees je nog selectiever, maar hoe dan ook vond ik het confronterend om te merken hoeveel bladzijden ik ongelezen omsloeg. Hele bijlagen legde ik ongelezen opzij. Al dat tijdrovende en zinloze geblader maakte fysiek duidelijk wat ik al langer vaststel: ik vind papieren kranten in een rap tempo uit de tijd raken. Nou horen we al langer dat de papieren krant op sterven na dood is, en inderdaad: de cijfers

geven een zeer somber beeld. In de VS verdween sinds 2007 ruim dertig procent van de banen in de krantenindustrie. En maar liefst 286 lokale kranten sloten daar hun deuren. Ook bij Nederlandse kranten gingen de afgelopen jaren duizenden banen verloren. Wat mij zo verbaast, is dat kranten nauwelijks lijken in te springen op een trend die nu al jaren zichtbaar is: personalisatie. Op internet kun je zo'n beetje alles op maat laten maken, maar veel digitale kranten zijn niet meer dan een replica van de papieren krant. Je kunt er niet doorklikken, niks highlighten en je kunt 'm niet aan je eigen wensen aanpassen. Wat ik zou willen is dit: een digitale betaalkrant voor op mijn pc, smartphone en tablet, met stukken uit de beste Nederlandse en buitenlandse kranten over onderwerpen en thema's die ik zelf kan bepalen door ze aan of uit te vinken. Er mag een algemeen gedeelte bij zitten, een keuze van de redactie om mij te verrassen,

maar verder bepaal ik graag zelf welke columnisten ik lees, welke bijvoegsels ik kan missen en in welke onderwerpen ik mij echt wil verdiepen. De voordelen lijken mij evident: geen hoge drukkosten meer, geen distributieproblemen, maar een flexibele, snel te actualiseren krant, die ik via mijn persoonlijk profiel doorlopend op maat kan snijden. Niet te dik, meer analyses en achtergronden dan nieuws, want voor veel nieuws ben ik al tevreden met de gratis nieuwsdiensten op internet. Ik zou willen kunnen highlighten, waar mogelijk wil ik worden verwezen naar digitale bronnen of dossiers



over hetzelfde onderwerp, ik zou graag makkelijk kunnen zoeken in het archief (geef mij alle stukken van de afgelopen drie maanden over het land waar ik nu ben of naartoe ga) en ik zou graag artikelen willen kunnen opslaan. Ik heb zelf niet zo'n behoefte om te reageren, maar ook dat is een trend, dus die mogelijkheid moet er zeker bij.

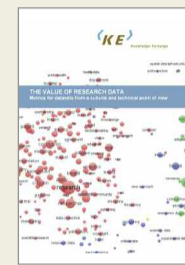
En ja, ik zou voor dit alles met liefde betalen. Waarschijnlijk net zo veel als voor mijn papieren krant, hoewel de beste op maat te snijden digitale krant niet van één krantenuitgever komt, maar per land of taalgebied van samenwerkende kwaliteitskranten. Het is te hopen dat de huidige kwaliteitskranten lang genoeg blijven bestaan om dit te realiseren.

Ewoud Sanders
Taalhistoricus en journalist.
Sanders is medewerker van NRC Handelsblad en Onze Taal.

foto Leo van Velzen

GELEZEN

Report 'The Value of Research Data. Metrics for datasets from a cultural and technical point of view' Rodrigo Costas, Ingeborg Meijer, Zohreh Zahedi and Paul Wouters of the Center for Science and Technology Studies (CWTS).
Leiden University
René van Horik



Gegevensuitwisseling is belangrijk voor wetenschappelijke vooruitgang en kennisbevordering. Er zijn echter nog steeds beper-

kingen en belemmeringen waardoor deze uitwisseling niet massaal plaatsvindt. Dit rapport geeft een aanzet om hier verbetering in aan te brengen. Centraal in het rapport staat een beloningssysteem op basis van datametrics, het stimuleren van het delen van onderzoeksdata. Ook behandelt het rapport de kosten van gegevenspublicatie. Het rapport adviseert een standaard te ontwikkelen voor het opslaan, publiceren, identificeren en citeren van datasets. Het rapport verkent de mogelijkheden om metrics toe te passen op datasets om een effectief beloningssysteem op te zetten. Dit beloningssysteem moet rekening houden met de belangen van degenen die bij het proces betrokken zijn. De publicatie bevat verder een overzicht van de huidige literatuur op het gebied van gegevensuitwisseling en data metrics. Ook de resultaten van gesprekken met de belangrijkste betrokkenen zijn opgenomen. Volgens zijn bestaande repositories en tools op het gebied van gegevensuitwisseling in kaart gebracht. Deze publicatie is interessant voor belanghebbenden in de wetenschap, voor onderzoeksfinanciers en wetenschappelijke instellingen. Het biedt inzicht in het instrumentarium dat beschikbaar is voor het stimuleren en belonen van gegevensuitwisseling en -publicatie binnen de eigen academische gemeenschap.

knowledge-exchange.info/datametrics

KORT

Nieuwe sponsor voor E-data&Research

De bestuurs- en redactieleden van dit blad verwelkomen het RIVM (Rijksinstituut voor Volksgezondheid en Milieu) als nieuwe sponsor van E-data&Research. Dit Nederlandse overheidsinstituut verzorgt informatie, monitoring en wetenschappelijke onderbouwing van het volksgezondheidsbeleid van de Nederlandse overheid. Wilt u ook meer weten over het sponsoren van dit blad? Kijk voor meer informatie op de website www.edata.nl of neem contact met ons op via edata@dans.knaw.nl.
(HB)
rivm.nl