



## E-DATA & RESEARCH

Jaargang 13 | nummer 2

Nieuwsbrief over data en onderzoek in de alfa- en gamma-wetenschappen.

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: CentERdata, CLARIAH, DANS, Huygens ING, de Koninklijke Bibliotheek en het Rijksmuseum.

## INHOUD



- 3** Van de Sompel: volg het online artefactenspoor
- 5** OpenINTEL, BBMRI-omics en PAN winnaars Dataprijis
- 6** ArtLives: digitale verrijking kunsthistorische publicaties
- 7** Digitale vaardigheden steeds belangrijker
- 7** In gesprek met jong talent Alex Brandsen over AGNES
- 8** Gastcolumnist Oberski over differential privacy

E-data wordt gratis toegezonden aan relaties van de stakeholders. Ook een uitgave ontvangen? Mail de redactie: [edata@dans.knaw.nl](mailto:edata@dans.knaw.nl).



Scan deze QR-code met een smartphone om de website van E-data te bezoeken.  
[edata.nl](http://edata.nl)

Fries-Nederlandse spraakherkenner ontwikkeld

# Speuren in de archieven van Omrop Fryslân

**Sinds afgelopen najaar zijn de radio-archieven van Omrop Fryslân digitaal doorzoekbaar, dankzij een spraakherkenner die zowel Fries als Nederlands herkent.**

*Erica Renckens*

Ruim 3.000 uur aan oude analoge radio-opnames van Omrop Fryslân zijn online te doorzoeken. De Friese omroep riep hiervoor de hulp in van spraak- en taaltechnologen van de Radboud Universiteit. Zij ontwikkelden binnen het project FAME een Nederlands-Friese spraakherkenner.

### Code switching

“In het Fries wordt veel geswitcht met het Nederlands”, vertelt projectleider Henk van den Heuvel. “Dat is erg lastig voor een spraakherkenner.” Om ervaring op te doen met dit zogenaamde ‘code switching’, bezocht één van de projectleden, spraaktechnoloog Emre Yilmaz onder andere Zuid-Afrika, waar het Afrikaans is doorspekt met Engelse woorden. “Je kunt het probleem grofweg op twee manieren benaderen: je maakt één herkenner voor beide talen of je stelt eerst van elk woord de taal vast en herkent dat

vervolgens. Uit Emres onderzoek bleek de eerste aanpak in ons geval het beste te werken.” De Friese spraakherkenner heeft nu een error rate van zo’n 23 procent. “Dat is een stuk lager dan we bij aanvang van het project verwachtten. Het lijkt misschien nog steeds hoog, maar het is laag genoeg om in het archief relevante fragmenten te kunnen vinden.”

Van den Heuvel verwacht dat de spraakherkenner ook buiten het project gebruikt zal worden. “Omdat het systeem is ontwikkeld voor omroepmateriaal, is het vocabulaire vrij breed. Onlangs heeft iemand het getest voor een zorgrobot en dat leek veelbelovend uit te pakken.”

### Doorontwikkeling

De zoekmachine is opgeleverd in drie versies voor verschillende doelgroepen: het grote publiek, journalisten en onderzoekers. “De journalisten kunnen de zoekresultaten ook downloaden. In de versie voor de onderzoekers is daarnaast ook sprekerherkenning beschikbaar. Het systeem clustert de spraak per spreker en probeert deze vervolgens te linken aan een van de 336 stemprofielen. Die zijn van mensen die regelmatig in de opnames terugkomen, zoals presentatoren.”

Hoewel het project inmiddels is afgerond, zal het zoekstelsel nog worden doorontwikkeld. Van den



*Radio-opnames op analoge banden en taperecorder van Omrop Fryslân. Inmiddels is het mogelijk de opnames digitaal terug te luisteren en te doorzoeken. credits Omrop Fryslân*

Heuvel: “Momenteel is alleen nog het analoge archief doorzoekbaar, met materiaal tot het jaar 2000. De private partners uit het project – Omrop Fryslân, Tresoar en Gridline – zullen ook nog het latere materiaal

doorzoekbaar maken, dat is van oorsprong al digitaal. De spraakherkenner zelf wordt daarvoor ook bij Tresoar ondergebracht.”

Wie zelf wil zoeken in het materiaal kan terecht op: [zoeken.fame.frl](http://zoeken.fame.frl)

Rijksmuseum sluit zich aan bij Linked.Art consortium

## Erfgoeddata van LOD naar LOUD

**In Linked.Art werkt een internationaal consortium aan de verbetering van de bruikbaarheid van Linked Open Data voor erfgoedobjecten. Onlangs sloot ook het Rijksmuseum zich aan.**

*Maarten Heerlien*

Afgelopen maand is het Rijksmuseum als eerste Nederlandse museum toegetreden tot het internationale consortium van 23 erfgoedinstellingen en onderzoeksinstituten uit Noord-Amerika, Europa en Azië dat uitvoering geeft aan het Linked.Art-project. Het consortium wordt geleid door Robert Sanderson, semantisch architect bij de J. Paul Getty Trust. Voor het Rijksmuseum biedt Linked.Art mooie kansen, aldus Saskia Scheltjens, Hoofd Research Services bij het Rijksmuseum: “Deze samenwerking kadert binnen de Memorandum of

Understanding tussen het Rijksmuseum en het Getty. Het geeft ons de gelegenheid om verder te bouwen aan onze expertise en deze kennis maximaal te delen met anderen op een internationaal niveau.”

### Bruikbaarheid vergroten

Doelstelling van het Linked.Art-project is om tot een concrete, op gebruikers gerichte toepassing te komen van CIDOC-CRM, het Conceptual Reference Model voor beschrijving en uitwisseling van informatie over erfgoedobjecten van de International Council of Museums (ICOM). Hoewel CIDOC-CRM al sinds 2006 een ISO-standaard is, blijft implementatie ervan door erfgoedinstellingen achter. Chris Dijkshoorn, databeheerder bij het Rijksmuseum en betrokken bij het ontwikkelteam van het project: “CIDOC-CRM is een expressief, maar zeer theoretisch model. Data is

met het model op verschillende manieren vast te leggen, wat de concrete toepassing van CIDOC-CRM hindert. Linked.Art richt zich op het bereiken van consensus over de toepassing van het model door middel van een zogeheten applicatieprofiel. Daarbij worden alleen de elementen uit CIDOC-CRM gebruikt die voor specifieke use-cases relevant zijn. Uiteindelijk vergroot dat de bruikbaarheid van cultureel erfgoed data.”

### Linked Open Usable Data

Linked.Art werkt dit applicatieprofiel uit volgens de principes van LOUD, Linked Open Usable Data. LOUD, een model van Sanderson, is een aanvulling op Tim Berners-Lee’s vijfsterrenmodel voor Linked Open Data. Berners-Lee’s model is gericht op dataproviders en minder op datagebruikers, waardoor vijfsterren datasets vaak onderbenut

blijven. LOUD richt zich op ontwikkelaars, de intermediërs tussen dataproviders en datagebruikers. Om de balans in datasets tussen bruikbaarheid en rijkheid te optimaliseren, formuleerde Sanderson vijf aanvullende sterren voor bruikbare Linked Open Data en evenzoveel ontwerpprincipes om daar concrete invulling aan te geven. Een belangrijke daarvan is de consequente uitdrukking van Linked Data in JSON-LD, een gebruiksvriendelijker alternatief voor RDF/XML.

De deelname van het Rijksmuseum aan Linked.Art, gecoördineerd door de afdeling Research Services, is een eerste stap in de intensivering van de samenwerking tussen deze afdeling en de afdeling Digital van de J. Paul Getty Trust. Saskia Scheltjens neemt namens het museum zitting in de projectstuurgroep. Linked.Art heeft een initiële doorlooptijd tot 2021.

[linked.art](http://linked.art)