



K O N I N K L I J K E N E D E R L A N D S E  
A K A D E M I E V A N W E T E N S C H A P P E N

## **FAIR Signposting biedt online navigatiehulp voor bots**

van de Sompel, Herbert

***published in***

E-data & Research  
2021

***document version***

Publisher's PDF, also known as Version of record

***document license***

CC BY

[Link to publication in KNAW Research Portal](#)

***citation for published version (APA)***

van de Sompel, H. (2021). FAIR Signposting biedt online navigatiehulp voor bots: Machinevriendelijke wegwijzers op en rond landingspagina's. *E-data & Research*, 16(1), 3. <https://edata.nl/2021/10/12/fair-signposting-biedt-online-navigatiehulp-voor-bots/>

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**E-mail address:**

[pure@knaw.nl](mailto:pure@knaw.nl)

Machievriendelijke wegwijzers op en rond landingspagina's

# FAIR Signposting biedt online navigatiehulp voor bots

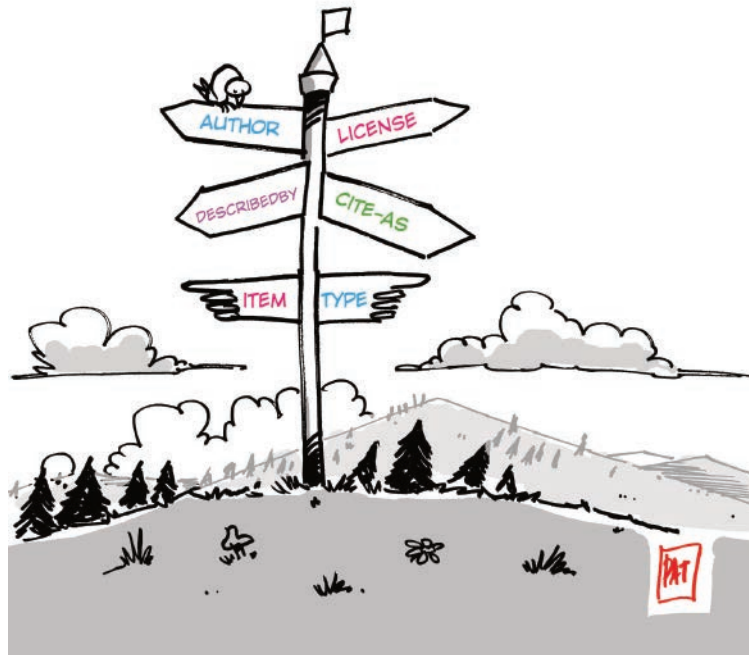
**FAIR Signposting biedt een online wegwijzer voor bots op zoek naar wetenschappelijke output. DANS realiseerde onlangs een implementatie die als inspiratie kan dienen voor andere wetenschappelijke repositories.**

Herbert Van de Sompel

Wie zich op het wetenschappelijke web begeeft, is vertrouwd met landingspagina's voor onderzoeksoutput. Ze komen in vele geuren en kleuren, het vinden van specifieke informatie op die pagina's is soms een uitdaging. Uiteindelijk wordt de link naar een artikel, dataset of een bibliografische citatie vaak wel gevonden. Met een beetje geluk wordt ook de Persistent Identifier (PID) gelokaliseerd om, in plaats van het webadres van de landingspagina, deze PID duurzaam naar het werk te laten verwijzen.

## Tallose links

Voor webbots is de situatie veel lastiger. Een bot ziet op een landingspagina tallose links en kan niet zomaar differentiëren naar welk soort informatie de links leiden en of ze nuttig zijn voor de taak die vol-



De implementatie van DANS kan als inspiratie dienen voor andere repositories. Credits: illustratie Patrick Hochstenbach

bracht moet worden. Diensten die gebruik maken van bots gaan hierdoor noodzakelijkerwijs heuristisch te werk en houden per type landingspagina bij welk soort informatie via welke link te vinden is. Een behoorlijke klus voor bijvoorbeeld referentiemanagers zoals Zotero en archiveringsdiensten zoals LOCKSS. En het is uiteraard een significante

drempel voor het creëren van nieuwe bot-gebaseerde diensten.

## Hulp bij navigatie

FAIR Signposting werd in het leven geroepen om dit probleem aan te pakken en bots te helpen bij het navigeren van het wetenschappelijke web. FAIR Signposting plaatst machinevriendelijke wegwijzers op en

rond landingspagina's. Technisch zijn dit links voorzien van een type-ring. Zo leidt de link met het type 'describedby' naar een bibliografische beschrijving, 'cite-as' naar de PID, en 'item' naar het artikel of de dataset.

## Dataverse

In het kader van het ODISSEI roadmap project implementeerde DANS 'FAIR Signposting' in Dataverse, een open source repository platform. Een aanpak met garanties voor stabiliteit op de lange termijn: de basale ingrediënten voor deze uniforme machinetoegang zijn links en link types gestandaardiseerd door de Internet Engineering Task Force. Door de implementatie van 'FAIR Signposting' zullen bots makkelijker hun weg vinden om en rond datasets in Dataverse installaties zonder gebruik te moeten maken van een specifieke API. Maar het wetenschappelijke web wordt pas écht machinevriendelijk als ook andere platforms 'FAIR Signposting' ondersteunen. DANS hoopt dan ook anderen te inspireren hetzelfde te doen voor hun wetenschappelijke repositories. Wees lief voor bots en zeg het voort! [signposting.org/FAIR](https://signposting.org/FAIR)

Centerdata deelt ervaring LISS panel

# Tips en tricks over werken met Speech to Text

**Wat zijn de voordelen en uitdagingen van audio-antwoorden in een online vragenlijst? Centerdata deelt de ervaring met spraaktechnologie Speech to Text.**

Joris Mulder

Het LISS panel, dat beheerd wordt door Centerdata, verzamelt sinds 2007 data door middel van online vragenlijstonderzoek. Panelleden beantwoorden de vragen door een antwoord aan te klikken of, in het geval van een open vraag, een antwoord te typen. Dit levert een scala aan waardevolle gegevens op voor onderzoekers en beleidsmakers.

## Twee experimenten

In aanvulling op deze klassieke methode van online dataverzameling kan het LISS panel ook *Speech to Text* in online vragenlijsten toepassen. Het afgelopen jaar zijn er twee experimenten uitgevoerd waarbij deze spraaktechnologie succesvol in online vragenlijsten werd ingezet. In deze experimenten konden respondenten hun open antwoorden in-



spreken in plaats van typen. De gegeven antwoorden zijn op twee manieren opgeslagen: als audiobestand en als transcriptie naar tekst. De winst van deze methode van dataverzameling is dat er, naast het primaire antwoord op de vraag, waardevolle aanvullende data worden verzameld. Denk bijvoorbeeld aan de toon van een respondent bij een antwoord, de linguïstische en verbale capaciteiten en de grootte van iemands woordenschat. Dit kan inzicht geven in het cognitief functioneren, de sociaaleconomische status of het verbaal redeneervermogen van een respondent. Tevens bieden

de data de mogelijkheid tot sentiment- of emotieanalyses (al dan niet gerelateerd aan een bepaald onderwerp waar mensen meer of minder binding mee hebben) en *Natural Language Processing* (het analyseren van natuurlijke taal met behulp van kunstmatige intelligentie).

## Aandachtspunten

Naast deze voordelen zijn er ook enkele nadelen en aandachtspunten. Zo zijn er soms technische beperkingen, bijvoorbeeld omdat een respondent geen werkende microfoon heeft of een te oude (niet geschikte) browser gebruikt. Ook wil een deel

van de panelleden überhaupt niet deelnemen aan deze vorm van dataverzameling. Hierdoor kan er (selectieve) non-respons ontstaan. Een belangrijk aandachtspunt is privacy. Een stem geldt als persoonlijke data en kan tot persoonsidentificatie leiden. Data die middels *Speech to Text* worden verzameld, dienen dan ook onder voorwaarden opgeslagen te worden. Het LISS panel zorgt er bovendien voor dat alle audiobestanden en transcripts door een medewerker worden gecontroleerd op eventuele persoonlijke informatie die men in het antwoord geeft (zoals een naam of plaatsnaam).

## Beschikbare data

De data van de *Speech to Text* experimenten komen eind 2021 beschikbaar in het LISS Data Archive en zijn te gebruiken voor eigen onderzoek. Voor meer informatie kan contact worden opgenomen met Joris Mulder, Survey Onderzoeker bij Centerdata, via [joris.mulder@centerdata.nl](mailto:joris.mulder@centerdata.nl).

[centerdata.nl](https://centerdata.nl)

## KORT

### Meld je aan voor ASReview Academy

Sneller systematisch literatuuronderzoek uitvoeren kan met ASReview. Deze software leert van gebruikers welke resultaten relevant zijn en verbetert hierdoor het zoekproces in de relevante databases. Om ASReview zo goed mogelijk te gebruiken, kunnen onderzoekers zich aanmelden voor een ASReview Academy, maandelijks georganiseerd in samenwerking met ODISSEI. Instromen kan elke maand, op verschillende niveaus (van beginners tot gevorderden) wordt in anderhalf uur steeds diepgaander kennisgemaakt met de software. (SZ) [odissei-data.nl/asreview-nl](https://odissei-data.nl/asreview-nl)

### Collecties als data bij KBR België

Met het project DATA-KBR-BE wil de KBR, de Koninklijke Bibliotheek van België, een verbeterde toegang op dataniveau van haar digitale collecties realiseren. Samen met de universiteiten van Gent en Antwerpen gaat de KBR onder andere de IT-infrastructuur optimaliseren en wordt een open data platform ontwikkeld. Hierlangs zullen de onderliggende databestanden van gedigitaliseerd erfgoed beschikbaar worden gesteld, wat als voedingsbodem dient voor onderzoek in de Digital Humanities. Met het project volgt de KBR het Amerikaanse Collections as Data Initiative ([collectionsasdata.github.io](https://collectionsasdata.github.io)). DATA-KBR-BE wordt gefinancierd door het Belgische instituut voor federaal wetenschapsbeleid BELSPO in het kader van het BRAIN 2.0 programma en loopt tot en met 2022. (MH) [kbr.be/nl/projecten/data](https://kbr.be/nl/projecten/data)

### Beurs voor OpenHuygens.nl

Het Onderzoeksfonds KNAW-instituten heeft het Huygens ING een beurs toegekend voor het versneld online overzetten van de collectie van het Huygens ING naar OpenHuygens.nl. Het Huygens ING beschikt over bijna tweehonderd resources op het gebied van Nederlandse letterkunde, geschiedenis en wetenschapsgeschiedenis. Samen vormen deze bronnen, teksten en datasets een fundamenteel onderdeel van de digitale infrastructuur voor de geesteswetenschappen. Met OpenHuygens.nl, een nieuwe data-omgeving die de collectie beter toegankelijk maakt voor innovatieve, digitale toepassingen, realiseert het Huygens ING duurzame toegang tot deze belangrijke bronnen voor onderzoek naar de Nederlandse geschiedenis en cultuur. De eerste versie van een werkend OpenHuygens.nl wordt in het voorjaar van 2022 gelanceerd. (MJ)