

Openbaarheid van bestuur kan vrijwel kosteloos beter

De opvolger van de *Wet openbaarheid van bestuur* (Wob) is de *Wet open overheid* (Woo); deze verplicht de overheid om documenten actief en open te publiceren. Dat lukt nu niet vanwege een verkeerde technische invulling van de Woo-dossiers en door de decentrale werkwijze. Een eenvoudige standaard, gebaseerd op de data-principes van FAIR, zou dit vrijwel kosteloos kunnen veranderen.

IN HET KORT

- Woo-dossiers voldoen niet aan de FAIR-principes van vindbaar, archiveerbaar, uitwisselbaar en herbruikbaar.
- Het invoeren van een standaard die naadloos in het werkproces van de Woo-jurist past kan de problemen oplossen.
- De extra kosten zijn marginaal, en de opbrengsten – zowel voor de maatschappij als voor de overheid – zijn groot.

MAIK LAROOIJ

Bachelor student aan de Universiteit van Amsterdam (UvA)

JAAP KAMPS

Universitair hoofd-docent aan de UvA

MAARTEN MARX

Universitair docent aan de UvA

Dit onderzoek is deels ondersteund door NWO via het ACCESS project onder nummer CISC.CC.016, en door de Universiteit van Amsterdam via Humane AI

Op 1 mei 2022 maakte de *Wet open overheid* (Woo) zijn intrede als vervanger van de *Wet openbaarheid van bestuur* (Wob). De nieuwe wet is bedoeld om de landelijke en lokale overheid transparanter te maken. De Woo regelt het recht op informatie over alles wat de overheid doet. Beide wetten stellen dat een burger een verzoek mag doen om informatie over ‘bestuurlijke aangelegenheden’, ofwel informatie over de voorbereiding en de uitvoering van het beleid van een bestuursorgaan. Nieuw in de Woo is de verplichting voor overheidsorganisaties om niet alleen op verzoek, maar ook zelf gefaseerd en actief informatie openbaar te maken.

De Woo is een zogenaamde *Freedom of* (nu vaker ook: *Access to*) *Information Act*, op dit moment van kracht in meer dan honderd landen. Worthy (2010) noemt zes veelgebruikte doelen van zo’n wet, waaronder het verhogen van de kwaliteit van beleidsmaatregelen en van het begrip daarvan bij de bevolking, en het verhogen van de deelname en het vertrouwen van burgers in de democratie. Deze wetten zijn ook een belangrijk instrument voor onderzoeksjournalisten, en zelfs voor Tweede Kamerleden om informatie boven water te krijgen. Naar het beleid rond covid werd er zo veel ‘gewobd’ dat het Ministerie van Volksgezondheid besloot de informatie direct te publiceren op een speciaal opgezette website wobcovid19.rijksoverheid.nl.

De recente wetswijziging verplicht de publicatie van de besluiten en de opgevraagde documenten, zowel door de centrale als door de lagere overheden, maar sleutelt niet aan de werkwijze van Wob-ambtenaren, noch aan de kwaliteit

van de dossiers. Maar al te vaak bestaan Wob en Woo dossiers uit ingescande, niet voor de computer leesbare, pdf-bestanden, zonder nuttige gestandaardiseerde metadata over het dossier of de afzonderlijke vrijgegeven documenten.

In dit artikel bestuderen we de *FAIRness* van de gepubliceerde dossiers en tonen we aan hoe onze openbaarheid van bestuur efficiënter, goedkoper en FAIRder kan. Hierbij keken we bewust niet naar de inhoud van de stukken, en dus niet naar de vraag of de informatiebehoefte, maar naar de technische kwaliteit van de Wob-besluiten, inclusief de vrijgegeven documenten. Dit artikel behandelt situaties die zowel op de Wob als op de Woo gebaseerd kunnen zijn. In dat geval spreken we van Wob/Woo-dossiers.

Drie voorwaarden voor doorzoekbaarheid

De vier principes zijn nogal abstract en niet eenvoudig te operationaliseren. Daarom beantwoorden we de vraag via een gedachte-experiment. We stellen ons voor dat we een zoekmachine – een gespecialiseerde Google of Woogle – willen maken voor Wob/Woo-dossiers. Vanzelfsprekend spelen de vier principes dan een grote rol. Een zoekmachine ordent enorme hoeveelheden documenten op hun relevantie, gegeven een zoekvraag. Dat ordenen gebeurt op basis van de metadata, en door de woorden in de vraag en in de documenten te vergelijken. Is het mogelijk om met de nu vrijgegeven Wob/Woo-dossiers zo’n zoekmachine te maken?

Onder een *Wob/Woo-dossier* verstaan we een bundel documenten met daarin het verzoek, het genomen besluit, de (inventaris)lijst met relevante documenten, en de daaruit, soms deels, vrijgegeven documenten. Met daarbij metadata over het dossier als geheel, en metadata per document. Zie figuur 1. Onze Woogle moet zowel hele dossiers als losse bestanden uit (vaak enorme) dossiers na een zoekvraag kunnen teruggeven, en natuurlijk goed geordend op relevantie wat betreft die vraag.

Om een zoekmachine te kunnen maken, moeten deze dossiers wel aan de volgende drie basisvoorwaarden voldoen. Ten eerste moet de logische informatie-eenheid overeenkomen met de technische bestands-eenheid. Google leidt je naar een Wikipedia-pagina, niet naar de hele Wikipedia. Het technische formaat van Wikipedia – elk lemma is een aparte pagina op het web – maakt dit mogelijk.

Ten tweede moeten de woorden in de documenten als woorden leesbaar zijn door een computer. Een situatie waarin dit niet het geval is, is wanneer men met Control-F zoekt naar een woord in een pdf-file en niks vindt, terwijl dat woord toch duidelijk op het scherm staat. Het tekstbe-

stand is dan feitelijk opgeslagen als een foto.

Ten derde moet er per informatiedrager een zekere minimale hoeveelheid metadata aanwezig zijn. Documenten vindbaar op Google hebben een titel, meestal een datum, een adres (de URL) en nog veel meer metadata, die de zoekmachine gebruikt om de resultaten op relevantie te ordenen en te presenteren.

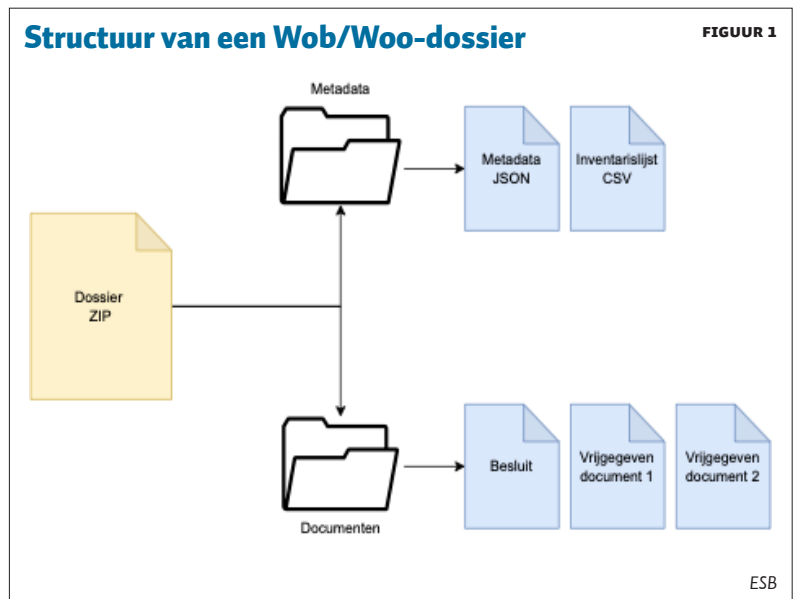
Wob/Woo-dossiers lastig doorzoekbaar

Het blijkt dat de Wob/Woo-dossiers lastig doorzoekbaar zijn voor een zoekmachine. Ten eerste voldoen de vrijgegeven Wob-documenten in overgrote meerderheid niet aan de eerste voorwaarde (de logische informatie-eenheid komt niet overeen met de technische bestands-eenheid). Het Wob/Woo-dossier blijkt in de praktijk vaak te bestaan uit drie pdf-bestanden: het besluit, de inventarislijst en een pdf met daarin alle vrijgegeven documenten achter elkaar geplakt, zonder voor de computer leesbare grenzen. Alle door *Open State* bekeken Wob-dossiers (alle 981 door ministeries gepubliceerde dossiers in de periode oktober 2020–september 2021) hebben deze vorm, evenals vrijwel ook alle dossiers vindbaar op het web die gepubliceerd zijn door lagere overheden. De gemeente Amsterdam en de provincie Gelderland vormen twee uitzonderingen. Zij plaatsen de verzameling documenten op een heel logische wijze in een *zip*-bestand.

Wij hebben de best beschikbare AI-technieken op basis van *machine learning* met neurale netwerken toegepast op het probleem van het automatisch weer opdelen in de oorspronkelijke documenten. Dit is niet foutloos te doen. Onze best presterende techniek had een pakkans van slechts vijftig procent bij een precisie van tachtig procent (Van Heusden et al., 2022).

Ten tweede blijken veel van de Wob-documenten scans van een print te zijn. De meeste scanners staan standaard zo ingesteld dat ze *optische tekenherkenning* toepassen, dus niet alleen een foto maken, maar ook zorgen dat de tekst voor een computer leesbaar is (bijvoorbeeld als Word-bestand). Bij de Wob-documenten gaat dit echter heel vaak niet goed. De vrijgegeven Wob-dossiers op *wobcovid19.rijksoverheid.nl* geven een goede indruk van dit probleem. In april 2022 bevatten deze 28.331 pagina's. Op bijna een kwart daarvan stond geen enkel voor de computer leesbaar woord, maar op vrijwel al die pagina's stond er wel tekst die met optische tekenherkenning zichtbaar gemaakt had kunnen worden. Het alsnog toepassen van een dergelijke herkenning, leverde in totaal meer dan een miljoen extra herkende woorden op. Op 77 procent van alle pagina's vond de optische tekenherkenning extra informatie die daarvoor niet voor de computer leesbaar was.

Ten derde blijkt de inventarislijst die bij bijna elk Wob-dossier zit, een tabel met op elke rij een document, en per kolom specifieke metadata voor elk document zoals de titel, het soort document (e-mail, whatsapp-bericht, Kamerstuk en dergelijke), hoe het is vrijgegeven, de eventuele weigeringsgrond, et cetera. Dit klinkt ideaal, en dat zou het inderdaad zijn als ten eerste die inventarislijsten als een Excel-bestand openbaar gemaakt werden en niet als een uitgeprinte en weer ingescande (en vaak onleesbare) tabel; ten tweede elke Wob/Woo-producent consequent



Aanwezigheid van standaard-attributen van Wob/Woo-dossiers, met gebruikte varianten (N = 346) TABEL 1

Attribuut	Aanwezig	Gebruikte varianten
Doc identifiër	75%	Nr; nummer; volgnummer; docnr; documentnr; id
Naam document	71%	Document; documentnaam; titeldocument; titeldoc; onderwerp; naamdocument; titel; naam
Beoordeling	67%	Beoordeling; beroordeling [sic!]; oordeel; beoordelingWob
Weigeringsgrond	68%	Weigeringsgrond; artikelWob; Wob; beslissingconform; Wobgrond; uitzonderingsgrond; artikel; Wobartikel; weigeringsgrondWob; weigeringsgronden; lakgrond; relevanteWobgronden; grond
Datum	51%	Datum; datumdocument
Soort	21%	Soort; soortdocument; type; categorie; documenttype; typedocument; soortstuk
Afzender	53%	Afzender; afzenders; van
Ontvanger	52%	Ontvanger; ontvangers; naar; aan

ESB

dezelfde namen voor de kolommen zou gebruiken, en ook consequent is in het benoemen van de waarden in de cellen; en ten derde alle Wob/Woo-producenten dat op dezelfde manier zouden doen. Jammer genoeg is dat niet het geval.

We hebben 2.703 Wob/Woo-dossiers opgehaald op *open.overheid.nl*. Bij slechts 436 stuks konden we daarbij op basis van de bestandsnaam een inventarislijst vinden, allemaal als tabel in pdf-formaat. Die konden 346 keer min of meer foutloos automatisch omgezet worden naar een spreadsheet. Tabel 1 toont hoe vaak we in deze 346 leesbare documenten basale metagegevens konden terugvinden, en op hoeveel verschillende manieren dezelfde informatie wordt weergegeven in de inventarislijsten.

Omdat de data op zo'n onhandige en niet-uniforme manier worden aangeleverd, is het opzetten van een zoekmachine een enorm lastig karwei. Met moderne technieken uit de kunstmatige intelligentie kunnen we een heel eind komen om de dossiers bruikbaar te maken, maar dit proces is nooit foutloos.

Makkelijk te verbeteren

Het kan gelukkig ook anders, door het probleem bij de bron aan te pakken. Wij hebben de indruk dat ambtenaren hun Wob/Woo-dossiers prachtig FAIR op hun eigen schijf hebben staan, maar dat er in de laatste publicatiestap iets misgaat. Want wat is er nou eigenlijk nodig om die dossiers FAIR te publiceren? De documenten digitaal (en dus machine-leesbaar) in een (zip-)mapje, en de metadata op een uniforme wijze in een spreadsheet, via unieke codes (lieft zogenaamde *permalinks* vergelijkbaar met een DOI) gekoppeld aan de losse documenten. Eigenlijk precies zoals in figuur 1.

Het lastigste hier is dat er heel veel Woo-ambtenaren zijn met ieder hun eigen werkwijze. Dus de eis van uniforme metadata is een coördinatieprobleem. Wij hebben dat simpel opgelost door gratis opensourcesoftware beschikbaar te stellen waarmee men heel handig een Woo-dossier opbouwt en automatisch uniform en FAIR publiceert (software staat op https://github.com/wooverheid/fair_woo_cms). De toegekende metadata zijn gebaseerd op voorstel door Open State, de provincie Noord-Holland en de VNG (Open State et al., 2021), en van de in de Woo vastgestelde verplichte metadata (KOOP, 2022).

Wob/Woo-dossiers kunnen dus vindbaar en archiverbaar gemaakt worden door ze machine-leesbaar en met uniforme metadata in een open formaat op open.overheid.nl te publiceren. Uitwisselbaarheid en herbruikbaarheid kunnen eenvoudig gegarandeerd worden door iedere Woo-ambtenaar dezelfde Woo-publicatiesoftware te laten gebruiken. Woo-ambtenaren gebruiken al speciale software voor hun Woo-dossiers, vooral om semi-automatisch persoonsgegevens te herkennen en die zwart te lakken (het is dus niet nodig om hiervoor documenten te printen en in te scannen).

Maatschappelijke opbrengsten

De extra kosten om de dossiers FAIR te publiceren zijn verwaarloosbaar, zeker in verhouding tot de kosten nodig voor het vinden en anonimiseren van de stukken. De opbrengsten kunnen erg hoog zijn.

Berenschot adviseert om Wob-documenten netjes op te slaan en voor de Wob-ambtenaar vindbaar te maken zodat ze herbruikbaar zijn voor een nieuw Wob-verzoek (en dus niet weer opnieuw geanonimiseerd hoeven te worden) (Enthoven et al., 2021). Prachtig natuurlijk, maar de winst lijkt in het niet te vallen bij de winst die te behalen is als burgers eenvoudig en goed in gepubliceerde Wob/Woo-dossiers kunnen zoeken en hun vraag beantwoord zien, zodat ze die dus niet opnieuw als Woo-verzoek hoeven te stellen. Niet alleen zou dit de overheid vele werkuren besparen, ook onderzoekscollectieven als *Follow the Money* zouden een hoop geld (en frustratie) besparen. Zij zijn een groot deel van hun tijd kwijt met het *reverse-engineeren* van de Wob/Woo-dossiers, waarbij precies de drie voorwaarden voor doorzoekbaarheid het struikelblok vormen voor het opbouwen van een bruikbaar (FAIR) digitaal dossier door zo'n journalist of collectief.

Tenslotte zijn de opbrengsten van open data en standaarden potentieel zeer groot, zoals ook aangestipt in de recente EU Data Act voorgesteld door de Europese Com-

missie. De Commissie verwacht dat de herbruikbaarheid van data 280 miljard Euro aan extra BNP zal opleveren in 2028. Precies hierom zet de EU met de hergebruik richtlijn 2019/1024 onze overheden aan om de hergebruik van overheidsinformatie zo eenvoudig mogelijk te maken. De laatste richtlijn is sinds 17 juli 2021 ook in Nederland van toepassing. Dit artikel toont aan dat de implementatie ervan nog te wensen overlaat, maar redelijk gemakkelijk beter kan.

Conclusie

De op 1 mei 2022 ingegane opvolger van de Wob, de Wet open overheid (Woo), moet, als ze haar naam eer wil aandoen, een omslag maken in de manier waarop de opgevraagde stukken openbaar worden gemaakt. Dat kan simpel en vrijwel kosteloos. De benodigde verandering is driedelig: publiceer Woo-dossiers echt als dossiers, maak ze echt machine-leesbaar, en zorg voor rijke uniforme metadata. De potentiële opbrengsten zijn groot, zowel economisch als maatschappelijk.

Aangezien de Woo-dossiers decentraal gemaakt worden, is een zekere mate van sturing noodzakelijk. Dat kan van bovenaf, door de Minister van Binnenlandse Zaken, of van onderop, door een autonoom *preferential attachment*-proces (*zwaan-kleef-aan*), gedreven door de wens om goed zichtbaar te zijn in zoets als een *Woogle* of een *Woobipedia* (Barabási en Pósfai, 2016).

Het invoeren van een eenvoudige standaard gebaseerd op de FAIR principes zou dus zorgen voor een efficiëntere, goedkopere en toegankelijker openbaarheid van bestuur

Literatuur

- Barabási, A.-L. en M. Pósfai (2016) *Network science*. Cambridge, VK: Cambridge University.
- Enthoven, G., H. Spanninga, C. Pino en A. Spruit (2021) *Verbeterpunten in de informatiehuishouding voor een tijdige en kwalitatief goede afhandeling van Wob-verzoeken*. IMI/Berenschot Rapport, 2 april. Te vinden op www.informatiehuishouding.nl.
- Heusden, R. van, J. Kamps en M. Marx (2022) *WooIR: A new open page stream segmentation dataset*. In: *ICTIR '22: Proceedings of the 2022 ACM SIGIR International Conference of Theory of Information Retrieval*. New York: Association for Computing Machinery, p. 24–33. Te vinden op dl.acm.org.
- KOOP (2022) *Documenten en metadata aanleveren voor actieve openbaarmaking via PLOOI (vo.3.2-alpha)*. Informatie te vinden op koop.gitlab.io.
- Open State, Provincie Noord Holland en VNG (2021) *Handreiking Open Wob: Wob-informatie publiceren volgens de Wobstandaard zodat iedereen er kennis van kan nemen*. Te vinden op kia.pleio.nl.
- Wilkinson, M.D., M. Dumontier, I.J.J. Aalbersberg et al. (2016) *The FAIR Guiding Principles for scientific data management and stewardship*. *Scientific Data*, 3(1), artikelnr. 160018.
- Worthy, B. (2010) *More open but not more trusted? The effect of the Freedom of Information Act 2000 on the United Kingdom Central Government*. *Government*, 23(4), 561–582.