

Tussen schrappen en open data

De portal Haagse Feiten ontsluit parlementaire informatie die toegankelijk is via overheidskanalen, maar die daar moeilijk vindbaar is. Een interview met Ronald Baan, een van de initiatiefnemers van deze startup.



Ronald Baan

Rob Feenstra

Transparantie. Het was een woord dat wel heel vaak viel tijdens het Hoe Open? Festival, dat 12 december werd gehouden in Utrecht. Op deze bijeenkomst hielden de deelnemers, voor het merendeel ambtenaren, zich bezig met de vraag hoe je een open overheid in praktijk brengt. Natuurlijk gebeurt al veel op dat gebied. Zo zijn er inmiddels een kleine tienduizend open datasets beschikbaar via data.overheid.nl. Maar de vraag die geregeld naar voren kwam was ‘wie gebruikt die data nu eigenlijk?’ Soms is dat heel duidelijk. Buienradar is bijvoorbeeld waarschijnlijk de bekendste gebruiker van open data. Maar vaak blijkt er weinig tot geen contact te zijn tussen de overheid, als aanbieder, en de personen, instellingen en bedrijven die

de data gebruiken. Wie is die gebruiker en wat doet hij met de data?

Tijdens de middagpauze spreek ik met Ronald Baan, een van de initiatiefnemers van Haagse Feiten (haagsefeiten.nl). Deze portal ontsluit parlementaire informatie (wetsvoorstellen, kamervragen, stemmingsuitslagen enzovoort) die weliswaar al wel toegankelijk is via overheidskanalen, maar die daar moeilijk vindbaar is. Haagse Feiten wil meerwaarde bieden door alle data full-text te ontsluiten en daar diensten aan te koppelen.

Op mijn vraag van welke open data Haagse Feiten gebruikt maakt, begint Ronald Baan te lachen. ‘Eigenlijk maken we helemaal geen gebruik van open data,’ zegt Baan. ‘De gegevens van de Eerste en

Wat heeft de geïnteresseerde burger aan Haagse Feiten?

BASIC ABONNEMENT (gratis)

- full-text doorzoeken van wetsvoorstellen, kamerstukken enzovoort vanaf 1995
- full-text doorzoeken van tweets door parlementariërs
- Relevante informatie op volkskrant.nl, nu.nl, nos.nl
- Selecteren op personen (wat zegt Ronald Plasterk over open data)
- Zoeken op / toegang tot een groot aantal onderwerpdossiers

Wat heeft de professional aan Haagse Feiten?

PREMIUMABONNEMENT (99 euro per maand)

- alle functies van het basic abonnement
- full-text doorzoeken van de verschillende parlementaire agenda's
- inzicht in de activiteiten van de Haagse spelers
- gepersonaliseerde functies (e-mailtendering, notities aan documenten toevoegen, werken in teams, opbouwen en monitoren van onderwerpen naar keuze)

rugkoppeling op gegeven hebben. Maar bij Rijksoverheid.nl en de Eerste Kamer blijft het angstig stil.'

We leven in een webcultuur waarin mensen gewend zijn dat alle informatie gratis is. De overheidsdata die jullie aanbieden is al vrij toegankelijk. Gebruikers betalen eigenlijk alleen voor het verzamelen en bewerken van die data. Zijn ze daartoe bereid?

'We hebben natuurlijk vooraf de behoefte aan parlementaire data gepeild bij een groep professioneel geïnteresseerden, zoals journalisten, beleidsmedewerkers van ministeries en lobbyisten. Ze willen krachtenveldanalyses kunnen maken, ze willen direct weten wanneer een nieuw stuk wordt gepubliceerd en ze willen weten welke kamerleden zich bezighouden met een bepaald onderwerp. Daarbij is het vooral belangrijk dat de informatie snel en correct getoond wordt.'

'We hebben er bewust voor gekozen om Haagse Feiten voor iedereen toegankelijk te maken. Liever veel klanten voor een lage prijs dan een paar voor een hoge prijs. Onze klanten betalen 99 euro per maand. Voor professionele gebruikers is dat een goede prijs. Zij hoeven zich niet meer te bekommeren om het verzamelen van data, maar kunnen zich volledig richten op wat zij moeten doen: het interpreteren van die data.'

'Voor de geïnteresseerde burger is dat bedrag uiteraard te hoog. Daarom is er een gratis basisversie waarmee hij toegang heeft tot het volledige archief. Zo geven wij invulling aan onze maatschappelijke wens, namelijk democratie op meerdere wijzen een impuls geven.'

Vóór je geld binnenkrijgt van betalende gebruikers moet je ze eerst iets bieden. Hoe financieer je die aanlooperperiode?

'Het MKB & Technofonds Flevoland investeert in onze startup. Zo konden we onze basisstructuur opzetten. Het gaat daarbij niet alleen om full-text zoeken, we willen de informatie ook in context aanbieden.'

'We zitten nu in een stadium waarin we de binnenkomende abonnementsgelden gebruiken om meer diensten te ontwikkelen waar onze klanten om vragen. Zo hebben we in samenwerking met onderzoeksbureau Toponderzoek de Haagse Feiten Raadpleging ontwikkeld. Daarmee is het

Tweede Kamer zijn weliswaar openbaar, maar ze worden alleen gepubliceerd op een website. Die informatie moeten we zelf van de websites schrappen en omzetten naar gestructureerde data. Achter Haagse Feiten zit een team van zeven personen en vier daarvan zijn ontwikkelaars. De helft van onze ontwikkelcapaciteit gaat zitten in het correct en compleet binnenkrijgen van de data. De Officiële Bekendmakingen, die worden aangeboden op Overheid.nl, zijn ook toegankelijk via een ftp-site waar de verschillende versies van de documenten in een directorystructuur staan opgeslagen. Met wat fantasie kun je zeggen dat dat open data is, maar het is totaal niet te vergelijken met de gestructureerde datasets in data.overheid.nl.'

Is het niet vreemd dat er uitgerekend op parlementair niveau geen open data wordt aangeboden, terwijl er via data.overheid.nl bijvoorbeeld wel een uitgebreide dataset beschikbaar is van de lantaarnpalen in Schagen?

'Inderdaad, vooral als je bedenkt hoeveel organisaties er zijn die met eigen tools of handmatig de data proberen binnen te halen. Bovendien staan de verschillende parlementaire sites waar we gebruik van maken, volledig op zichzelf. Neem een simpel antwoord op een kamervraag. Dat verschijnt eerst op Rijksoverheid.nl. Een paar uur tot een dag later komt het op

tweedekamer.nl, weer een dag later op Officiële Bekendmakingen en als het voor de Eerste Kamer relevant is, verschijnt het daar ook op. Vaak wordt hetzelfde document dan ook nog in verschillende opmaak gepubliceerd. Daarbij komt dat de documenten vaak zo snel mogelijk worden gepubliceerd, en dat gebeurt nogal eens in rudimentaire vorm. Pas later verschijnt de officiële publicatie. Haagse Feiten moet op basis van tekstvergelijking proberen te traceren wat het verband is tussen al die documenten. Het zou enorm veel werk schelen als die gegevens als "echte" en goed gestructureerde open data werden aangeboden.'

Hebben jullie hierover contact met de mensen rond het parlement die zich hiermee zouden moeten bezighouden?

'Zeker, ik heb meer dan tien jaar voor de Tweede Kamer gewerkt. Er wordt daar al jaren gesproken over een Application Programming Interface (API) waarmee iedereen informatie kan ophalen. Uiteindelijk gaat die API er binnenkort komen. Dat is een goede zaak want veel data wordt daardoor beter toegankelijk, maar we weten nu al dat niet alles wat Haagse Feiten via de website ontsluit ook via de API zal worden aangeboden.'

'We hebben ook contact met Officiële Bekendmakingen. Zij krijgen binnenkort een nieuwe interface waar we al een te-

mogelijk om meer dan 10.000 openbaar bestuurders te benaderen met een vraag of idee. Zeker voor lobbyisten is dat heel nuttig. Verder denken we bijvoorbeeld ook aan een soort retrospectieve kieswijzer waarbij het verkiezingsprogramma wordt gelegd naast de uitspraken en het stemgedrag van de afgelopen vier jaar.'

Behalve parlementaire gegevens bieden jullie ook biografische informatie uit Wikipedia aan en alle tweets van alle kamerleden. Haalt een product dat 'Feiten' in de naam heeft staan daarmee niet subjectieve gegevens binnen?

'We willen zelf geen content beheren. Haagse Feiten presenteert alleen wat er is. Bij de biografieën staat Wikipedia nadrukkelijk als bron aangegeven, dus dat is voor onze klanten duidelijk. We gaan ervan uit dat de informatie redelijk betrouwbaar is. Doorgaans trekt een minister of kamerlid zelf aan de bel als Wikipedia aperte onjuistheden bevat. En voor de tweets geldt natuurlijk dat het duidelijk is dat het gaat om de mening van de betreffende twitteraar. En overigens is het mogelijk om de twitter-zoekfunctie aan of uit te zetten. Maar die tweets, het zijn er inmiddels zo'n 400.000, geven een mooie verrijking bij de informatie die we al aanbieden.'

Hoe zie je de toekomst van Haagse Feiten? Verwacht je dat je het makkelijker of zwaarder krijgt wanneer er in jullie segment meer informatie beschikbaar komt als open data?

'We zijn nu ruim een half jaar aan de gang en we zijn nog volop bezig om ons verder te ontwikkelen. Het aantal klanten stijgt gestaag, maar we streven naar meer groei. Wanneer de overheid haar data meer open en beter gestructureerd aanbiedt, zullen wij onze software daarop aanpassen. Die inspanning weegt ruimschoots op tegen de eenvoudigere en betere verwerking die dat oplevert. Ik ben niet bang voor meer concurrentie, want de essentie van het gebruik van open data zit niet zozeer in het binnenhalen ervan, maar wat je ermee doet, en ik denk dat we daar nu al een flinke slag in hebben gemaakt.'

Rob Feenstra is projectleider/consultant bij de Universitaire Bibliotheken Leiden en heeft als aandachtsgebieden bibliotheeksystemen en digitale bibliotheek.

VOGIN IP-lezing 2017

9 maart
OBA in
Amsterdam

Staat
in het teken
van:

web search &
enterprise search | social
search & business search |
semantic search & systematic search
| goed zoeken & slecht zoeken | deep
web & dark web | bloggen & vloggen
| taxonomieën & metadata | irc & nntp
| altmetrics & infopolitics | efficiëntie
& evidentie | signposting & memento
| webarchieven & mailarchieven
| factchecking & factfinding
| sharepoint & google |
analyseren & classificeren

BRAVE NEW SEARCH WORLD

vogin-ip-lezing.net