

## Een onderzoek naar de stand van zaken

# Digitaal duurzaam publiceren van Woo-dossiers

**In hoeverre publiceren de provincies en het Rijk hun Woo-dossiers op een digitaal duurzame wijze? Er is grote vooruitgang in het openbaar maken van Woo-dossiers, maar er zijn ook flink wat beperkingen in de manier waarop de dossiers worden gepubliceerd. We zien duidelijk verschillen tussen de provincies. Vooral nog lukt het geen enkele provincie om aan alle minimale digitale duurzaamheidseisen te voldoen. Door het adopteren van *best practices* is er veel winst te behalen, meestal door middel van eenvoudige aanpassingen zonder noemenswaardige extra kosten.**

Sinds 1 mei 2022 is de Wet open overheid (Woo) van kracht. Deze wet verplicht overheden om een breed scala aan documenten, waaronder hun reacties op gerichte verzoeken om informatie gedaan door burgers, de vroegere Wob-verzoeken, actief te publiceren. Die reacties bestaan uit het verzoek, het genomen besluit, de (gedeeltelijk) vrijgegeven documenten en een inventarislijst daarvan. We noemen dit geheel een *Woo-dossier*. Er zijn in Nederland meer dan duizend bestuursorganen<sup>1</sup> die aan zo'n verzoek om informatie moeten voldoen. Slechts een deel daarvan publiceert de afgehandelde Woo-dossiers nu al: alle ministeries, via open.overheid.nl, vrijwel alle provincies en steeds meer gemeenten.

Eerder onderzoek heeft zich vooral gericht op het niet, of uiterst traag, voldoen aan Woo-verzoeken.<sup>2</sup> Wij analyseren de *digitale duurzaamheid* van de gepubliceerde Woo-dossiers, waarbij we in het bijzonder kijken naar de *herbruikbaarheid* van de verstrekte informatie. Het belang en grote potentieel van herbruikbare overheidsinformatie wordt sterk benadrukt in de EU-richtlijn Hergebruik,<sup>3</sup> in Nederland van kracht sinds juli 2021. Deze richtlijn, de twee eerdere EU-richtlijnen die ook worden aangehaald in artikel 2.4 lid 3 van de Woo, dat lid en de DUTO-eisen vastgesteld in de Archiefwet vormen de juridische meetlat waarmee de digitale duurzaamheid kan worden vastgesteld. In essentie zijn al die bepalingen nadere preciseringen van de vier FAIR-principes: overheidsinformatie moet vindbaar, toegankelijk, koppelbaar en herbruikbaar zijn.<sup>4</sup>

We analyseren drie kernonderdelen van digitale duurzaamheid: machineleesbaarheid, metadata en export. De motivatie voor dit onderzoek kwam voort uit een sterke wens naar hergebruik. We wilden alle via de Woo en Wob opgevraagde documenten via één zoekportaal, een *Woog*, snel en makkelijk beschikbaar maken voor iedereen. Om zo'n zoekportaal te maken heb je die documenten nodig (export), je moet ze automatisch kunnen verwerken (machineleesbaarheid), ze plaatsen in de juiste context (metadata), en ernaar kunnen verwijzen. We zullen zien dat dit met flink wat kunst- en vliegwerk mogelijk is, maar niet optimaal, en zeker niet duurzaam. Ten slotte laten we zien dat het helemaal niet zo moeilijk is om de stukken volgens artikel 2.4.3 van de Woo te publiceren. Men

hoeft alleen de *best practices* die verschillende overheden al gebruiken, te combineren.

### Analyse

We beginnen met de positieve kant: er is ontegenzeggelijk grote vooruitgang in de openbaarheid van Woo- en Wob-dossiers sinds de invoering van de Woo. Negen van de twaalf provincies maken de stukken die zijn vrijgegeven na een Woo-verzoek openbaar. Daarnaast maakt nog één provincie alleen een samenvatting van het verzoek en het besluit openbaar. De periode waarover de stukken openbaar worden gemaakt, varieert enorm. De eerste dossiers komen uit eind 2015. Naarmate we dichterbij het heden komen, neemt het volume per jaar toe. Alle provincies die publiceren, publiceren de stukken van geheel 2022, behalve één, die net na de invoering van de Woo begint. Soms zijn de stukken van eerdere jaren op Archiefweb te vinden. Bij alle provincies die publiceren was via Google de pagina met de Woo-dossiers heel makkelijk en snel te vinden met de zoekvraag 'provincie NAAM Woo'. Altijd zat de juiste pagina bij de eerste vijf hits. Ook de pagina waar wordt uitgelegd hoe men een Woo-verzoek kan indienen, is heel snel te vinden.

Het publiceren van Woo-dossiers is een zeer belangrijke eerste stap, maar er blijkt nog grote winst te behalen in de manier waarop de dossiers worden gepubliceerd. Het begrip digitale duurzaamheid wordt in de Woo niet precies omschreven. Wel staan er in artikel 2.4.3a van de Woo twee kernbegrippen hiervan, *metadata* en *machineleesbaarheid*.

### Metadata

De minimale metadata van een Woo-dossier zijn een titel en een korte beschrijving, de datums van verzoek en besluit, de tekst van het gedane verzoek en het genomen besluit. In een zoekmachine kan dat er dan uitzien als in figuur 1.

Elke provincie die Woo-dossiers publiceert, geeft daar in ieder geval een titel (meestal tussen de drie en zes woorden) en een datum bij. In de meerderheid van de gevallen is (zonder het besluit te lezen) niet duidelijk om welke datum het gaat

## Wob-verzoek over bestemmingsplan en notulen raadsvergaderingen inz. uitrit Industrierrein Numansdorp Einsteinstraat met Rijksstraatweg (N488)

Wob-verzoek om notulen raadsvergaderingen over uitrit Industrierrein Numansdorp Einsteinstraat met Rijksstraatweg (N488). Bestemmingsplan van – over uitrit Industrierrein Numansdorp Einsteinstraat met Rijksstraatweg (N488).

ZI - Dossier na Wob/Wob-verzoek Verzoek: 2019-08-30 Besluit: 2019-10-03 Gemeente Hoeksche Waard

Hits: 1 2 3 4

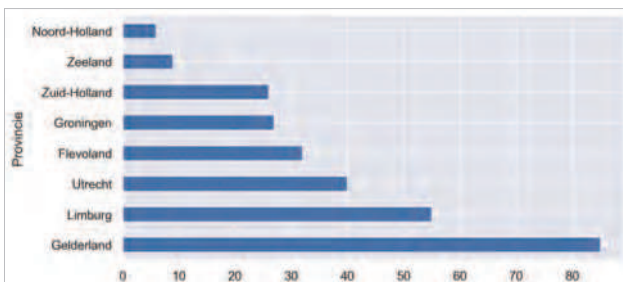
Figuur 1. Voorbeeld van een digitale duurzaam Wob-dossier.

(verzoek, besluit of publicatie). Twee provincies geven een korte, informatieve beschrijving. Twee provincies vermelden expliciet de datum van het besluit en ook twee die van het verzoek. Slechts één provincie geeft allebei. Ook twee provincies vermelden het gedane verzoek (in parafrase), en drie geven het genomen besluit in een paar woorden (gedeeltelijk openbaar, afgewezen, reeds openbaar et cetera). Geen van de tien provincies verstrekt dus al deze metadata. Maar belangrijker is dat voor elk van deze zes cruciale metadata er minstens twee provincies zijn die ze verstrekken.

### Machineleesbaarheid

Machineleesbaarheid gaat over twee zaken. Zijn de metadata machineleesbaar (in een vast omschreven gestructureerd technisch formaat)? En zijn de verzochte en vrijgegeven documenten machineleesbaar? Dit betekent minimaal dat: de woorden die een mens in een bestand ziet staan, ook door de computer gelezen kunnen worden. Dit is makkelijk na te gaan door met Control (of Command) F te controleren of een woord dat in de tekst staat ook gevonden wordt; de afzonderlijke documenten als losse bestanden worden verstrekt. Figuur 2 toont het percentage bladzijden per provincie waar geen enkel door de computer leesbaar karakter op staat.<sup>5</sup> Twee procent (zoals bij de gemeente Waalwijk en het ministerie van VWS) is normaal (dat zijn lege pagina's, pagina's met alleen plaatjes, of totaal zwartgelakte pagina's).

Zoals gezegd bestaat een Wob-dossier uit vier elementen: (a) het verzoek, (b) het besluit, (c) de inventarislijst, en (d) de vrijgegeven documenten. Het verzoek en het besluit zijn allebei brieven en de inventarislijst is een *spreadsheet* (Excel), met daarin voor elk relevant document een aantal eigenschappen.



Figuur 2. Percentage bladzijden zonder tekst in Wob-dossiers uit 2022 (N=49.664), uitgesplitst per provincie.

Het is een vreemde gewoonte binnen de Wob- en Wob-wereld om al deze documenten aan elkaar te plakken tot één (vaak enorm) pdf-bestand, en dat dan openbaar te maken. De grenzen tussen de losse onderdelen worden niet, of in ieder geval niet machineleesbaar, aangegeven. Voor de mens is het vaak al lastig te zien waar het ene stuk ophoudt en het volgende begint, voor de computer is dit een zeer lastig probleem, dat zelfs met de modernste neurale *deep machine learning*-technieken niet aanvaardbaar op te lossen is.<sup>6</sup>

Gelukkig kan het ook anders. Zeven van de tien provincies publiceren hun besluit los, drie provincies ook het verzoek, en twee provincies eveneens de inventarislijst (zij het jammer genoeg niet in Excel maar als pdf). In totaal drie van de tien provincies publiceren de vrijgegeven documenten heel handig in een mapje met daarin de losse bestanden, een zogenoemd zip-bestand.

De nadelen van aan elkaar geplakte bestanden zijn groot. Ten eerste is het veel lastiger voor een zoekalgoritme om te bepalen of een document relevant is voor een zoekvraag als het ingebed zit in allemaal andere documenten. Ten tweede is het voor een gebruiker enorm lastig om zelf het voor de zoekvraag relevante deel-document terug te vinden. Ten derde duurt het ophalen door de gebruiker van een (veel te) groot pdf-bestand erg lang, het kost veel energie en bandbreedte, en de pdf-lezer kan erg traag worden of zelfs niet meer reageren. Dit laatste euvel wordt nog erger als de pdf niet gewoon digitaal is, maar bestaat uit foto's van pagina's. Zo'n bestand is vele malen groter (in aantal bytes) en onhandelbaarder dan nodig.

### Beschikbaarheid

We noemen een bestand *beschikbaar* als het zonder tussenkomst van een mens is op te halen door de *crawler* van een zoekmachine. Een platform dat een actueel en volledig overzicht wil bieden van alle in Nederland gepubliceerde Wob-dossiers zal die dan, net als Google, regelmatig moeten ophalen en indexeren. Met meer dan duizend bestuursorganen die onder de Wob vallen kan dat alleen geautomatiseerd. De UvA heeft in de herfst van 2022 deze automatische toegankelijkheid van de Wob-dossiers bij de provincies bekeken.<sup>7</sup> De volgende punten maakten automatisch actueel houden erg lastig: de lijst met Wob-dossiers is niet uniform en de (html-)opmaak wisselt jaarlijks; automatisch downloaden wordt doelbewust tegengewerkt; bestanden staan op Google Drive of andere niet goed toegankelijke externe providers; en ten slotte nietszeggende bestandsnamen. »

» **Verwijsbaarheid**

Kunnen we duurzaam, dus met een *persistente identifier*, verwijzen naar een Woo-dossier? Geen van de provincies gebruikt persistente identifiers. Ook op het platform open.overheid.nl worden deze niet gebruikt. Wetten op <https://wetten.overheid.nl> en parlementaire stukken op <https://zoek.officielebekendmakingen.nl/> hebben die wel. Ook voor Woo-dossiers is zo'n verwijssysteem makkelijk op te zetten. De zoekmachine voor Woo-dossiers Woogle gebruikt handige hiërarchische *digital object identifiers* (doi), waarmee de naam en de locatie van een digitaal object losgekoppeld worden. Dit is ook nodig: overheden hebben de neiging documenten die ze op het web hebben gezet later weer weg te halen. Dit levert een ongewenste dode link op. Door niet direct naar een URL te verwijzen, maar de locatie te *resolven* vanuit de doi, kan in dat geval besloten worden om de lokale kopie te laten zien. De doi's bij Woogle hebben de volgende vorm: nl.bestuursorgaan.woo-categorie.jaar.rank, bijvoorbeeld nl.gm1963.2i.2022.22, is het 22e Woo-dossier (code 2i) uit 2022 van de gemeente Hoeksche Waard (TOOI-code gm1963). De doi van een document uit dit dossier heeft de doi van het dossier als *prefix*.

Het kunnen downloaden en up-to-date houden van een zoekplatform als Woogle of PLOOI is met zo'n *handle*-systeem ook meteen opgelost. De plek waar de doi's doorverwezen worden (*resolved*) kent alle actieve doi's en kan die dus openbaar maken. Iemand die alle nieuwe dossiers wil hebben, kan dus de eigen doi's vergelijken met de lijst en het verschil ophalen.

**Ons onderzoek toont aan dat er met relatief kleine aanpassingen nog veel te verbeteren is aan de manier waarop Woo-dossiers worden gepubliceerd**

Zo'n overzicht is verplicht gesteld in artikel 9, lid 1 van de EU-richtlijn 2019/1024.

**De ministeries**

Hoe verhouden de gepubliceerde Woo-dossiers van de provincies zich tot die van de rijksoverheid? Woo-verzoeken afgehandeld door een ministerie worden gepubliceerd op open.overheid.nl, vrijwel allemaal op dezelfde wijze. Als voorbeeld bekijken we een verzoek over de toewijzing van nieuwbouwoopwoningen afgehandeld door het ministerie van BZK.<sup>8</sup> Van de zes eerdergenoemde basis-metadata is er eigenlijk slechts één aanwezig, de titel. Vaak bestaat die uit twee à vier sleutelwoorden. In ons voorbeeld is dat *toewijzing nieuwbouwoopwoningen*. Er wordt ook wel een omschrijving van het

**Tabel 1. Overzicht per provincie van de digitale duurzaamheid van haar Woo-dossiers, gemeten op 23-10-2022.**

Provincie	url	titel	beschrijving	datum	datum verzoek	datum besluit
Groningen	<a href="https://www.provinciegroningen.nl/contact/woo-besluiten/">https://www.provinciegroningen.nl/contact/woo-besluiten/</a>	1				
Friesland	<a href="https://www.fryslan.frl/publieksinformatie">https://www.fryslan.frl/publieksinformatie</a>					
Drenthe	<a href="https://www.provincie.drenthe.nl/loket/wet-regelgeving/wet-open-overheid/woo-verzoeken/">https://www.provincie.drenthe.nl/loket/wet-regelgeving/wet-open-overheid/woo-verzoeken/</a>	1		1		
Overijssel	<a href="https://www.overijssel.nl/loket/bezwaar-klacht-info/gepubliceerde-beslissingen-wob-woo-verzoeken/">https://www.overijssel.nl/loket/bezwaar-klacht-info/gepubliceerde-beslissingen-wob-woo-verzoeken/</a>	1	1		1	
Flevoland	<a href="https://www.flevoland.nl/loket/wet-open-overheid-(woo)">https://www.flevoland.nl/loket/wet-open-overheid-(woo)</a>	1				1
Gelderland	<a href="https://www.gelderland.nl/woo-verzoeken">https://www.gelderland.nl/woo-verzoeken</a>	1		1		
Utrecht	<a href="https://www.provincie-utrecht.nl/loket/informatieverzoek-wob/wob-besluiten">https://www.provincie-utrecht.nl/loket/informatieverzoek-wob/wob-besluiten</a>	1		1		
Noord-Holland	<a href="https://www.noord-holland.nl/Loket/Woo_verzoek">https://www.noord-holland.nl/Loket/Woo_verzoek</a>	1		1		
Zuid-Holland	<a href="https://www.zuid-holland.nl/politiek-bestuur/bestuur-zh/gedeputeerde-staten/besluiten/#&amp;ajax=true&amp;facet_wob=10&amp;pager_page=0&amp;zoeken_term=&amp;date_from=&amp;date_to=">https://www.zuid-holland.nl/politiek-bestuur/bestuur-zh/gedeputeerde-staten/besluiten/#&amp;ajax=true&amp;facet_wob=10&amp;pager_page=0&amp;zoeken_term=&amp;date_from=&amp;date_to=</a>	1	1		1	1
Zeeland	<a href="https://www.zeeland.nl/loket/woo-verzoek">https://www.zeeland.nl/loket/woo-verzoek</a>	1		1		
Noord-Brabant	<a href="https://www.brabant.nl/loket/verzoek-overheidsinformatie/afgeronde-wooverzoeken">https://www.brabant.nl/loket/verzoek-overheidsinformatie/afgeronde-wooverzoeken</a>					
Limburg	<a href="https://www.limburg.nl/@748/informatieverzoek/">https://www.limburg.nl/@748/informatieverzoek/</a>	1		1		
TOTAAL		10	2	6	2	2
OpenOverheid	<a href="https://open.overheid.nl/zoekresultaten?informatiesoort=c_e0865e4d&amp;sort=date-desc&amp;page=1">https://open.overheid.nl/zoekresultaten?informatiesoort=c_e0865e4d&amp;sort=date-desc&amp;page=1</a>	1	1	1		
Hoeksche Waard	<a href="https://www.gemeentehw.nl/wob-verzoeken/">https://www.gemeentehw.nl/wob-verzoeken/</a>	1	1	1	1	1

dossier gegeven, maar die lijkt gebaseerd op een sjabloon, met daarin de woorden uit de titel ingevuld. Er worden twee datums gegeven, *publicatiedatum* en *documentdatum*, maar niet de gewenste datums van verzoek en besluit. Wat het verzoek was en wat er besloten is, wordt niet machineleesbaar in metadata aangegeven. Dit staat vanzelfsprekend wel in het besluit. Daarnaast wordt er een thema, in dit geval *woningmarkt*, meegegeven. De machineleesbaarheid van de gepubliceerde documenten (de pdf's) is voor dit voorbeeld uitstekend voor wat de tekst betreft.<sup>9</sup> Jammer genoeg worden alle vrijgegeven documenten wel aan elkaar geplakt.

## Conclusie

De provincies (noch de ministeries) houden zich bij het publiceren van hun Woo-dossiers aan de digitale duurzaamheidseis (artikel 2.4.3) in de Woo zelf. Daarmee wordt de herbruikbaarheid van die dossiers zeker niet bevorderd, zoals voorgeschreven in de EU-richtlijn 2019/1024. Deze situatie is eenvoudig te veranderen met een combinatie van *best practices* die al binnen de provincies worden toegepast. De Woo-dossiers gepubliceerd door de gemeente Hoeksche Waard zijn een ander prachtig voorbeeld van digitale duurzaamheid.

Woo-dossiers zijn slechts een van de zeventien informatie-categorieën die onder de Woo vallen. Woo-dossiers vallen samen met vergaderstukken, bereikbaarheidsgegevens en convenanten onder de stukken die als eerste actief en digitaal duurzaam openbaar gemaakt gaan worden. Ons onderzoek toont aan dat er met relatief kleine aanpassingen nog veel te verbeteren is aan de manier waarop Woo-dossiers worden

gepubliceerd. Dezelfde aanpassingen gelden onverkort ook voor alle andere Woo-categorieën. Dit kan een grote bijdrage leveren aan de bruikbaarheid en herbruikbaarheid van openbare overheidsstukken door burgers, en het bereiken van de doelstellingen van de Woo.

De resultaten van ons onderzoek zijn samengevat in tabel 1, ook beschikbaar als spreadsheet. Dit geeft de stand van zaken weer op 23 oktober 2022. |

## NOTEN

- 1 Zie: Kamerbrief Min BZK, 8 juli 2022, voortgang Woo.
- 2 Zie: IMI & OSF (2022) en IMI & OSF (2023).
- 3 Zie: EU Richtlijn 2019/1024 inzake open data en het hergebruik van overheidsinformatie.
- 4 Zie: Wilkinson (2016).
- 5 Deze figuur is per pagina, wat de meest eerlijke vergelijking is tussen provincies. Als we kijken naar documenten of hele dossiers zonder tekst is het beeld wat positiever (bijvoorbeeld voor Zeeland en Zuid-Holland is in 50% van de dossiers 'iets' leesbaar). Dit verhult echter bijvoorbeeld dossiers waarvan slechts het verzoek of besluit machineleesbaar is, maar alle geleverde stukken niet.
- 6 Zie: Wiedemann et al. (2021) en Van Heusden et al. (2021).
- 7 Zie: Elchik (2023).
- 8 BZK Woo-verzoek Toewijzing nieuwbouw-koopwoningen, en gereconstrueerde versie in WooGle.nl.mnre1034.2i.2022.62.
- 9 Dit is een positieve uitzondering, zie Marx & Kamps, 'Stukken vrijgegeven onder de Woo zijn vaak niet goed toegankelijk', *Od* (april 2023).

	verzoek-tekst	besluit-tekst	stukken beschikbaar	verzoek los	besluit los	inventarislijst los	stukken los	zip	fulltext score	vanaf	tot
			1		1			1		2022-02-08	
			1	1	1					2017-04-17	
	1	1								2019-01-03	
		1	1							2015-08-12	
			1		1			1		2022-05-25	
			1	1	1	1				2020-01-21	2022-12-07
			1		1					2016-01-20	
	1	1	1		1		1			2021-10-19	
			1							2019-10-19	
			1	1	1	1		1		2022-01-04	
	2	3	9	3	7	2	1	3			
			1		1					2010-01-07	
	1	1	1	1	1		1				