

---

# **Onderzoeksrapport naar de digitale toegankelijkheid van documenten die de overheid publiceert onder de Woo**

**Maarten Marx, IRLab, Universiteit van Amsterdam**

**januari 2024**  
Dit onderzoek is uitgevoerd in opdracht van het Adviescollege Openbaarheid en  
Informatiehuishouding (ACOI).



<b>1</b>	<b>Introductie</b>	<b>1</b>
<b>2</b>	<b>Machineleesbaarheid en de Fairiscore</b>	<b>3</b>
2.1	De Fairiscore . . . . .	3
2.2	De resultaten . . . . .	4
2.3	De oorzaak . . . . .	5
2.4	Hoe dan wel? . . . . .	6
<b>3</b>	<b>Doel en bereik van het onderzoek</b>	<b>7</b>
<b>4</b>	<b>Adviezen naar aanleiding van het onderzoek</b>	<b>9</b>
4.1	De manier van openbaarmaken . . . . .	9
4.2	De werkwijze van het openbaarmaken . . . . .	11
<b>5</b>	<b>Operationalisatie</b>	<b>15</b>
5.1	De onderzoeksvraag . . . . .	15
5.2	Hoe rapporteren we . . . . .	16
5.3	Voorbeeld: vindbaarheid en toegankelijkheid van beslisnota's . . . . .	16
<b>6</b>	<b>Wat betekent FAIRness voor elke informaticategorie</b>	<b>23</b>
6.1	Beantwoording van een Woo-verzoek: Woo dossiers (2i) . . . . .	23
6.2	Convenanten (2f) . . . . .	26
6.3	Beschikkingen (2k) . . . . .	28
6.4	Beleid en bestuurstukken (2c) . . . . .	29
6.5	Agendas en besluitenlijsten (2d) . . . . .	29
6.6	Organisatie en werkwijze en Bereikbaarheidsgegevens (1e) . . . . .	30
<b>7</b>	<b>Bevindingen FAIRness per informaticategorie</b>	<b>31</b>
7.1	Besluiten op Woo verzoeken . . . . .	31
7.2	Bereikbaarheidsgegevens en Woo-index . . . . .	42
7.3	Convenanten . . . . .	46
7.4	Beschikkingen . . . . .	51
7.5	Beleid en bestuurstukken . . . . .	55
7.6	Blijvende beschikbaarheid . . . . .	57
<b>8</b>	<b>Voorstellen tot andere werkwijzes</b>	<b>59</b>
8.1	Maak echt werk van die machine leesbaarheid . . . . .	59

8.2	Produceer toegankelijke documenten volgens de WCAG norm. . . . .	60
8.3	Hergebruik en publiceer slim . . . . .	61
<b>9</b>	<b>Verder lezen</b>	<b>63</b>
<b>10</b>	<b>Colofon</b>	<b>65</b>

---

## Introductie

---

We onderzoeken de uitvoering van de Wet open overheid (Woo) door alle bestuursorganen, van ministeries tot gemeentes en waterschappen. We richten ons daarbij op de allerlaatste stap, de publicatie van actief of passief openbaar gemaakte stukken. En dan beperken we ons nog tot alleen de *vorm* van die stukken. De inhoud en al helemaal de betekenis daarvan blijft buiten beschouwing.

Onze vraag is simpel: *wat is de kwaliteit van de digitale toegankelijkheid van die stukken?* Zijn ze na publicatie terug te vinden? Zijn ze toegankelijk, ook voor mensen met een beperking? Zijn ze her te gebruiken, ook voor andere doelen dan op voorhand beoogd? En is de informatie in die stukken makkelijk te verbinden met informatie uit andere bronnen?

Die kwaliteit kan vaak nog veel beter, maar soms gaat het ook wel goed.

Dit is zo ondanks het feit dat de overheid zich, via Europese richtlijnen en ook via de Woo zelf, aan al deze kwaliteitseisen gecommitteerd heeft. Uit gesprekken en onderzoek blijkt dat het in beginsel geen onwil is van ambtenaren om digitaal toegankelijk te publiceren, maar dat er wel onmacht is doordat de expertise hiervoor simpelweg ontbreekt, en mensen hulp nodig hebben bij het bepalen van de juiste werkwijze. Ook werken zij met de middelen die zij hebben, zoals de beschikbare laksoftware. Ons advies zou dus simpel “*overheid, hou je aan je eigen regels*” kunnen zijn, maar dat biedt geen handelingsperspectief.

Dat is er gelukkig wel. De kwaliteit van de openbaar gemaakte stukken is weliswaar vaak erg laag, dit is vaak ook onnodig het geval, en dus relatief eenvoudig te verhelpen. Het is echt mogelijk om een heleboel van de onvolkomendheden die we in dit rapport zullen beschrijven te voorkomen. En met behulp van AI kan er ook een heleboel gerepareerd worden. Wij zien niet alleen de technische mogelijkheden, en we beschrijven die ook in dit rapport, we zien ook een enorme bereid- en gedrevenheid bij Woo-ambtenaren om het goed te doen. De wil is er zeker.

Ons advies is dan ook: “*overheid, biedt ze de weg*”. Maak het makkelijk, ja gewoon vanzelfsprekend, om de stukken in de hoge, zelf voorgeschreven, kwaliteit te openbaren.

Werk daarbij op twee niveaus: Heel erg decentraal, direct op de ambtenaar met direct inzetbare, AI-gedreven “kijk-na-en-verbeter-samen” software. En juist heel centraal door een alles overkoepelend Woo platform aan te bieden waar iedereen alle stukken uit de 17 Woo-categorieën van alle bestuursorganen die onder de Woo vallen makkelijk en snel kan vinden en tot zich nemen. Dat is geen grootheidswaanzin. Met behulp van AI en een moderne, op vertrouwen, moderatie en *checks and balances* gebaseerde Wikipedia-achtige werkwijze is dat haalbaar. Wat een heerlijk wenkend perspectief.

---

---

### Machineleesbaarheid en de Fairiscore

---

Als opwarmer voor dit best lange rapport behandelen we een centraal begrip, *machine leesbaarheid*. Artikel 2.4.3 van de Woo stelt dat stukken vrijgegeven onder de Woo machine leesbaar moeten zijn en voorzien van metadata. Als dat zo is moet je stukken kunnen terugvinden en aan elkaar kunnen koppelen. Simpel gezegd test men bij machineleesbaarheid of een computer hetzelfde “ziet” of “leest” als een mens die een document (uitgeprint of van het scherm) leest. Dit heeft directe invloed op de vindbaarheid, de toegankelijkheid en de mogelijkheid tot hergebruik van die documenten<sup>1</sup>.

Machineleesbaarheid is een lastig begrip om precies te maken en dus om op te meten. Hoe kunnen we op miljoenen onder de Woo gepubliceerde pagina's nagaan of mens en machine de tekst die erop staat op dezelfde manier “zien”? Voordat we deze vraag beantwoorden is het goed om te kijken waarom dit überhaupt een probleem is. Eigenlijk is het vreemd dat een computer iets anders zou lezen dan een mens in bijvoorbeeld een Word document. Elke letter die je ziet staan is toch ingetypt? Dit is natuurlijk zo en voor “gewone” Word, PDF, chat of mail bestanden is dit ook helemaal geen issue. Het verschil tussen mens- en machine-leesbaarheid ontstaat wanneer een document wordt *ingescanned* (dit betekent dat er van elke bladzijde een foto gemaakt wordt) waarna vervolgens wordt getracht de woorden die erin stonden weer terug te halen via een automatisch, door AI gedreven, proces dat letters herkent in afbeeldingen. Dit proces heet *optische karakter herkenning (OCR)*. Deze scan&ocr techniek wordt voornamelijk ingezet bij documenten waarin iets weggelakt moet worden.

#### 2.1 De Fairiscore

Op basis hiervan hebben we een automatisch uit te voeren controle op machine leesbaarheid gemaakt: we kijken of een pagina geheel uit een afbeelding bestaat, hét teken dat het document later opnieuw is ingescand, en of er machineleesbare tekst in staat. Als er tekst inzit gaan we op zoek naar aanwijzingen dat er OCR is toegepast. Dat zie je aan gekke foutjes, bijvoorbeeld dat er “OekraTne” in een tekst staat (waar in het echt op de plek van die T natuurlijk een i stond). Zoeken met de term *Oekratne in Google* levert flink wat hits naar overheidsdocumenten op. We geven elke pagina een score van A tot E, vergelijkbaar met de Nutriscore voor voedsel. De slechtste, zogeheten *Fairiscore* E krijgt een document met geen enkele machineleesbare letter erin en waarbij elke pagina uit een afbeelding bestaat. De beste score A is voor documenten die niet gescanned zijn en waar dus tekst in te vinden is. De één na beste score B kan een ingescand document al krijgen als de OCR voor meer dan drie kwart net zo goed is als bereikt kan worden met de beste gratis beschikbare OCR. Score C en D zitten tussen de uitersten in; een document is bijvoorbeeld ingescand, en er is achteraf OCR toegepast, maar die

---

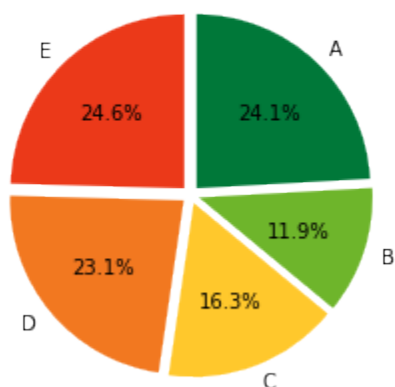
<sup>1</sup> Stukken waarvoor deze 4 zaken goed gaan noemen we *FAIR*. De 4 letters in FAIR staat (in het Engels) achtereenvolgens voor vindbaar, toegankelijk, koppelbaar en herbruikbaar.

bevat veel foutjes. Of het is een aan elkaar geplakt document, bestaande uit meerdere bijlagen, waarbij niet elke bijlage machine-leesbaar is. Volgens de wet zouden we mogen verwachten dat nieuwe stukken zonder of met minimale tekst redactie score A hebben en alle andere stukken A of hooguit score B.

## 2.2 De resultaten

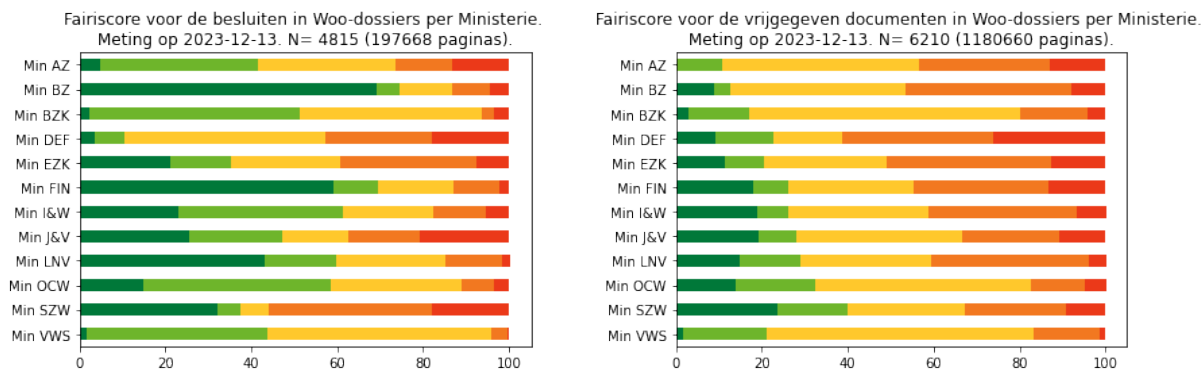
We hebben de Fairiscore voor drie soorten documenten berekend. Ten eerste alle documenten vrijgegeven na een Woo-verzoek beschikbaar op [Woogle](#). Dat zijn 51 duizend documenten met in totaal 1.4 miljoen pagina's gebundeld in 9.500 Woo-dossiers vrijgegeven door 44 bestuursorganen. De verdeling van die scores over deze documenten staat hieronder. Slechts 36% heeft score A of B.

Woogle Fairiscore verdeling voor Documenten in Woo dossiers, N=51109.



**Figuur: Verdeling van de Fairiscores voor alle documenten vrijgegeven na een Woo-verzoek die beschikbaar zijn op [Woogle](#).**

Om beter te kunnen vergelijken beperken we ons nu verder tot stukken gepubliceerd door ministeries op [overheid.nl](#). We maken onderscheid tussen de vrijgegeven stukken en de besluiten. Het besluit is een onlangs door een Woo-jurist opgestelde brief aan de verzoeker met minimale tekst-redactie. Een document waarvoor het dus onnodig is de scan&ocr techniek toe te passen. De resultaten staan in de twee tabellen hieronder. Zoals verwacht zijn de scores over de hele linie wat beter voor de besluiten dan voor de vrijgegeven stukken.

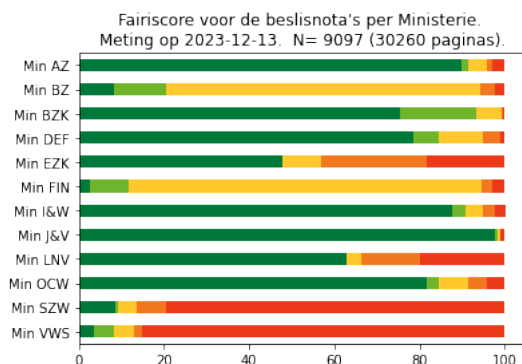


**Tabel: Verdeling van de Fairiscores voor alle besluiten en voor alle vrijgegeven documenten na een Woo-verzoek**



## ingediend bij een ministerie die beschikbaar zijn op **Woogle**.

Tenslotte hebben we dezelfde analyse uitgevoerd op een type documenten die niet onder de verplicht actief openbaar te maken documenten vallen: beslisnota's. Sinds 1 juli 2021 zijn beslisnota's over wetgeving en beleidsvorming bij Eerste en Tweede Kamerstukken openbaar en worden deze gepubliceerd op [open.overheid.nl](https://open.overheid.nl). Beslisnota's zijn allemaal heel recente stukken van gemiddeld 3 pagina's met minieme tekst-redactie. We zien hier een heel ander beeld, met vier ministeries die het erg slecht doen, en de anderen juist een stuk beter dan bij de besluiten. Maar eerlijk gezegd hadden natuurlijk alle balkjes donkergroen moeten zijn.



Tabel: Verdeling van de Fairiscores voor alle beslisnotas die beschikbaar zijn op **Woogle**, per ministerie.

## 2.3 De oorzaak

De oorzaak achter deze slechte scores is de scan&ocr methode die gebruikt wordt in veel laksoftware. Die software helpt de Woo jurist bij het verwijderen van persoonsgegevens en bespaart daarmee veel tijd. Jammer genoeg zien we dat die software heel vaak niet goed is afgesteld (alle documenten met score E bevatten helemaal geen tekst en daar stond de OCR dus uit), maar ook dat deze software wordt ingezet op korte net zelf gemaakte documenten met minimale redactie (rond de 1-3 weggelakte persoonsgegevens per document). We geloven dat gebruikers zich onvoldoende bewust zijn van de feilbaarheid van AI technieken zoals OCR. Het is anders toch niet denkbaar dat men een prachtige brief waar zoveel tijd en aandacht aan is besteed zo laat verminken<sup>2</sup>? Ook is het inscannen van PDF of Word documenten helemaal niet nodig om veilig te kunnen lakken<sup>3</sup>.

Er zijn gelukkig ook voorbeelden waar het wel goed gaat. Bij de gemeente [Hoeksche Waard](#) scoort 44% van de 711 documenten uit Woo-verzoeken een Fairiscore van A en nog eens 20% een B. Dit zijn stukken vanaf 2019. Van de 403 documenten uit Woo verzoeken uit 2021 en 2022 van de [Provincie Zuid Holland](#) scoort zelfs 44% een A en 43% een B.

<sup>2</sup> Op [deze blogpost op wooverheid.nl](#) staat een voorbeeld van zo'n besluit dat erg verminkt is door OCR. Iemand zonder zicht die dit moet laten voorlezen snapt er niets van.

<sup>3</sup> Beslissingen van Amerikaanse rechtbanken, vol met zwartgelakte persoonsgegevens van bijvoorbeeld getuigen, worden rechtstreeks in de PDF gelakt en blijven dus echte digitale documenten. De weggelakte informatie is veilig en kan niet meer naar boven gehaald worden. Ook onderzoek aan de Universiteit van Amsterdam naar lakken direct in PDFs laat zien dat dit eenvoudig én veilig kan, ook samen met het geven van laksuggesties.

---

## 2.4 Hoe dan wel?

Die vraag beantwoorden we in dit rapport. De conclusie van dit onderzoek naar de staat van de Woo documenten is tweeledig: op dit moment is er nog veel ruimte voor verbetering en wordt door de overheden geen recht gedaan aan de eigen richtlijnen. Maar oude stukken kunnen met AI heel goed gerepareerd worden, en kleine cultuurveranderingen kunnen al heel grote positieve effecten op de kwaliteit van de nieuwe stukken hebben. Beide maatregelen kunnen met weinig kosten, zeker gezien de enorme te behalen voordelen uitgevoerd worden. De uiteenlopende Fairiscores en een aantal lagere overheden laten zien dat dit geen luchtfietsrij en ook niet onhaalbaar duur is.

Goed machineleesbare documenten zijn een eerste vereiste voor een handig platform voor overheidsinformatie zoals PLOOI of [Woogle](#). De Woo documenten goed machineleesbaar te produceren is een haalbare en enorm nuttige eerste stap.

---

### Doel en bereik van het onderzoek

---

We onderzoeken in hoeverre documenten die onder de Woo openbaar zijn gemaakt én zijn gepubliceerd op internet voldoen aan artikel 2.4 lid 3 van de Woo. Hierbij nemen we de FAIR-data principes<sup>1</sup> als uitgangspunt: zijn de stukken vindbaar, toegankelijk, aan andere gegevens te koppelen, en herbruikbaar? Het bereik is zo breed mogelijk: in principe alle openbaar gemaakte en gepubliceerde documenten van alle bestuursorganen die onder de Woo vallen. Wel beperken we ons tot stukken die redelijkerwijs semi-automatisch te verzamelen en te analyseren zijn.

Naast *opmeten*, en daarbij aangeven wat er juist goed en wat er nog niet zo goed gaat, zullen we ook concrete technische aanbevelingen doen om zaken te verbeteren. We beperken ons tot technieken die schalen<sup>2</sup>, werken zonder menselijke tussenkomst, en eenvoudig en met weinig kosten in te voeren zijn.

De doelstelling van het onderzoek is daarmee tweeledig:

1. Verkrijgen van inzicht in de mate waarin bestuursorganen documenten digitaal toegankelijk openbaar maken op hun publicatiekanalen via internet.
2. Het doen van concrete technische aanbevelingen om de toegankelijkheid van overheidsinformatie snel en laagdrempelig te verbeteren.

Dit onderzoek betreft het opzetten en uitvoeren van een nulmeting. Het onderzoek is dan ook zo opgezet dat het in de toekomst makkelijk herhaald kan worden.

---

<sup>1</sup> Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

<sup>2</sup> Dit houdt in dat de kosten van die technieken slechts logaritmisch of lineair toenemen in de hoeveelheid verwerkte data.

---

---

## Adviezen naar aanleiding van het onderzoek

---

Dit hoofdstuk bevat de belangrijkste algemene adviezen. Deze zijn gebaseerd op het onderzoek dat in latere hoofdstukken beschreven staat. Ze vallen in 2 delen uitéén, over de manier van openbaarmaken, en over het proces voorafgaand aan het openbaarmaken. We vaten ze hier puntsgewijs samen.

1. Op welke wijze openbaar te maken?
  1. Als men openbaar maakt in PDF, dan een digitale en toegankelijke PDF. Vermijd de lakstraat zoveel mogelijk. Data zoals bijvoorbeeld spreadsheets nooit in PDF.
  2. Verbind openbaargemaakte informatie expliciet aan elkaar via hyperlinks. Maak er een web van informatie van in plaats van losstaande stukken.
  3. Plaats informatie liefst op één plek.
2. Voorafgaand aan het openbaarmaken
  1. Gebruik AI volop. Niet om achteraf dingen te repareren, maar liefst vooraf, bijvoorbeeld om de Woonambtenaar te helpen zich te richten op de inhoud.
  2. Maak het oplossen van gemaakte foutjes ook ná publicatie mogelijk en makkelijk.
  3. Sla de handen inéén. Deel oplossingen en leer van elkaar.

In hoofdstuk 8 werken we een aantal van deze adviezen heel concreet en direct inzetbaar uit.

## 4.1 De manier van openbaarmaken

### 4.1.1 Publiceren in PDF

Vrijwel alle stukken vrijgegeven onder de Woon worden in het PDF formaat aangeboden. PDF is een formaat ontwikkeld om stukken te printen. Sindsdien wordt het natuurlijk ook veel gebruikt om stukken te delen, en er daarbij zeker van te zijn dat het stuk er exact hetzelfde uitziet bij de ontvanger als bij de verzender. Er kleven echter wel nadelen aan het PDF formaat vanuit het oogpunt van FAIR data. Daar gaan deze deeladviezen over

---

## Vermijd PDF als het niet nodig of niet geëigend is. Dring het zeker niet op.

Voor (semi-)gestructureerde data zoals een *spreadsheet* of een *kaart* is het vreselijk zonde om die vrij te geven als PDF, omdat dan ontzettend veel waarde van de data verloren gaat. Toch wordt nog steeds vrijwel elke inventarislijst van een Woo-dossier, een *spreadsheet*, los als PDF of als bijlage in een PDF openbaar gemaakt<sup>1</sup>.

Vanuit het oogpunt van toegankelijkheid voor visueel of anderszins gehandicapten is het PDF formaat vaak inferieur aan bijvoorbeeld HTML, de *semantisch gestructureerde en opgemaakte taal van het World Wide Web*. Een beschikking of een convenant opgesteld in HTML moet dus zeker niet (verplicht) worden omgezet in PDF. De website <https://www.officielebekendmakingen.nl> doet dit ontzettend goed door te proberen al haar stukken zowel in HTML als in PDF formaat aan te bieden.

## Als men publiceert in PDF, hou de PDF dan zo digitaal mogelijk

Probeer alle waarde die in het digitale (vaak Word) document zit te behouden bij het overzetten of exporteren naar PDF. Dit is niet moeilijk, maar we zien heel vaak dat het niet goed gaat. Met veelal desastreuze gevolgen voor de vindbaarheid, de toegankelijkheid en de herbruikbaarheid van het document.

## Volg de instructies, vooral voor toegankelijkheid

Het is erg lastig om een PDF te maken die voldoet aan alle WCAG normen. Maar bij een eenvoudig document als een brief of een besluit is dit, zeker als het zelf gemaakt is in bijvoorbeeld MS Word, echt wel mogelijk door gewoon de regels te volgen. Twee voorbeelden van regels die al een groot effect hebben zijn:

1. Bekijk de metadata en vul degewenst aan. De taal van het document hoort ingesteld te zijn (veelal Nederlands); het document hoort een titel in de metadata te hebben; een (geanonimiseerde) auteur is gewenst, bijvoorbeeld “Woo-team Gemeente Wijdereen”.
2. Transformeer op de juiste wijze naar PDF. Dus vanuit Word niet Control P en *bewaar als PDF*, maar *exporteer als PDF* met de juiste toegankelijkheidskeuzes.

Het is enorm leerzaam om zelfgemaakte PDFs door de [pdfchecker.nl](https://www.pdfchecker.nl) van Forum Standaardisatie te halen, en te leren van de fouten. Heel veel daarvan zijn met de twee bovenstaande regels op te lossen. Vaste instellingen in Word (liefst gemaakt voor de hele organisatie) zorgen ervoor dat deze zaken *by default* goed gaan.

## Vermijd de lakstraat waar mogelijk

In ons onderzoek zijn we duizenden zelf geproduceerde documenten tegengekomen die “door de lakstraat gehaald zijn”, met vaak fatale en volstrekt onnodige gevolgen voor de FAIRness. De lakstraat gebruikt heel erg vaak AI technieken die niet onfeilbaar zijn, en is bovendien vaak gewoon slecht ingesteld.

Veel laksoftware gebruikt de *scan en OCR<sup>2</sup> techniek*, met de volgende drie nadelen:

1. slechte toegankelijkheid,
2. verlies aan informatie door OCR fouten of het niet aanzetten van de OCR, en
3. veel te grote bestanden.

Lak software heeft dus veelal een slecht effect op de digitale duurzaamheid van documenten. Vaak, als een document dat moet worden vrijgegeven alleen nog in PDF beschikbaar is, kan het niet anders en moet de laksoftware ingezet worden. Maar dat hoeft niet bij een document waarvan de *productie in eigen hand is*, zoals een beslisnota of een Woo-besluit. Die kunnen in veel gevallen met een slim ontwerp, en vaak direct vanuit MS Word, geanonimiseerd geschreven worden.

---

<sup>1</sup> Een gunstige uitzondering vormt de net nieuwe “wobcovid” website <https://open.minvws.nl>.

<sup>2</sup> OCR staat voor *Optische karakterherkenning*, een AI techniek om tekst te herkennen in afbeeldingen. Het *scannen* van een document is niets meer dan fotos nemen van elke bladzijde.

---

## 4.1.2 Informatie leeft in een netwerk, biedt het ook zo aan

De Woo draait om openheid (in het buitenland wordt het vaak *vrijheid* genoemd) van *informatie*, niet om openheid van documenten, en zeker niet alleen van PDF documenten. Al die overheidsinformatie staat in verbinding met elkaar, vrijwel elk stuk<sup>3</sup> verwijst naar andere stukjes informatie die ook een plekje hebben binnen het informatieaanbod van de Woo. Maar die verbintenissen zijn in de regel **niet expliciet via een hyperlink** gemaakt. Ons advies is om dat wel te doen<sup>4</sup>.

## 4.1.3 Maak informatie vindbaar en beschikbaar vanaf één plek.

Er is na het BIT advies over PLOOI<sup>5</sup> expliciet gekozen om informatie per bestuursorgaan decentraal aan te bieden. Dat is niet zo handig als het beschikbaar hebben van alles op één plek (d.w.z. één website waar alle stukken van alle bestuursorganen op uniforme wijze beschikbaar en tegelijk doorzoekbaar zijn<sup>6</sup>), maar de Woo-index lost dat dan weer op. Daar vindt men voor elk bestuursorgaan en elke Woo-informatie categorie waar de desbetreffende documenten dan te vinden zijn. Dat klinkt ideaal, maar de praktijk laat zien dat dit lastig volledig te krijgen is<sup>7</sup>. Ons advies is dan ook: Elk bestuursorgaan plaatst haar informatie liefst zoveel mogelijk op één plek; dat maakt het voor iedereen makkelijker het echt terug te vinden. En het is ook makkelijker de Woo-index volledig te krijgen (men hoeft dan slechts 1 link per informatie categorie in plaats van een aantal in te voeren).

## 4.2 De werkwijze van het openbaarmaken

### 4.2.1 Het gebruik van technische hulpmiddelen, in het bijzonder AI

#### Zet AI niet in om zelf gemaakte foutjes op te ruimen, maar vermijd ze liever als dat kan

AI kan heel veel, en steeds meer taken die tot voor heel kort alleen door mensen gedaan konden worden. Maar juist daardoor is toepassing van AI, in tegenstelling tot bijvoorbeeld een database programma als SQL, *nooit zonder fouten*. En ook al komen die niet veel voor, bij grote aantallen kan dat toch oplopen. Optische karakter herkenning (OCR) is een AI techniek die door laksoftware wordt ingezet om de “fout” van het inscannen weer teniet te doen. Heel goed als het niet anders kan, maar probeer het te vermijden waar dat kan.

Bewustzijn over de negatieve gevolgen van de lakstraat kan ertoe bijdragen dat die techniek alleen wordt ingezet waar het echt nodig is.

---

<sup>3</sup> Document zoals bedoeld in Artikel 2.1 Wet open overheid

<sup>4</sup> Neem als voorbeeld deze publicatie in de Staatscourant: Mededeling van toetreding van het College van de gemeente Altena tot het Convenant LSI 2017 ten behoeve van samenwerking in interventieteams, ex artikel 19, 2e lid, van het Convenant LSI 2017, Ministerie van Sociale Zaken en Werkgelegenheid. De pagina bevat wel hyperlinks, maar dat is allemaal *boilerplate*. Terwijl vrije informatie gemaakt vanuit een netwerk gedachte minimaal hyperlinks zou bevatten naar het convenant zelf en de “spelers”: de gemeente Altena, het Ministerie SZW, de Landelijke Stuurgroep Interventieteams (LSI).

<sup>5</sup> Zie <https://www.adviescollegeicttoetsing.nl/documenten/publicaties/2022/11/28/bit-advies-plooi>.

<sup>6</sup> Eigenlijk een Google voor alle Woo-documenten dus. Je kunt ze daar makkelijk opzoeken, en de zoekmachine *verwijst* je dan naar de plek waar ze op te halen zijn.

<sup>7</sup> De Woo index van het Ministerie van Defensie verwijst voor de Convenanten alléén naar [zoek.officielebekendmakingen.nl](http://zoek.officielebekendmakingen.nl). Echter op [open.overheid.nl](http://open.overheid.nl) vinden we ook convenanten van Defensie. En ook op [www.rijksoverheid.nl](http://www.rijksoverheid.nl) vinden we convenanten van ministeries.

---

## Zet AI juist wel in als hulp van de Woo-jurist

De Woo-jurist die zich buigt over Woo-verzoeken kan aan twee kanten enorm door AI gedreven software geholpen worden:

1. Bij het vinden van de voor een verzoek relevante stukken (met zogenaamde *e-discovery tools*). Dit is vaak een andere manier van zoeken dan we met Google doen. Vaak zoekt men naar een speld in de hooiberg, en is men niet tevreden met één antwoord, maar moet alles naar boven komen, en vermeldt in de inventarislijst. Dit vergt dus een andere manier van zoeken, waarbij AI enorm kan helpen<sup>8</sup>.
2. Bij het veilig publiceren van die stukken. Laat AI suggesties doen, per weigeringsgrond, voor stukjes die weggelakt moeten worden. Dit gebeurt *in samenspraak met de Woo-jurist*, die suggesties accordeert of afwijst. In dit proces leert de AI-software continu bij van de interactie met de jurist.
3. Bij het opstellen van een inventarislijst met rijke metadata per vrijgegeven document. Denk hierbij aan het type document (en dan niet `.docx` of `.ppt`, maar semantische categorieën als email, beleidsstuk, kamervraag, etc); de toegepaste weigeringsgronden en hoe vaak die zijn toegepast; datums; en de “key spelers” in het document (bijvoorbeeld de ondertekenaars van een convenant, in genormaliseerde vorm (bijv. hun Too of KvK codes)).

### 4.2.2 Maak het oplossen van gemaakte foutjes achteraf mogelijk en makkelijk

We zien in de data dat heel veel informatie verloren is gegaan tijdens het lakproces. We bedoelen hier informatie die bewust *niet* weggelakt is, en die dus openbaar gemaakt mocht worden. Het lijkt soms alsof men bewust de lakstraat veel te streng afstelt, en hiermee feitelijk ook niet weggelakte informatie in ieder geval digitaal verwijderd, allicht uit angst om foutjes te maken.

Als het makkelijk is om een stukje tekst te herstellen dat gemist is bij het lakken, dan wordt het werkproces eenvoudiger en minder angstig. Het kan hopelijk sneller, en met een veel betere (meer FAIR) kwaliteit van de documenten.

Dit vereist dus dat een vrijgegeven stuk makkelijk veranderd kan worden, en opnieuw *op dezelfde plek* geplaatst kan worden. Als organisaties als PLOOI kopieën maken, moeten die ook automatisch veranderd worden, liefst ook met aanpassing van de zoekindex.

Dit is technisch best lastig, maar niet onoverkomelijk.

### 4.2.3 Hergebruik meerwerk dat gedaan moet worden voor de Woo, en sla de handen ineen.

De gemeenten springen eruit wat betreft de bereikbaarheidsgegevens op [organisaties.overheid.nl](https://organisaties.overheid.nl). Informatie lijkt volledig, geactualiseerd, en vooral ook nuttig. We vinden vrijwel alle namen, functies, rollen en contactgegevens van de raadsleden, bestuurders, en belangrijke ambtenaren. Het verschil met de waterschappen, provincies en gemeentelijke regelingen valt erg op. Allicht komt dit omdat de gemeentes al die gegevens *centraal via hun koepel de VNG opslaan*, en voor verschillende nuttige doelen inzetten. Hier is de Woo-index er maar één van. Ook voor deze index zijn de gemeentes zelf verantwoordelijk, *net als voor de bereikbaarheidsgegevens*. Dit voorbeeld toont ons twee lessen:

1. Sla de handen ineen, en zorg voor een handige plek waar bestuursorganen makkelijk en snel gevraagde gegevens kunnen invullen, aanvullen en veranderen.
2. Zorg dat de ingevulde gegevens makkelijk *herbruikbaar* zijn, en eigenlijk al gebruikt worden voor een aantal, voor de bestuursorganen zelf ook, nuttige toepassingen. Wees hierbij creatief: bijvoorbeeld: elk bestuursorgaan met een raad of een commissie wil aan het eind van het jaar een overzicht opstellen van het verloop in die raad: wie kwam erbij en wie ging weg. Die data is aanwezig in de centrale databank, want keurig bijgehouden. Maak een simpele applicatie waarmee een bestuursorgaan de mutaties in een jaar kan inzien. Dit biedt een sterke intrinsieke motivatie

---

<sup>8</sup> AI kan bijvoorbeeld de “pakkans” voor relevante stukken verhogen door het gebruik van synoniemen van zoektermen. Zo vindt men stukken waarin de zoekterm niet voorkomt maar die wel gaan over het begrip of concept achter de zoekterm.



---

om die mutaties nog sneller en beter door te geven. Een voorbeeld is het [overzicht van de afgehandelde Woo-verzoeken van de gemeente Nijmegen](#). Allemaal automatisch gemaakt vanuit de toch al ingevoerde metadata (die allemaal, behalve de tooi-categorie, gewoon in het besluit staan).

Een ander voorbeeld van samenwerking, hergebruik en de handen inéén slaan is het [OpenWebConcept](#) initiatief opgezet door een stel enthousiaste bestuursorganen die besloten het heft zelf in handen te nemen. Deze gemeentes, zoals Buren, Hoeksche Waard, en de BAR gemeentes, lopen veelal mijlenver voor als het gaat om publiceren van stukken onder de Woo. Dit komt omdat ze werken vanuit een meent (*commons* of *common ground*) gedachte waarin de één iets laat ontwikkelen en betaald en dat meteen aan alle anderen gratis beschikbaar stelt. Die verbeteren het weer en stellen dat weer beschikbaar, of veranderen het voor een volgende Woo informatiecategorie, enzovoort. Innovatie gaat zo heel snel, tegen lage kosten, en met een grote mate van voldoening en werkvreugde.

---

In dit hoofdstuk beperken we het onderzoek. We benoemen de soorten documenten die we gaan onderzoeken en van welke bestuursorganen. We benoemen kort wát we dan onderzoeken, en hoe we meten en rapporteren. Daarna maken we dit duidelijk met behulp van een uitgebreid voorbeeld over de vindbaarheid en toegankelijkheid van de beslisnota's van de ministeries.

**Welke bestuursorganen.** We bekijken stukken van de ministeries, de provincies, de waterschappen, de gemeentes en een flink aantal andere (zelfstandige) bestuursorganen.

**Welke informatiecategorieën.** We bekijken stukken van de vier informatiecategorieën die als eerste volgens de Woo gepubliceerd moeten worden: de bereikbaarheidsgegevens, convenanten, besluiten op Woo-verzoeken, en de vergaderstukken (deze laatste alleen van de decentrale overheden). Daarnaast bekijken we de Woo-index, beschikkingen en de agendas en besluitenlijsten (ook deze laatste weer alleen van de decentrale overheden).

Ons voorbeeld over de beslisnota's gaat over een type documenten die niet onder de Woo vallen, maar die dikwijls op een heel vergelijkbare manier als Woo-verzoeken worden gepubliceerd.

## 5.1 De onderzoeksvraag

De onderzoeksvraag van dit rapport is

Voldoen de onder de Woo openbaar gemaakte stukken aan artikel 2.4 lid 3 van de Woo<sup>1</sup>?

Dit artikel stelt dat Woo-bestanden “in elektronische vorm, in een machinaal leesbaar open formaat, samen met de meta-data” gepubliceerd worden. Dit wordt verder gespecificeerd met verwijzing naar een aantal (Europese) richtlijnen<sup>2</sup>.

Een andere gerelateerde richtlijn voor onze vraagstelling is het *Tijdelijk besluit digitale toegankelijkheid overheid*<sup>3</sup>, waarin staat dat overheidssites moeten voldoen aan toegankelijkheidseisen zoals vastgelegd in EN 301 549/WCAG 2.1. In dit onderzoek richten we ons *niet* op websites, maar alleen op gepubliceerde stukken. Die zijn voor het overgrote deel in het PDF formaat. Die vallen onder wat *kantoorbestanden* genoemd worden<sup>4</sup>.

---

<sup>1</sup> <https://wetten.overheid.nl/jci1.3:c:BWBR0045754&hoofdstuk=2&artikel=2.4&z=2023-04-01&g=2023-04-01>

<sup>2</sup> Waaronder de hergebruik richtlijn uit 2003 die ondertussen ook een Nederlandse invulling heeft gekregen in de Wet Hergebruik Overheidsinformatie <https://wetten.overheid.nl/BWBR0036795/2021-07-01>.

<sup>3</sup> <https://wetten.overheid.nl/BWBR0040936/2018-07-01>

<sup>4</sup> Deze website helpt bij het duidelijk maken waar PDFs aan moeten voldoen: <https://www.digitoeankelijk.nl/aan-de-slag/tips/kantoorbestanden>.

---

We zullen kort aandacht besteden aan WCAG, maar omdat Woo PDFs daar in overgrote mate heel slecht aan voldoen richten we ons eerst op de toegankelijkheidseisen die makkelijker te bereiken zijn (zoals de machineleesbaarheid die in de Woo zelf genoemd wordt).

### 5.1.1 Wat meten we op

Zoals gezegd hebben we gekozen om de stukken te beoordelen op de FAIR-data principes: zijn de stukken vindbaar, toegankelijk, aan andere gegevens te koppelen, en herbruikbaar? Machine leesbaarheid en metadata zijn voor al deze principes cruciaal. De principes zijn nog steeds best abstract, maar kunnen als volgt heel concreet gemaakt worden:

1. **Vindbaarheid** Ga dit na voor een PDF document door een woord dat je ziet staan in de tekst op te zoeken met Control F of iets te selecteren en te kopiëren met Control C.
2. **Toegankelijkheid** Laat een tekst voorlezen; als je niks hoort of vreemd gebrabbel is er iets mis met de machine leesbaarheid. Vergroot de tekst extreem.; als de lijnen in letters een zaagtand vorm krijgen of bovenop een raster van puntjes liggen is er sprake van een afbeelding van tekst.
3. **Koppelbaarheid** Bevat de PDF metadata in de vorm van *tags*<sup>5</sup> (verplicht in WCAG)? Staat de metadata van de stukken in een ISO-standaard formaat, en worden standaard ontologieën<sup>6</sup> gebruikt? Is basale metadata überhaupt aanwezig?
4. **Hergebruik** Kan de data (de tekst en de metadata) zonder enorme inspanning worden hergebruikt *voor toepassingen waarvoor de stukken in eerste instantie totaal niet bedoeld waren?*<sup>7</sup>

## 5.2 Hoe rapporteren we

We rapporteren veelal *comparatief*, waarmee we de resultaten voor één bestuursorgaan plaatsen in de context van de andere organen van hetzelfde organisatietype. We zullen vaak nogal negatieve resultaten laten zien, omdat de werkelijkheid nou eenmaal zo is. Dan proberen we altijd een voorbeeld daarnaast te zetten dat laat zien hoe het ook wel goed kan.

## 5.3 Voorbeeld: vindbaarheid en toegankelijkheid van beslisnota's

We laten zien hoe we toetsen op toegankelijkheid en vindbaarheid, en illustreren dat met het voorbeeld van de beslisnota's vrijgegeven op [open.overheid.nl](https://open.overheid.nl).

### 5.3.1 Toegankelijkheid en vindbaarheid

Onze analyse gaat uit van het volgende antwoord op de vraag “Waarom plaatst men iets op het internet?”.

Vanuit de wens dat een stuk gelezen en, mits relevant voor een zoekvraag, gevonden zal worden.

We kijken dus of door de overheid gepubliceerde stukken inderdaad door iedereen makkelijk gelezen kunnen worden, en of ze (terug)gevonden kunnen worden. We gebruiken hiervoor de volgende methode:

1. We selecteren een organisatietype, zoals de ministeries, of de provincies.
2. We selecteren een bepaald soort documenten, bijvoorbeeld *beschikkingen*.

---

<sup>5</sup> Hiermee wordt semantische informatie over de opmaak van het document aangegeven (bijvoorbeeld dat een stukje vetgedrukte tekst in een grote lettertype de kop van een nieuwe sectie is).

<sup>6</sup> Dit zijn internationaal afgesproken standaard manieren om over bepaalde zaken te spreken. Bijvoorbeeld de [Dublin Core standaard](#).

<sup>7</sup> Dit vinden mensen vaak erg vergezocht, maar dat valt best mee. Zo is Google Translate ontwikkeld op basis van de simultaan vertaalde Handelingen van het Europees Parlement. En wie had 20 jaar geleden aan Siri of ChatGPT gedacht bij het schrijven van een lemma voor Wikipedia?

- 
3. Voor dat organisatietype en soort documenten halen we zoveel mogelijk al die (meestal PDF) documenten op, met zoveel mogelijk metadata.
  4. Dan openen we elk document en proberen vast te stellen of het document via de “scan-en-ocr” methode gemaakt is. Met de volgende vragen
    1. Staat er machine-leesbare tekst in?
    2. Bestaan de paginas uit afbeeldingen?
    3. Vinden we verhaspelde woorden in het document?
    4. Zien we tekst-redactie (weggelakte stukjes)?
  5. We rapporteren dan per bestuursorgaan, en geven positieve en negatieve voorbeelden.
  6. We geven aan hoe bestuursorganen dit soort documenten FAIR zouden kunnen publiceren.

### 5.3.2 Beslisnota's

In een beslisnota leggen ambtenaren conceptbeslissingen voor aan de bewindspersoon, met alle afwegingen om tot een beleidskeuze te komen. Sinds 1 juli 2021 zijn beslisnota's over wetgeving en beleidsvorming bij Eerste en Tweede Kamerstukken openbaar en worden deze gepubliceerd [[Wikipedia](#)].

#### Eigenschappen van het corpus

Eind September 2023 hebben we alle beslisnota's, gepubliceerd door ministeries, die op [open.overheid.nl](#) stonden opgehaald. Inclusief bijbehorende documenten waren deze 7.919 PDF files, als volgt verdeeld over de ministeries:

Departement	Aantal beslisnota's
Ministerie Van Algemene Zaken	61
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	920
Ministerie Van Buitenlandse Zaken	786
Ministerie Van Defensie	197
Ministerie Van Economische Zaken En Klimaat	724
Ministerie Van Financiën	800
Ministerie Van Infrastructuur En Waterstaat	898
Ministerie Van Justitie En Veiligheid	741
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	479
Ministerie Van Onderwijs, Cultuur En Wetenschap	665
Ministerie Van Sociale Zaken En Werkgelegenheid	503
Ministerie Van Volksgezondheid, Welzijn En Sport	1300

**Tabel: aantal documenten geklassificeerd als beslisnota per ministerie. Opgehaald van [open.overheid.nl](#) eind September 2023.**

Beslisnota's zijn over het algemeen erg kort: 80% bestaat uit 1 tot 3 paginas. De overige 20% hebben er iets meer.

---

## Informatie die niet gedeeld mag worden

Voor zover we steekproefsgewijs hebben gezien worden er alleen persoonsgegevens weggelakt. Dit betreft eigenlijk vrijwel altijd gegevens over de opsteller(s) van de nota. We zien stukjes lak dan ook nooit *in de tekst zelf*, maar wel in de aanhef, in het briefhoofd, en soms ook in de salutatatie.

Onze “zwartlakdetector” laat zien dat gemiddeld rond de 1-2% van de tekst is weggelakt.

### 5.3.3 Vraag 1: machineleesbare tekst.

We testen of een pagina machineleesbare tekst heeft. Wij testen dat automatisch, maar de test komt overeen met de volgende makkelijk zelf uit te voeren test: selecteer alle tekst op een pagina (Control A), kopieer die (Control C), en plak die in een nieuw document (Control V). Als er **niets verschijnt** in het nieuwe document, bevat de oorspronkelijke pagina geen machine leesbare tekst.

De tekst is dan niet voor te lezen door een computer, er kan niet (in ieder geval niet zonder iets extra's te doen) in gezocht worden met Control F, en een zoekmachine als Google kan de inhoud van die pagina niet (in ieder geval niet zonder iets extra's te doen) indexereren. Zoeken (zowel via Google, als door middel van Control F in het document) op een woord dat op die pagina staat, zal dus in veel gevallen geen treffer opleveren.

We vinden dat 24.7% (N=26.980) van de paginas, en 23.2% van alle beslisnota's (N=7.919) geen enkel machine leesbaar karakter bevat. Voor de afzonderlijke ministeries zijn die percentages voor de beslisnota's als hieronder. De meeste doen het dus heel goed.

Departement	% bestanden zonder tekst
Ministerie Van Algemene Zaken	3.3
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	1.0
Ministerie Van Buitenlandse Zaken	2.9
Ministerie Van Defensie	0.5
Ministerie Van Economische Zaken En Klimaat	19.5
Ministerie Van Financiën	2.9
Ministerie Van Infrastructuur En Waterstaat	3.8
Ministerie Van Justitie En Veiligheid	5.1
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	19.6
Ministerie Van Onderwijs, Cultuur En Wetenschap	6.5
Ministerie Van Sociale Zaken En Werkgelegenheid	78.5
Ministerie Van Volksgezondheid, Welzijn En Sport	83.9

**Tabel: aantal documenten geklassificeerd als beslisnota zonder ook maar 1 karakter machine leesbare tekst, per ministerie als percentage van het totaal aantal beslisnota's per ministerie.**

### 5.3.4 Vraag 2: een echt document of een afbeelding?

Alle beslisnota's zijn PDF documenten. Die zijn op veel manieren te maken. De meest eenvoudige is via *save as PDF* (Control P) vanuit een tekst-verwerker als Word<sup>8</sup>. Daar komt een prima machine- en mens-leesbaar document uit. Dit is goed zelf te checken door de hierboven beschreven test voor machine-leesbaarheid. Mens-leesbaarheid is na te gaan door het document extreem te vergroten in een PDF-viewer. De letters in het document blijven **perfect leesbaar** bij extreme vergroting.

Dit is de kwaliteit die we verwachten van een recent gemaakt en gepubliceerd document. Toch is dat bij veel beslisnota's niet het geval. We zien hier dat paginas bestaan uit **afbeeldingen van letters**. Dit is makkelijk zelf na te gaan door de

---

<sup>8</sup> Dit is zeker niet de geëigende manier, en ook zeker niet de manier om WCAG *toegankelijke* PDFs uit Word te produceren.

extreme-vergroting-test (vergroot een PDF tot 400% of meer, iets wat visueel gehandicapten dikwijls moeten doen om iets te kunnen lezen). Letters “groeien” niet mee, maar worden korrelig en (zeer) slecht leesbaar.

Deze test hebben we geautomatiseerd door per pagina alle afbeeldingen uit de PDF te halen en hun gezamenlijke oppervlakte te berekenen. Komt die overeen met de standaard oppervlakte van een A4, dan bestaat dus de hele pagina uit één of meerdere afbeeldingen.

In 44.3% van de beslisnota's bestaan **alle paginas** uit afbeeldingen. In 53% is dat voor minstens 1 pagina het geval. Het gaat dus ook heel vaak goed: in 47% van de beslisnota's is de PDF waarschijnlijk prima mens- en machine-leesbaar.

Hieronder het percentage beslisnota's waarin alle paginas afbeeldingen zijn, per ministerie. Vergeleken met het vorige overzicht, zien we bij veel meer ministeries hoge percentages. Dit is de scan en OCR techniek.

Department	% beslisnota's waarbij elke bladzijde volledig uit een afbeelding bestaat
Ministerie Van Algemene Zaken	9.8
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	19.1
Ministerie Van Buitenlandse Zaken	89
Ministerie Van Defensie	17.8
Ministerie Van Economische Zaken En Klimaat	19.1
Ministerie Van Financiën	82.2
Ministerie Van Infrastructuur En Waterstaat	2.3
Ministerie Van Justitie En Veiligheid	2.2
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	19.3
Ministerie Van Onderwijs, Cultuur En Wetenschap	6.5
Ministerie Van Sociale Zaken En Werkgelegenheid	79.7
Ministerie Van Volksgezondheid, Welzijn En Sport	94.1

**Tabel: Het percentage beslisnota's waarin alle paginas afbeeldingen zijn, per ministerie.**

De software waarmee een PDF gemaakt is is terug te vinden via Control I in de metadata. Hier zien we de 10 meest gebruikte software, en hoeveel van de beslisnota's daarmee gemaakt zijn:

pdfCreator	aantal beslisnota's
Aspose	6653
Adobe	2079
FineReader	1866
Microsoft®	264
ABBYY	187
Acrobat	114
OmniPage	75
ZyLAB	54
PScript5.dll	43
ScandAll	9

**Tabel: Top 10 meest gebruikte software om een beslisnota PDF mee te maken, met het aantal documenten dat daarmee gemaakt is.**

---

### 5.3.5 De Scan en OCR techniek

Veel door de overheid gebruikte *laksoftware* gebruikt de “scan en ocr” techniek om een document te publiceren. Hierbij wordt er van elke bladzijde een afbeelding gemaakt (de *scan*), waarna men vervolgens probeert de tekst die op die pagina staat weer tevoorschijn te halen door *optische-karakter-herkenning* (OCR) op die afbeelding toe te passen.

Voor stukken die alleen nog op papier voorhanden zijn is dit een geweldige uitkomst. Voor recente, digitaal vervaardigde en beschikbare stukken heeft dit drie grote nadelen:

1. de toegankelijkheid voor slecht(er) zienden neemt enorm af
  - de PDF verliest vaak haar metadata
  - de letters worden *gerasterd* en groeien dus niet goed mee bij vergroting
  - alle semantische codes (verplicht in PDFs volgens de WCAG2 richtlijnen) verdwijnen. Hierdoor wordt voorlezen en bladeren veel moeilijker. (Semantische codes geven bijvoorbeeld aan of een stukje tekst een kop of een paragraaf is of het met nadruk moet worden uitgesproken, etc).
2. de vindbaarheid neemt af, omdat het OCR proces nooit foutloos is
3. de herbruikbaarheid neemt af, door de fouten in de OCR, maar ook doordat bestanden enorm veel groter worden (tot wel 100 keer zo groot als nodig om de tekst op te slaan).

#### Verhaspelde woorden

Het is niet zo eenvoudig om automatisch te tellen hoeveel OCR fouten er gemaakt worden. We noemen een woord dat veranderd wordt door de OCR *verhaspeld*. Wij checken op verhaspelde woorden door alle woorden uit een beslisnota te halen, en dan te kijken of dat bestaande Nederlandse woorden zijn. Dit laatste testen we door te kijken of een woord voorkomt in een lijst van 400.000 Nederlandse woorden van de Taal Unie of dat het woord in de Handelingen der Staten Generaal tussen 2005 en 2022 voorkomt. Samen gaat het dan om meer dan 650.000 unieke woorden.

Natuurlijk komen hier woorden *niet* in voor. Grappig genoeg is *beslisnota* zelf zo'n woord.

We geven nu een ruw beeld van de tellingen van de verhaspelde woorden. We vonden iets meer dan 100.000 unieke woorden die niet in onze woordenlijsten zaten. Hiervan kwamen er 75 duizend maar 1 keer voor. In totaal kwamen die niet bestaande woorden zo'n 380 duizend keer voor.

Opstellers van beslisnota's zijn vaak heel creatief met taal, want veel van de door ons gevonden woorden waren complexe samenstellingen van bestaande woorden, zoals *Arbeidsextensiviteit*, *Alleenstaandeouderen*, en *Autowrakkenrichtlijn*.

De volgende voorbeelden zijn wel echte OCR-fouten:

- *privmuaeicdtybewedsaitanergdbiionnrgggen of vteit^alliehatareQldipioilatiekebelangen*
- Een veel voorkomende verhaspeling is het vervangen van een l door een i, zoals in *Toesiagen* (241 keer) en *Bijjage* (261 keer) en *Bijjagen* (633)
- De OCR verliest soms de spaties zoals in *Randvoorwaardelijkewerkzaamhedenvoordebeveiligingenveiligheid-  
vanbestaandeoffshoreoperatiessleepbotenwerkbottenvoorsleepensalvagedienstenoilspillrecovery-surveillanceevenalsSearchandRescue*

Over het algemeen valt het best mee, en is de OCR van zeer hoge kwaliteit.



---

### 5.3.6 File size van de PDF

We zagen boven dat 47% van de beslisnota's geen enkele pagina heeft die helemaal uit een plaatje bestaat, en dat bij 44% juist elke pagina uit een plaatje bestaat. Een pagina van zo'n PDF zonder die plaatjes is gemiddeld 67 KiloBytes groot. Die gescande beslisnota's met een plaatje van elke pagina zijn daarentegen gemiddeld 470 KiloBytes per pagina groot. Dat is meer dan 7 keer zo groot.

Ter vergelijking, dit *hele hoofdstuk* als PDF is 76 KiloBytes groot, en het originele bestand waar die PDF uit gemaakt is maar 14KB. En als we dat met zip *compressen* nog geen 5KB.

Voor de informatieinhoud hoeven we alleen de *compressed* originele file op te slaan. Als we dat dan mooi als PDF willen laten zien kan dat *on-demand*. Dat zou betekenen dat we een pagina bijna 100 keer zo klein kunnen opslaan als nu gebeurt met de scan-en-ocr techniek.

### 5.3.7 Voorbeelden hoe het niet moet, en hoe het ook kan

#### Zo moet het niet: Onnodig scan en OCR

In deze beslisnota gaat het over Oekraïne, maar dat woord is in de machine leesbare tekst verhaspeld tot OekraTne. Ook wordt de tekst lastig leesbaar bij extreme vergroting.

De nota zelf zegt expliciet dat alle informatie in de nota openbaar gemaakt kan worden: *Informatie die niet openbaar gemaakt kan worden N v t*. Natuurlijk zijn er persoonlijke gegevens weggelakt, maar dit had ook “by design” gekund. Het is dus onnodig en jammer dat er voor dit document gekozen is voor de “scan en OCR” optie.

#### Zo kan het ook: Digitaal geproduceerde PDF

Deze beslisnota is geproduceerd door MS Word, is zeer goed te vergroten, alle tekst staat er netjes in, en ook hier zijn, we citeren, *gegevens van de steller en de ondertekenaar [...] gelakt vanwege bescherming van de persoonlijke levenssfeer*.

### 5.3.8 Restrictie tot 2023

In 2023 zijn 4.430 beslisnota's gepubliceerd. Als we de analyse beperken tot stukken uit 2023 zien we eigenlijk vrijwel hetzelfde beeld, met één groot verschil. Het aantal PDFs waarbij elk pagina een afbeelding is bij het Ministerie van Defensie enorm gedaald naar nog maar 1.5%. Heel goed nieuws dus.

### 5.3.9 Advies

Beslisnota's zijn recent digitaal vervaardigde stukken, en horen dus ook gewoon als een net digitaal PDF document openbaar gemaakt te worden. Ook al moeten er delen worden weggelakt, juist voor beslisnota's is dat vaak *boilerplate*, en kan dat heel makkelijk op het moment dat de nota gemaakt wordt gebeuren. De nota's van bijvoorbeeld het ministerie van Defensie laten mooi zien hoe dat kan. Zelfs de meestal erg strenge <https://pdfchecker.nl/> heeft maar een paar aanmerkingen op de WCAG 2.1 toegankelijkheid van het *gegeven voorbeeld*. Ook die zijn eenvoudig te verbeteren.

---

---

## Wat betekent FAIRness voor elke informatiecategorie

---

In dit hoofdstuk maken we concreet wat we bedoelen met een *FAIR document*. Dus een document dat vindbaar, toegankelijk, aan andere gegevens te koppelen, en herbruikbaar is. Dit doen we voor elke informatiecategorie die in dit rapport besproken wordt. De operationalisatie bestaat uit een check- en een wensen-lijst en positieve en negatieve voorbeelden.

We behandelen achtereenvolgens Woo-dossiers (2i), Convenanten (2f), Beschikkingen (2k), Beleid en bestuurstukken (2c), Agendas en besluitenlijsten (2d), en de Bereikbaarheidsgegevens (1e).

### 6.1 Beantwoording van een Woo-verzoek: Woo dossiers (2i)

De besluiten na een Woo-verzoek vormen een kernonderdeel van de Woo en zijn exemplarisch voor veel van de andere informatiecategorieën. Veel van wat we hier zullen schrijven gaat dus ook op voor stukken van andere informatiecategorieën, en zullen we daar dan ook niet meer herhalen. Ze zijn ook speciaal omdat het om complexe bundels van informatie gaat, die we *Woo dossiers* noemen.

Hoe ziet zo'n Woo-dossier er conceptueel uit, en hoe vertaalt dat zich in een FAIR technische oplossing?

Dat is beschreven in Larooij et al. (2023)<sup>1</sup> en dat volgen we hier. Een Woo dossier bestaat uit vier onderdelen:

1. een verzoek
2. een besluit
3. een lijst met alle voor het verzoek relevante stukken; de *inventarislijst*
4. de concrete stukken uit die lijst die geheel of gedeeltelijk openbaar gemaakt zijn.

Deze opdeling geeft meteen ook de minimaal gewenste metadata voor een Woo dossier:

1. de datums van verzoek en van besluit<sup>2</sup>
2. de verwoording van het verzoek en het besluit

---

<sup>1</sup> Larooij, M. et al. (2023), ESB, 108(4817), 36-39, [https://esb.nu/036-039\\_larooijkamp-2](https://esb.nu/036-039_larooijkamp-2)

<sup>2</sup> Deze vormden de kern van de bevindingen in de *Ondragelijk Traag* en *Matglas* rapporten van IMI en OpenState. Voor die rapporten zijn die datums allemaal handmatig uit de besluiten gehaald. Een enorm tijdrovend karwei.

- het besluit ook als keuze uit een lijstje gestandaardiseerde termen (als (gedeeltelijk) openbaar gemaakt, afgewezen, niet ontvankelijk, etc)

3. de inventarislijst heeft gestandaardiseerde kolommen met gestandaardiseerde waardes, met in ieder geval

1. naam document
2. mate van openbaarmaking
3. toegepaste weigeringsgronden
4. soort document (semantisch, niet technisch; dus *email*, *spreadsheet* of *rapport* en niet *PDF* of *.xlsx* of *Word document*)
5. mits beschikbaar, datum van het document
6. mits beschikbaar, opsteller/auteur/verantwoordelijke van het document.

4. De vrijgegeven stukken tenslotte worden in een bundel *los van elkaar* aangeboden.

- Dat kan met kleine aantallen als losse bestanden. Voor grote aantallen is bundelen in een zip bestand de meest eenvoudige en ideale oplossing<sup>3</sup>.

De inventarislijst wordt als *spreadsheet* in open formaat aangeboden (*csv*) en er bestaat een machine-verwerkbare link tussen de rijen in het spreadsheet en de vrijgegeven files. Hierdoor kan de *metadata* uit de inventarislijst automatisch aan de vrijgegeven stukken gekoppeld worden.

Hoe dit technisch kan worden vormgegeven is te zien in dit [voorbeeld dossier uit Barendrecht](#). Klikkt men op [Download volledig dossier \(ZIP\)](#), dan krijgt men inderdaad een zip bestand met de informatie zoals in de afbeelding hieronder. In dit voorbeeld zit alle [metadata van dit dossier](#) in een zogenaamd Dublin Core JSON bestand, een machine-leesbare taal die gebruik maakt van gestandaardiseerde namen voor de attributen.

Item	Size	Type
nl.gm0489.2i.2023.3	--	Folder
BESLUIT_Woo-besluit-geanonimiseerd.pdf	700 KB	PDF Document
VERZOEK_Woo-verzoek-geanonimiseerd.pdf.pdf	106 KB	PDF Document
Metadata	--	Folder
nl.gm0489.2i.2023.3-inventaris.csv	1 KB	comma...d valu
nl.gm0489.2i.2023.3.json	3 KB	JSON File
Bijlagen	--	Folder
04Exploitatievergunning-Diggels-geanonimiseerd.pdf	916 KB	PDF Document
04Exploitatievergunning-PaPi-geanonimiseerd.pdf	660 KB	PDF Document
04Exploitatievergunning-Uitspanning-geanonimiseerd.pdf	732 KB	PDF Document
nl.gm0489.2i.2023.3.zip	2.7 MB	ZIP archive

**Figuur: Inhoud van het zip bestand van een Woo dossier uit Barendrecht.**

### 6.1.1 FAIRness van de vrijgegeven stukken

FAIRness is het beste te zien als een concrete invulling van de volgende drie beleidsprincipes. Die principes kunnen gezien worden als de informatiehuishoudingskundige vertaling van het bekende [Bijbelse gebod](#), *Wat jij niet wilt...*

Openbaar uw stukken zoals jij ze graag had aangetroffen in uw zoektocht naar alle relevante stukken voor een Woo-verzoek.

<sup>3</sup> Deze oplossing wordt al door meerdere bestuursorganen toegepast.

---

## De 3 principes

1. Maak openbaar in het geëigende formaat.
  - Voorbeeld: een spreadsheet dus niet “uitprinten” als PDF, maar exporteren in het open .csv formaat (en dat eventueel lakken)
2. Anonimiseer *by design*, of probeer persoonsgegevens te vermijden, of markeer die meteen bij het opstellen van een stuk.
3. Pas de “lakstraat” alleen toe op niet zelf gemaakte stukken, en zorg voor de juiste afstelling.
  - dus liever geen afbeeldingen van teksten maken, en die inscannen, maar direct de PDF bewerken,
  - en als dat niet kan, gebruik dan goede OCR software (Tesseract is gratis en geeft vaak veel minder fouten dan nu veel gebruikte OCR software), stel die goed in (bijv, geef aan dat de tekst op de afbeeldingen in het Nederlands is), en zet de OCR ook aan.

### 6.1.2 De voorbeelden

#### Woo dossiers

De Woo dossiers uit de gemeentes [Barendrecht](#) en [Nijmegen](#) voldoen vrijwel geheel aan de eerder gegeven beschrijving. Deze twee links komen uit de [Woo-index](#). Voor Nijmegen is het niet verwonderlijk dat ze zo goed voldoen: zij publiceren via een *tool* (gewoon een invulformulier plus een manier om een folder op de PC van een Woo-jurist te *uploaden*) gebaseerd op het boven geschetste model. De [openpub API](#) voor Woo-dossiers die Barendrecht en verschillende andere gemeentes gebruiken is heel vergelijkbaar.

Het enige dat nog beter kan is de inventarislijst en de koppeling met de bundel vrijgegeven documenten. Die koppeling zou niet pas op het moment van publiceren gemaakt moeten worden maar liever, ondersteund door software en een werkwijze, op het moment van informatie vergaren.

We hebben de Woo-juristen verantwoordelijk voor deze voorbeelden gevraagd naar hun ervaring met deze wijze van publiceren:

“Het klaarmaken voor publicatie van een afgerond Woo-dossier is eigenlijk zo gedaan. Het formulier waarmee we de metadata invullen werkt makkelijk en maakt een laatste check mogelijk.” [Suzanne Vloet, Woo-jurist, Gem. Nijmegen]

“Het publiceren van een afgerond Woo-verzoek kost slechts enkele minuten, terwijl de (tijds)winst door de publicatie groot is. Meer informatie is op een makkelijke manier vindbaar voor de burger. Dit versterkt het vertrouwen van de burger in de overheid en draagt bij aan de democratische rechtsstaat.” [Patty van Woerkom, Woo-adviseur, gemeenten Barendrecht, Albrandswaard en Ridderkerk]

#### Vindbare en toegankelijke stukken

Bij de Provincie Flevoland vonden we dit mooie [voorbeeld van een Woo besluit](#) dat geanonimiseerd is *by design*. Ook [dit besluit van het Ministerie van SZW](#) lijkt direct uit een tekstverwerker te komen en is waarschijnlijk tijdens het maakproces geanonimiseerd. Het verschil in *vindbaarheid* en *toegankelijkheid* (de F en de A in FAIR) is mooi te testen door ze te vergelijken met een [ander besluit](#) (dit staat op de [vernieuwe Woo website](#) van het Ministerie van VWS) door

1. een woord dat je ziet staan in de tekst op te zoeken met Control F of iets te selecteren en te kopiëren met Control C.
2. je voor te stellen dat je visueel gehandicapt bent en de tekst alleen met extreme vergroting kunt lezen.

We vonden zelfs de twee uitersten binnen één bestuursorgaan, de Raad voor de Rechtspraak. Een verzoek dat afgewezen wordt, zoals [dit](#) wordt keurig als HTML gepubliceerd (dus vind- en doorzoekbaar, en zeer toegankelijk), maar wordt het

---

toegewezen, zoals [dit](#) dan wordt het besluit als scan, zonder OCR, gepubliceerd. Beide besluiten zijn opgebouwd vanuit hetzelfde sjabloon.

## PDF is niet zaligmakend

Het lijkt alsof binnen de Woo-wereld openbaar maken gelijk staat aan vrijgeven in PDF formaat. Dat dit de gangbare praktijk is heeft allerlei oorzaken en redenen, maar het kan ook anders. We proberen nu een paar redenen die we horen te ontcrachten:

- met PDF vries je een versie vast, een Word bestand kan je zo veranderen
  - dit is onwaar, het is alleen wat lastiger om een PDF te bewerken. Een mooi voorbeeld is de [PDFoppepper](#) waarmee onder andere machine leesbare tekst aan een PDF wordt toegevoegd.
- PDF is een open machine-leesbaar en -verwerkbaar formaat
  - dit is ten dele waar, een PDF kán die eigenschappen hebben en is dan inderdaad ideaal te verwerken. Maar men kan bijna alles in een PDF stoppen, en dus ook een prima machine leesbaar Word document veranderen in een PDF die in essentie bestaat uit een zip bestand met afbeeldingen van elke bladzijde en niks meer (dus eigenlijk niet veel anders dan een verzameling vakantiekiekjes).

Hier nog een mooi voorbeeld van een [Woo dossier uit Waalwijk](#) waarin een aantal bijlagen in het originele formaat als echte Excel sheets worden aangeboden. Om ongewenste veranderingen te voorkomen zijn die als *read only* stukken bewaard.

Een tegenvoorbeeld over publiceren in het juiste formaat en dus niet altijd in PDF is het [Woo verzoek waarin expliciet om tabellen](#) gevraagd wordt. Inderdaad wordt er een enorm Excel bestand vrijgegeven, met bepaalde informatie consequent weggelakt, *maar in PDF*. Geprint als PDF beslaat [dit excel bestand](#) meer dan 70 pagina's. Dit is niet behulpzaam aan de verzoeker en kan slechts met zeer veel moeite als data hergebruikt worden.

## Officiële bekendmakingen: het goede voorbeeld

Stukken gepubliceerd op [officielebekendmakingen.nl](#) zoals deze [beschikking van het Hoogheemraadschap van Delfland](#) worden in zelfs 4 bestandsformaten aangeboden: in HTML en PDF, maar ook in 2 expliciet bewerkbare formaten: het open Word-formaat ODT, en een XML volgens een zelf vormgegeven opmaak schema. Daarnaast wordt uitgebreide metadata zowel netjes in een HTML tabel getoond als beschikbaar gesteld in een XML bestand volgens een gestandaardiseerd schema, dat deels gebaseerd is op de Dublin Core standaard.

## 6.2 Convenanten (2f)

Een convenant wordt als volgt beschreven in de [convenanten beslishulp](#) (blz 7):

Grofweg is een convenant een schriftelijke set van afspraken tussen een overheidsorganisatie en één of meer andere partijen. Het doel van de afspraken is het voorbereiden of realiseren van beleid of het uitvoeren van een overheidstaak.

Een convenant is bij uitstek een zogenaamd *semi-gestructureerd document*<sup>4</sup>, het bevat tekst (de afspraken), maar ook allerlei meer "linked data achtige" gegevens zoals

- de betrokken partijen, zowel de ondertekenaars als de organisaties die ze vertegenwoordigen;
- de datum en plaats van ondertekening;

---

<sup>4</sup> Op Wikipedia wordt dit als volgt omschreven: *Semi-structured data*[1] is a form of structured data that does not obey the tabular structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Therefore, it is also known as self-describing structure. Een helder voorbeeld is dit stukje uit het Waterschapsblad weergegeven in XML.

- de looptijden van de afspraken.

Deze gegevens zouden we graag in een gestructureerd formaat beschikbaar hebben, eigenlijk als *metadata*. Deze gegevens hebben de vorm van zogenaamde *linked data* omdat ze bestaan uit onderling verbonden entiteiten. Linked data werkt alleen als er gestandaardiseerde waarden gebruikt worden voor objecten en relaties. Gelukkig heeft elk overheidsorgaan dat onder de Woo valt een unieke code, de *tooi code* (bijvoorbeeld gm1963 voor de gemeente Hoeksche Waard). Datums en periodes zijn ook (ISO)-gestandaardiseerd. Voor de organisaties met wie een afspraak gemaakt wordt moet er dan iets handigs gekozen worden. Vaak wordt de Wikipedia URL gebruikt, maar een KvK-nummer of iets dergelijks zou ook goed kunnen.

## 6.2.1 Samenvattend: model voor convenanten

- Het echte convenant als toegankelijke machine leesbare digital born PDF/UA. Geen scans of OCR. Haal zonodig persoonsgegevens weg *by design*.
- Metadata over bovengenoemde objecten en attributen van het convenant in een gestandaardiseerd formaat.
- De manier waarop de BAR-gemeentes hun convenanten beschrijven zit daar al heel dicht tegenaan, zie bijvoorbeeld de metadata van dit [afvalwater akkoord](#) hieronder. De BAR-gemeentes gebruiken het model voor convenanten van OpenWebConcept.



The screenshot shows a web browser window with the URL `ridderkerk.nl/convenanten/afvalwaterakkoord/`. The page features the Ridderkerk logo and navigation options like 'Direct regelen en informatie' and 'Bestuu'. The main content is a table with the following metadata:

Onderwerp	Afvalwaterakkoord
Samenvatting	In dit bestuursakkoord is afgesproken door de koepelorganisaties, dat het waterschap en de gemeenten moeten samenwerken in de afvalwaterketen "als ware het één bedrijf".
Beleidssterrein	Milieu
Duur	2022 tot 2025
Datum ondertekening	21 december 2022
Partijen	<ul style="list-style-type: none"> <li>• Waterschap Hollandse Delta</li> </ul>
Bijlagen	<ul style="list-style-type: none"> <li>• <a href="https://openpub.ridderkerk.nl/wp-content/uploads/sites/3/2023/02/ondertekend-afvalwaterakkoord-Ridderkerk-2022.pdf">https://openpub.ridderkerk.nl/wp-content/uploads/sites/3/2023/02/ondertekend-afvalwaterakkoord-Ridderkerk-2022.pdf</a></li> </ul>

**Figuur: Voorbeeld metadata van een Convenant uit de gemeente Ridderkerk.**

---

## 6.3 Beschikkingen (2k)

Voor de beschikkingen is nog geen handreiking beschikbaar, ook geen concept. We volgen hier daarom kort het lemma uit [Wikipedia](#). In de Algemene wet bestuursrecht (Awb) wordt onder beschikking verstaan “*een besluit dat niet van algemene strekking is, met inbegrip van het afwijzen van een aanvraag daarvan*” (artikel 1:3 lid 2, Awb).

In de data zien we vooral vergunningen, subsidie-verleningen en dwangsommen:

- een **vergunning of ontheffing** is een officiële (noodzakelijke) toestemming van de overheid om een bepaalde activiteit uit te voeren.
- een **dwangsom** is de verplichting een bepaalde geldsom te betalen als niet volgens een bepaalde gerechtelijke of bestuurlijke beslissing wordt gehandeld.

In de data zien we erg weinig metadata **specifiek over de beschikking** bij beschikkingen, zie bijvoorbeeld [deze beschikking](#). Eigenlijk de enige metadata die echt bij de beschikking hoort en die we vaak tegenkomen is een geografische aanduiding. Dat kan zijn in de vorm van een adres, maar ook in de vorm van 2 geo-coördinaten of zelfs een gebied aangeduid met geo-coördinaten. De Geo aanduiding in de metadata van het voorbeeld hierboven is

```
<meta
  name="OVERHEIDop.gebiedsmarkering"
  data-scheme="Adres"
  content="Hooilandseweg 5, 9982CA, Uithuizermeeden, Gemeente Het Hogeland"
/>
```

[Artikel 3.3 a van de Wet open overheid](#) geeft heel precies de gewenste attributen in de metadata aan. Die attributen komen in feite neer op de antwoorden op de volgende vragen:

### 6.3.1 Wie, wat, waar, wanneer, hoeveel, waarom?

Net als bij de convenanten is er bij de beschikkingen natuurlijk ook sprake van steeds weer vaste eigenschappen, uitgedrukt als antwoorden op bovenstaande vragen.

- wanneer gaat de beschikking in, en voor welke periode geldt ze?
- aan wie is de vergunning of subsidie verleend, of dwangsom opgelegd (voor zover de bescherming van de persoonlijke levenssfeer hieraan niet in de weg staat)?
- over welk gebied gaat het?
- om welk bedrag gaat het?
- om welke activiteit gaat het?
- met welke motivatie is de beschikking opgelegd?

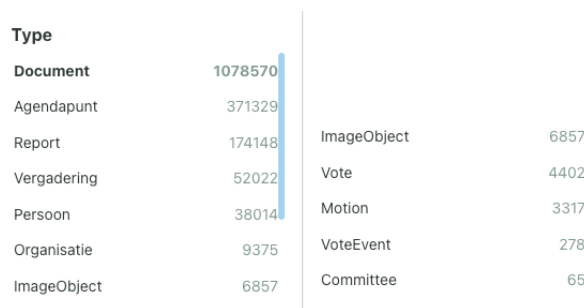
In het onderzoek naar de gepubliceerde beschikkingen kijken we hiernaar, en daarnaast of *het soort beschikking* in de metadata op een gestandaardiseerde manier is weergegeven.



---

## 6.4 Beleid en bestuurstukken (2c)

Voor de vergaderstukken van decentrale overheden (Woo Artikel 3.3 lid 2 sub c) is er een [werkdefinitie](#) vastgesteld. De werkdefinitie beschrijft wat er met een vergaderstuk bedoeld wordt maar specificceert niet *hoe* die stukken gepubliceerd moeten worden. De [Open API Specificatie voor Open Raadsinformatie](#) beschrijft de standaard voor het ontsluiten van raadsinformatie van gemeenten. In de handreiking wordt het voldoen aan deze standaard beschreven als een eerste stap. Het daargenoemde [informatiemodel](#) is een best complex *ER model*, waar in de praktijk van de website [openraadsinformatie.nl](#) niet veel valt terug te zien. De metadata van elk document daar bestaat uit drie velden: een (verder ongespecificeerde) datum, de naam van het bestuursorgaan, en het type document met op het moment van schrijven de volgende waarden en bijbehorende aantallen:



**Figuur: Soorten documenten en hun aantal op openraadsinformatie.nl.**

Omdat de stukken uit deze categorie enorm verschillen is deze metadata misschien ook wel het hoogst haalbare. In het [informatiemodel](#) wordt per document type een veel uitgebreider model behandeld.

Wat opvalt aan bovenstaand lijstje is de enorme hoeveelheid documenten van type *Document*. Dit betekent natuurlijk gewoon dat het juiste type niet bekend is.

## 6.5 Agendas en besluitenlijsten (2d)

Voor deze categorie is medio oktober alleen nog de [concept werkdefinitie](#) beschikbaar. Er zijn op het moment van het concept (2023-09-06) geen handreikingen voor deze categorie beschikbaar. In Woogie wordt duidelijk hoe weinig metadata er voor een [agenda beschikbaar is bij de Eerste Kamer](#). Dat is eigenlijk alleen een titel en 2 datums. Het onhandige is dat de datum waar het echt om gaat, namelijk de datum van de vergadering waar dit de agenda van is, niet in de gestandaardiseerde metadata staat, maar alleen in spreektaal in de titel.

### 6.5.1 Minimale metadata voor categorie 2c en 2d

- Naam en tooi code van het bestuursorgaan
- Semantisch type van het document, uit een vaste gestandaardiseerde lijst
- Datum van de **gebeurtenis** (de vergadering, de stemming, etc), mits relevant
- Datum van publicatie.

---

## 6.6 Organisatie en werkwijze en Bereikbaarheidsgegevens (1e)

Voor deze categorieën is een uitgebreide (concept) *handleiding* beschikbaar. Het bestaat eigenlijk uit 2 delen.

In de *Woo-index* zetten bestuursorganen URLs naar de plaatsen waar ze de verschillende Woo informatiecategorieën publiceren, plus de contact gegevens van de “Woo ambtenaar”. De Woo-index van *Zuid Holland* is een mooi ingevuld voorbeeld. Volgens de invulster Marianne de Nooij, senior beleidsmedewerker Open Overheid, was het invullen in een half uurtje gepiept<sup>5</sup>.

In de *bereikbaarheidsgegevens* staan allerlei persoonsgegevens over mensen werkzaam bij bestuursorganen, en over deelorganisaties binnen bestuursorganen. Dit is een complex (XML) model, met erg veel vrijheid voor de bestuursorganen, zoals we zullen zien als we gaan kijken wat er in zit. De handleiding erkent die vrijheid ook. Dit deel heeft verschillende namen, de oude *staatsalmanak* maar ook het *Register van Overheidsorganisaties (ROO)*. Dit register wordt deels vanuit andere databases gevuld. Veel bereikbaarheidsgegevens van gemeentes bijvoorbeeld worden via de databank VNG Ledengegevens aangeleverd aan het ROO. We zullen dit terugzien in de analyse van het ROO.

---

<sup>5</sup> Binnenlands Bestuur 2023-10-09 <https://www.binnenlandsbestuur.nl/digitaal/niet-alle-links-de-woo-index-zijn-even-bruikbaar>

---

## Bevindingen FAIRness per informatiecategorie

---

Dit hoofdstuk bevat de weerslag van ons onderzoek, en is dan ook het langst. Het is opgedeeld in secties per informatiecategorie plus een extra sectie die gaat over de blijvende beschikbaarheid van de stukken. Per informatiecategorie behandelen we dan de verschillende typen bestuursorganen. We kijken achtereenvolgens naar de besluiten op Woo-verzoeken, de bereikbaarheidsgegevens en de Woo-index, de convenanten, de beschikkingen, en de beleids- en bestuurs-stukken, agenda en besluitenlijsten van de lagere overheden.

### 7.1 Besluiten op Woo verzoeken

#### 7.1.1 Bekeken data

We hebben alle ministeries, alle provincies, alle waterschappen, alle ZBO's volgens [organisaties.overheid.nl](https://organisaties.overheid.nl), en alle 37 gemeenten en 9 ZBO's die eind September hun links naar hun Woo-besluiten op de [Woo-index](#) hadden staan via een gestandaardiseerd protocol bekeken. Dat ging als volgt: we zochten de website op waarop het bestuursorgaan haar Woo-besluiten plaatst, daarin namen we het 5de meest recente verzoek, en daarvoor keken we welke delen van het eerder geschetste Woo-dossier model aanwezig waren. Ook analyseerden we steeds de technische staat van 1 besluit en 1 vrijgegeven stuk. Eigenlijk alle websites die we vonden hadden een erg uniforme manier van publiceren. We keken na of het gekozen dossier een *outlier* was door nog een paar omliggende dossier te bekijken. Was het echt een outlier dan kozen we een dossier dat wel representatief was.

Daarnaast hebben we alle Woo-dossiers van betrokken bestuursorganen bekeken die aanwezig zijn in [Woogle](#). Per type bestuursorgaan geven we hieronder aan om welke data het precies gaat.

---

## 7.1.2 Ministeries

Alle ministeries publiceren hun Woo-dossiers op identieke wijze via [open.overheid.nl](https://open.overheid.nl) maar ook via [www.rijksoverheid.nl](https://www.rijksoverheid.nl). Wij hebben de dossiers uit het eerste portaal bekeken.

We vergelijken de hier gepubliceerde dossiers met het Woo dossier model. Dan zien we dat elk dossier heel handig een eigen pagina heeft met daarop wat metadata: een titel, de datum van publicatie en een “documentdatum”, en een korte beschrijving van het verzoek. We missen de datums van verzoek en besluit, en welk besluit er nou eigenlijk genomen is. Wat nou precies welk vrijgegeven stuk is, is niet altijd even duidelijk, zie het [onderstaande voorbeeld](#).

### Publicaties

 <a href="#">Publicatie op open.overheid.nl</a> Bestand: Besluit Woo verzoek.pdf
 <a href="#">Publicatie op open.overheid.nl</a> bundel
 <a href="#">Publicatie op open.overheid.nl</a> bundel
 <a href="#">Publicatie op open.overheid.nl</a> bundel

**Fig: Lijst van vrijgegeven stukken op dit Woo dossier op [open.overheid.nl](https://open.overheid.nl).**

We gaan nu in detail kijken naar de *besluiten* genomen door de ministeries.

## De brief met het besluit

Op 2023-09-27 hebben we van [open.overheid.nl](https://open.overheid.nl) alle Woo-verzoeken gepubliceerd door ministeries opgehaald: in totaal 4.374 besluitbrieven met in totaal 92.300 paginas. Bijna 80% van de besluiten bestaat uit 3-16 paginas. De laatste 20 % hebben er soms flink wat meer. Dit zijn dan veelal besluiten waarbij de (lange) inventarislijst of de vrijgegeven documenten als bijlage is toegevoegd aan de PDF van het besluit.

	2017	2018	2019	2020	2021	2022	2023
Ministerie Van Algemene Zaken	0	0	0	29	41	24	27
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	0	0	0	73	82	99	73
Ministerie Van Buitenlandse Zaken	0	0	0	56	50	81	38
Ministerie Van Defensie	0	0	1	26	51	63	38
Ministerie Van Economische Zaken En Klimaat	0	0	9	90	121	95	75
Ministerie Van Financiën	0	1	2	95	127	145	150
Ministerie Van Infrastructuur En Waterstaat	0	0	15	83	101	144	124
Ministerie Van Justitie En Veiligheid	1	0	2	95	109	147	128
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	2	4	7	204	261	277	201
Ministerie Van Onderwijs, Cultuur En Wetenschap	0	1	0	54	68	59	43
Ministerie Van Sociale Zaken En Werkgelegenheid	0	0	2	66	71	54	27
Ministerie Van Volksgezondheid, Welzijn En Sport	0	3	9	73	71	69	37

**Tabel: Aantal Woo-dossiers per ministerie per jaar.**

## Percentage documenten zonder machine leesbare tekst

Het algemeen percentage van 8% van de besluiten zonder machine leesbare tekst zegt niet veel want dit komt eigenlijk maar door een 4-tal ministeries, namelijk AZ, DEF, J&V en SZW. De rest zit allemaal onder de 5%. In 2023 zit alleen nog J&V rond de 20%

We concluderen dat dit eigenlijk heel goed gaat. Bijna alle ministeries zet nu standaard de “OCR aan”, of publiceert direct digitale stukken. We moeten hierbij wel aanmerken dat deze test alleen kijkt of er minstens één machine leesbaar karakter in de file staat. Het is dus best makkelijk om voor deze test te slagen.

	alle jaren	2023
Ministerie Van Algemene Zaken	15.7	0
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	4.6	1.4
Ministerie Van Buitenlandse Zaken	4.9	0
Ministerie Van Defensie	20.1	0
Ministerie Van Economische Zaken En Klimaat	8.2	2.7
Ministerie Van Financiën	4.8	0
Ministerie Van Infrastructuur En Waterstaat	7.7	0
Ministerie Van Justitie En Veiligheid	21.2	20.3
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	2.7	0
Ministerie Van Onderwijs, Cultuur En Wetenschap	4	0
Ministerie Van Sociale Zaken En Werkgelegenheid	19.5	0
Ministerie Van Volksgezondheid, Welzijn En Sport	0	0

**Tabel: Percentage besluiten op Woo dossiers zonder machine leesbare tekst, per ministerie, voor de gehele periode en alleen voor besluiten uit 2023.**

---

## Ingescande documenten

Hoe vaak is een besluitbrief (let op, dat is geen vrijgegeven, gevonden en vaak ouder, document, dit is een besluit dat net vers digitaal gemaakt is) nou *niet digitaal*? Dit meten we op door te tellen hoeveel van de paginas in zo'n brief volledig worden ingenomen door een afbeelding (dit is dan dus vrijwel altijd een afbeelding van de tekst die op die pagina van die brief staat).

Van alle 4.374 besluitbrieven bestaat 48% volledig uit afbeeldingen, 21% heeft beide soorten paginas, en gelukkig ook nog 31% van de brieven heeft géén één ingescande bladzijde.

Hieronder de uitsplitsing per ministerie. Behalve bij het ministerie SZW is hier geen duidelijke verbetering of verslechtering in 2023 te zien.

Ministerie	alle jaren	2023
Ministerie Van Algemene Zaken	92.6	96.3
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	95.3	95.7
Ministerie Van Buitenlandse Zaken	13.3	5.3
Ministerie Van Defensie	92.2	100
Ministerie Van Economische Zaken En Klimaat	46.9	32
Ministerie Van Financiën	21.7	20
Ministerie Van Infrastructuur En Waterstaat	50.8	55.6
Ministerie Van Justitie En Veiligheid	65.2	60.6
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	12.2	10.9
Ministerie Van Onderwijs, Cultuur En Wetenschap	65.6	76.7
Ministerie Van Sociale Zaken En Werkgelegenheid	58.9	0
Ministerie Van Volksgezondheid, Welzijn En Sport	90.5	75.7

**Tabel: Het percentage van de besluitbrieven per ministerie waarbij elke pagina een scan is, voor alle jaren en voor 2023.**

Ingescande documenten zijn ook onnodig groot: de gemiddelde *filesize* in Bytes van files zonder plaatjes is 36 KiloByte per pagina en van files waarbij elke pagina een plaatje is, 325 Kb, bijna 10 keer zo groot.

Het aantal besluiten dat volledig uit scans van pagina's bestaat is nog heel erg hoog, en in feite totaal onnodig. Let op, dit is het percentage besluiten waarbij **elke bladzijde volledig uit een afbeelding bestaat**. We zien soms PDFs waarbij de eerste paar bladzijden niet gescand zijn, maar daarna allemaal wel. Die tellen we in deze tabel dus niet eens mee.

## OCR en verhaspelingen

We zagen dat nu vrijwel altijd de OCR wordt aangezet om te proberen om de tekst die verloren is gegaan door het inscannen weer boven water te krijgen. Dit gaat natuurlijk heel vaak heel goed, maar het blijft AI, en maakt dus onherroepelijk fouten. Die zijn lastig met grote precisie op te sporen. OCR fouten kunnen een verhaspeling van 1 of een paar letters in een woord zijn. Een mens kan het vaak nog prima lezen/ontcijferen, bijvoorbeeld *OekraTne* waarbij de *i* als een *T* is aangezien. Let wel, een zoekmachine als Google repareert dit niet, en indexeert gewoon het woord dat het ziet staan. Met zoeken op *Oekraïne* (diacrieten hoeven niet bij Google), vind je een document waarin Oekraïne zo verhaspeld is dus niet terug<sup>1</sup>. OCR gaat soms ook de mist in met spaties, zoals te zien is bij deze voorbeelden in Woo-besluiten (deze tekst ziet een mens dus niet als ze de PDF leest (want ze leest een afbeelding), maar voor de computer staat dit er):

```
'MiriisterevanAlgemeneZaken': 1,  
'tus20C01250aEDerHaagKabernetMirnster': 1,  
'Bnnenho25flAADrHas': 1,
```

(Vervolgt op volgende pagina)

---

<sup>1</sup> Zoeken met Google op de term "Oekratne" levert een lijst van hits verwijzend naar websites van de Nederlandse overheid. Deze OCR fout komt dus veelvuldig voor.

```
'besluitinzakeAS1Lenexportvergunning': 1,
'Bijbriefvan26februari2020': 1,
'ontvangenop2maart2020': 1,
'ministerievvanAlgemeneZaken': 1,
'namenseenverzoekop': 1,
'grondvandeWetopenbaarheidvanbestuur': 1,
'alledocumentenoverAS1Lin': 1,
'combinatiemetdetermExtremeUltraviolet': 1,
'ofexportvergunningChina': 1,
'ofexportkredetverzekeringChina': 1,
'Uwverzoekbetreftdenode1januari2019totenmet26februari2020': 1,
```

## OCR verbeteren

We zien dat als we de OCR *opnieuw doen* met de (gratis) Tesseract OCR software, eerst ontwikkeld door HP en later overgenomen door Google, dat de kwaliteit omhoog gaat. De echt gekke fouten (vooral het missen van spaties) verdwijnen, en ook veel van de typische “OCR missers” (zoals een i en een l verwarren). De zoekmachine Woogole herhaalt voor elke pagina in elk document de OCR met Tesseract en vergroot daarmee de vindbaarheid.

## Conclusie en aanbeveling

Anonimiseer de besluiten op Woo-verzoeken *by design*, door in het brief-sjabloon *Geachte ..weg te lakken naam..* te vervangen door het anonieme *Geachte verzoeker*, en net zo voor de persoonsgegevens van de Woo-ambtenaar. Dan hoeft de brief niet door de laksoftware en kan direct vanuit de tekstverwerker geprint worden. Met als resultaat een keurige, goed machine leesbare PDF die, mits goed ingesteld, voldoet aan de PDF toegankelijkheidsnormen, waarbij de machine leesbare tekst exact gelijk is aan de mens-leesbare, en die ook nog eens vaak zo'n 10 keer kleiner (in Mb) is dan het vergelijkbare document uit de “scan-straat”.

## Alle vrijgegeven documenten in Woo-dossiers

We doen nu de “geen-tekst” en “pagina is een scan” analyse over alle documenten uit Woo-dossiers. Dat gaat om 630 duizend pagina's. Eerst kijken we hoeveel procent van de tekst per pagina is weggeklakt. Dat ligt per ministerie tussen de 1 en 6.6 procent gemiddeld.

	percentage (zwart)gelakte tekst
Ministerie Van Algemene Zaken	3
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	4.9
Ministerie Van Buitenlandse Zaken	3.2
Ministerie Van Defensie	5.1
Ministerie Van Economische Zaken En Klimaat	6.6
Ministerie Van Financiën	1.1
Ministerie Van Infrastructuur En Waterstaat	3.5
Ministerie Van Justitie En Veiligheid	3.1
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	5.1
Ministerie Van Onderwijs, Cultuur En Wetenschap	3.7
Ministerie Van Sociale Zaken En Werkgelegenheid	2.5
Ministerie Van Volksgezondheid, Welzijn En Sport	7.1

**Tabel: Gemiddeld percentage tekst dat is weggelakt per pagina, gegroepeerd per ministerie**

Over alle bestanden in Woo-dossiers gemeten heeft 18% van de paginas helemaal geen machine leesbare tekst en 8% van de bestanden. We zagen iets vergelijkbaars bij de besluitbrieven. Het aantal paginas dat een scan is is, zoals we kunnen verwachten, ontzettend hoog.

	2017	2018	2019	2020	2021	2022	2023
Ministerie Van Algemene Zaken	-	-	-	100	92	74	98
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	-	-	-	97	99	98	99
Ministerie Van Buitenlandse Zaken	-	-	-	82	88	80	90
Ministerie Van Defensie	-	-	100.0	77	99	87	73
Ministerie Van Economische Zaken En Klimaat	-	-	94.0	65	48	36	57
Ministerie Van Financiën	-	100.0	71.0	89	76	75	86
Ministerie Van Infrastructuur En Waterstaat	-	-	61.0	72	58	64	65
Ministerie Van Justitie En Veiligheid	98.0	-	100.0	94	65	82	68
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	50.0	46.0	9.0	44	33	40	52
Ministerie Van Onderwijs, Cultuur En Wetenschap	-	100.0	-	84	91	57	96
Ministerie Van Sociale Zaken En Werkgelegenheid	-	-	11.0	40	32	67	60
Ministerie Van Volksgezondheid, Welzijn En Sport	-	100.0	100.0	99	98	98	96

**Tabel: Percentage bladzijden die een scan zijn (N=660.000), uitgesplitst per jaar en ministerie.**

Voor we naar de andere bestuursorganen kijken vergelijken we de 4 typen bestuursorganen waar we Woo dossiers van hebben. 25% van de paginas (N=1.2 miljoen) heeft helemaal geen tekst, en 29% van alle bestanden (N=48 duizend). Het percentage bladzijden dat een scan is is 61%. Uitgesplitst over de jaren per soort organisatie ziet dat er als volgt uit:

organisatie type	2008	2010	2011	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
ZBO	-	-	100.0	97.0	99.0	89.0	83.0	99	65	51	48	45	69	-
gemeente	100.0	100.0	-	100.0	99.0	85.0	100.0	99	73	52	43	44	54	84.0
ministerie	-	-	-	-	-	-	-	57	92	43	77	64	65	71.0
provincie	-	-	-	-	-	-	67.0	86	74	51	26	25	60	-

**Tabel: Percentage bladzijden die een scan zijn (N=1.2 miljoen), uitgesplitst per jaar en type bestuursorgaan.**



---

### 7.1.3 Provincies

Een kwaliteits meting voor de provincies is in de herfst van 2022 uitgevoerd en beschreven in het [Archievenblad van Mei 2023](#). We hebben die analyse in de herfst van 2023 herhaald. De resultaten zijn zeer vergelijkbaar, en we behandelen ze dus kort en puntsgewijs.

#### Samenvattend

Negen van de 12 provincies publiceren Woo-dossiers. Friesland en Noord Brabant doen het niet. Overijssel geeft alleen een overzicht zonder de stukken. Hoe ziet een Woo-dossier er dan uit? Acht van de 9 provincies publiceren hun besluiten als apart bestand. Vier doen dat ook met het verzoek en 6 met de inventarislijst. Drie stoppen hun stukken netjes los in een zip bestand en één biedt ze los aan in een lijst. Dus jammer genoeg plakken 5 van de 9 hun vrijgegeven stukken nog aan elkaar.

Welke metadata geven ze dan? Acht (van de 10) bieden een datum aan, maar bij slechts 2 is het duidelijk dat het om de verzoeksdatum gaat en bij 1 dat het om de besluitsdatum gaat. Maar 2 (Overijssel en Zuid Holland) bieden de tekst van het verzoek en het besluit aan. Wat metadata betreft kunnen de provincies best wat leren van een gemeente als [Hoeksche Waard](#).

De [Woo-dossiers van de Provincie Zuid Holland](#) springen eruit. Netjes in een lijstje, handig doorzoekbaar, met mooie metadata, en de stukken gewoon los van elkaar, en vaak goed machine leesbaar.

#### Pluspunten

- **De besluiten zijn vaak beschikbaar als *digital born PDF documenten*.** Ze zijn dus prima machine leesbaar en vermijden de *scan en OCR* aanpak. De tekst is *vector based* wat betekent dat je enorm kunt inzoomen zonder kwaliteitsverlies. Dit geldt voor de provincies Groningen, Drente, Flevoland, Zuid-Holland, en Limburg. Flevoland valt op met een besluit dat *by design* geanonimiseerd is (Bijv *Geachte Verzoeker*). Zuid-Holland ook met een besluit dat zo uit Word lijkt te komen (mede door de extensie *.docx.pdf*; in de metadata van de PDF staat ook dat de content gemaakt is door *Microsoft Word 2016*).
- Al 3 van de 9 provincies die Woo-dossiers publiceren, publiceren de vrijgegeven documenten los in een **zip bestand**.

#### Aandachtspunten

- **Metadata blijft zorgelijk**, te veel ieder op haar eigen manier, zowel in wat er beschikbaar is, als de gebruikte naamgeving
- Crawl en indexeren door een zoekmachine blijft een breekbaar proces, zoals ook beschreven in [Elchik, \(2023\)<sup>2</sup>](#).

### 7.1.4 Gemeenten

Alle 37 gemeenten die eind september een link naar de pagina met hun Woo-dossiers in de Woo-index hadden ingevuld zijn bezocht en geanalyseerd. Een klein aantal had een verkeerde link of een link naar de plek om een Woo-verzoek in te dienen. Het merendeel had 1-50 Woo-dossiers, drie hadden er tussen de 50 en 100, en ook 3 meer dan 500 (Den Haag, Rotterdam en Amsterdam). De gemeente Utrecht had haar link nog niet ingevuld maar heeft al wel meer 1151 Woo-dossiers op Woogel staan.

De aanwezigheid van de documenten verschilt nog enorm. Soms is er alleen een beschrijving en geen enkel document, soms wel het besluit maar niet de vrijgegeven documenten. In die gevallen kan men wel altijd een mailtje sturen met het verzoek om de stukken te mogen ontvangen.

---

<sup>2</sup> [https://scripties.uba.uva.nl/search?id=record\\_52720](https://scripties.uba.uva.nl/search?id=record_52720)

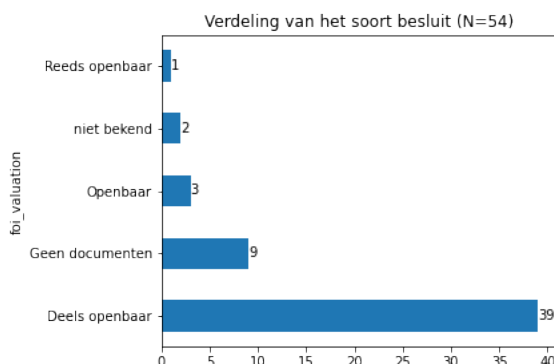
Als we iets uitzoemen, valt op dat gemeentes die gebruik maken van een gedeeld systeem eruit springen: dat zijn de gemeentes die het openpub/openwebconcept gebruiken, zij die Woogle gebruiken, en de gemeentes met een iBabs systeem en website. We geven drie voorbeelden:

- Nijmegen met Woogle
- Albrandswaard met openpub
- Zevenaar met iBabs

Wij weten van openpub en Woogle dat die gebruik maken van een handig invulformulier. Het lijkt erop dat de aanwezigheid van de (optionele) vragen naar metadata helpt, want het wordt ook trouw ingevuld. Bij iBabs zie je ook dat metadata consequent wordt ingevuld. Ook bieden deze systemen de mogelijkheid om de stukken los en getypeerd aan te leveren, en de gemeentes doen dat dan ook.

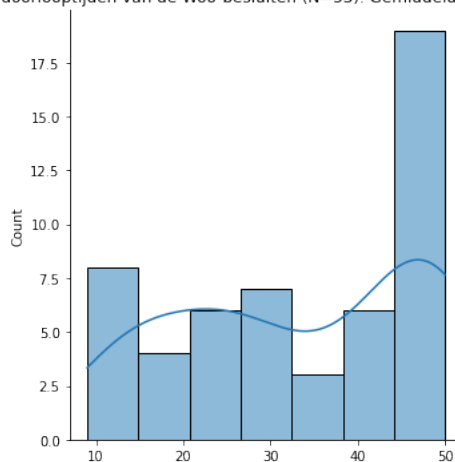
## Hergebruik de eigen metadata

Het invullen van metadata geeft niet alleen een helder overzicht op de website en vergemakkelijkt het zoeken, het biedt ook overzicht voor een bestuursorgaan zelf. Hieronder<sup>3</sup> een paar voorbeelden die met simpele Excel commando's gemaakt zijn op de 54 Woo-dossiers uit 2023 van de Gemeente Nijmegen.



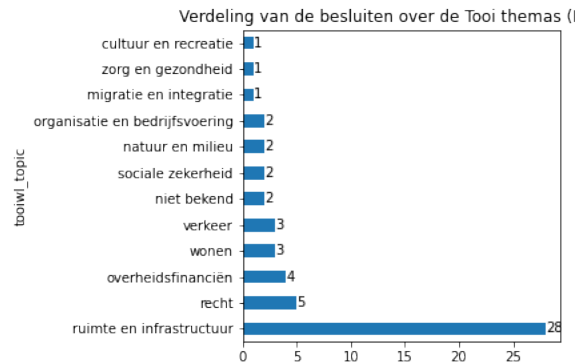
**Figuur: Verdeling van de 54 Woo dossiers uit 2023 van de Gemeente Nijmegen over 5 verschillende soorten besluiten.**

Histogram van de doorlooptijden van de Woo-besluiten (N=53). Gemiddelde beslistermijn is 33 dagen.

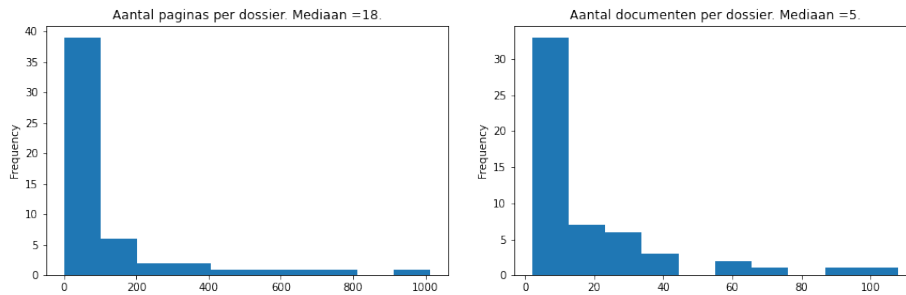


<sup>3</sup> De *pandas* code waarmee de grafiekjes gemaakt zijn uit de metadata is te vinden op <https://wooverheid.nl/2023/10/19/de-kracht-van-metadata/>.

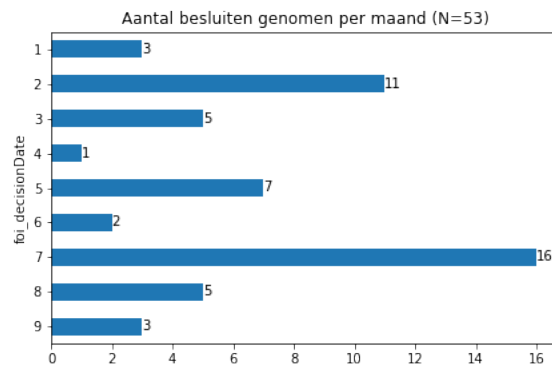
**Figuur: Histogram van de doorlooptijden van de 53 Woo dossiers uit 2023 van de Gemeente Nijmegen, gebaseerd op 2 datums, die van het verzoek en van het besluit.**



**Figuur: Verdeling van de dossiers over de verschillende thema's volgens de Tooi waardelijst.**



**Figuur: lengte van dossiers, in aantal bladzijden en aantal documenten. Automatisch bepaald.**



**Figuur: Hoeveel besluiten worden er in welke maand gepubliceerd? Op basis van besluit datum.**

---

## Kwaliteit van de bestanden

De kwaliteit van de bestanden in termen van machine leesbaarheid en toegankelijkheid varieert zoals ook bij de andere bestuursorganen enorm. Ook hier geldt dat samenwerken en leren van goede voorbeelden veel positiefs met relatief weinig inspanning kan opleveren. De openpub/openwebconcept aanpak lijkt dat uit te wijzen.

### 7.1.5 Waterschappen

Op één na hadden alle 21 waterschappen een pagina waarop werd aangegeven hoe een Woo-verzoek ingediend kan worden. Maar 8 van de 21 hadden ook een pagina met daarop de besluiten genomen over Woo-verzoeken. De aantallen zijn allemaal nog erg laag (maximaal 20). We zien dezelfde verscheidenheid als bij andere bestuursorganen, en soms ook houtje-touwtje oplossingen zoals het publiceren van de stukken op Google Drive.

Waar de waterschappen voorop lopen in hun bestuursinformatie en hun raadstukken lijken ze wat achter te lopen op het gebied van de Woo-dossiers. Dit lijkt een kwestie van tijd te zijn. Voor de 11 waterschappen die het ibabs systeem gebruiken is het eenvoudig om te schakelen naar een goede overzichtelijke presentatie zoals bijvoorbeeld het ibabs [Woo-overzicht van de gemeente Almelo](#).

### 7.1.6 Zelfstandige bestuursorganen (zbo's)

Alle bestuursorganen die als zbo's gekwalificeerd worden op [organisaties.overheid.nl](https://organisaties.overheid.nl) zijn onderzocht. Op het moment van onderzoek ging het om 153 stuks. Op het moment van schrijven staan er al [156 zbo's in Woog](#). Er is naar drie dingen gekeken:

1. Heeft de zbo een website?
2. Zo ja, is er een Woo contact persoon, of andere informatie over de Woo (zoals over het indienen van een Woo verzoek) te vinden?
3. Publiceert de zbo ook Woo-dossiers op haar website?

Op basis hiervan zijn de zbo's in 4 groepen opgedeeld:

Groep	Aantal	Beschrijving
1	49	Zbo's zonder digitaal publicatiekanaal
2	46	Zbo's met digitaal publicatiekanaal zonder informatie Woo
3	29	Zbo's met alleen bereikbaarheid/contact-gegevens voor indienen Woo-verzoek (incl. zbo's met korte beschrijvingen van Woo-verzoeken zonder beschikbaarheid van documenten)
4	29	Zbo's met bereikbaarheid/contact-gegevens voor indienen Woo-verzoek én publicaties van Woo-dossiers

#### Groep 1: Zbo's zonder digitaal publicatiekanaal

Deze groep omvat 49 zbo's die geen website beschikbaar hebben, met als gevolg geen mogelijkheid tot onderzoek naar de omgang met Woo. Hieruit zijn er 47 die een cluster van zbo's vormen of onderdeel zijn van een cluster<sup>4</sup>. Deze clusters van zbo's vallen niet onder de werking van de kaderwet zelfstandige bestuursorganen. In het zbo-register worden de overige twee zbo's; de Garantiewetcommissie en het Bureau Beheer Landbouwgronden (BBL) wel expliciet geclassificeerd onder de Woo.

Overige opmerkingen bij groep 1:

---

<sup>4</sup> Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2022, 5 april). *Contactgegevens zelfstandige bestuursorganen Overheid.nl*. Register van Overheidsorganisaties. [https://organisaties.overheid.nl/Zelfstandige\\_bestuursorganen](https://organisaties.overheid.nl/Zelfstandige_bestuursorganen)

- 
- 17 zbo's zijn een cluster van keuringsinstanties.
  - 2 zbo's zijn gerelateerd aan de Wet langdurige zorg (Wlz).
  - 2 zbo's zijn representatieve rechtspersonen.

## **Groep 2: Zbo's met digitaal publicatiekanaal zonder informatie Woo**

Deze groep omvat 46 zbo's die wel een website hebben, maar waar de informatie over Woo ontbreekt. Met 'informatie' wordt verwezen naar de bereikbaarheid/contact-gegevens voor het indienen van een Woo-verzoek, uitleg over de werking van Woo en publicaties van Woo-dossiers. In de lijst van zbo's in deze groep zit een groot aandeel van de stichtingen (13 van de 21) die vallen onder verschillende ministeries. Daarbij bevat deze lijst alle registratiecommissies die vallen onder het ministerie van Volksgezondheid, Welzijn en Sport (VWS)<sup>5</sup>.

## **Groep 3: Zbo's met alleen bereikbaarheid/contact-gegevens voor indienen Woo-verzoek**

Deze groep omvat 29 zbo's die een website hebben met bereikbaarheid/contact-gegevens voor het indienen van een Woo-verzoek. Wat opvallend is, is dat in deze lijst ook enkele grote zbo's voorkomen, zoals de Kamer van Koophandel, de Sociale Verzekeringsbank en het Zorginstituut. Verder zijn er geen opmerkingen over de zbo's in deze groep.

## **Groep 4: Zbo's met bereikbaarheid/contact-gegevens voor indienen Woo-verzoek én publicaties**

Deze groep omvat 29 zbo's die een website hebben met bereikbaarheid/contact-gegevens voor het indienen van een Woo-verzoek én tevens publicaties bevatten van Woo-dossiers. De zbo's kunnen ingedeeld worden naar het aantal beschikbare publicaties van Woo-dossiers: 9 hebben 1-5 publicaties; ook 9 hebben er 6-20; 6 tussen de 20 en 50; 2 tussen de 50 en 100, en 3 meer dan 100.

De publicaties worden door iedere zbo op een andere wijze gepresenteerd op hun website met afwisselende hoeveelheid metadata. We onderscheiden 4 publicatievormen:

1. (9 zbo's) Woo-verzoek/Woo-stukken onder elkaar in lijst van (pdf-)documenten
2. (6 zbo's) Woo-verzoek/Woo-stukken ingedeeld per dossier (op basis van titel of datum) zonder eigen URL
3. (7 zbo's) Woo-verzoek/Woo-stukken ingedeeld per dossier met eigen URL, een duidelijke titel en het Woo-besluit los beschikbaar
4. (7 zbo's) Woo-verzoek/Woo-stukken ingedeeld per dossier met eigen URL, een duidelijke titel, Woo-stukken los beschikbaar (incl. het Woo-besluit los) en full-tekst doorzoekbare documenten

Ten slotte nog wat losse opmerkingen over deze laatste groep. Het aantal zbo's in groep 4 of dat meer dan 20 publicaties bevat heeft geen relatie met de ministerie waar het onder valt. De aanwezigheid van metadata is per zbo in alle groepen afwisselend, met name de publicatiedatum, verzoekdatum en besluitdatum. Van het aantal zbo's in groep 4 hebben drie meer dan 20 publicaties (~43%). Er zijn een aantal (3) zbo's die verdeeld zijn over de groepen 1 en 3 waarbij er wel publicaties beschikbaar zijn op de website, maar de optie voor het indienen van een Woo-verzoek ontbreekt. In de volgende tabel geven we aan welk percentage van de 29 zbo's in groep 4 de basale metadata van een Woo dossier bevatte.

---

<sup>5</sup> Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2023, 23 oktober). *Contactgegevens Volksgezondheid, Welzijn en sport Overheid.nl*. Register van Overheidsorganisaties. [https://organisaties.overheid.nl/8591/Volksgezondheid\\_Welzijn\\_en\\_Sport](https://organisaties.overheid.nl/8591/Volksgezondheid_Welzijn_en_Sport)

Eigenschap van Woo dossier	Percentage (N=29)
titel	65.5
beschrijving	48.3
datum	69
datum verzoek	20.7
datum besluit	27.6
verzoek-tekst	41.4
besluit-tekst	20.7
Stukken beschikbaar	79.3
verzoek los beschikbaar	17.2
besluit los beschikbaar	65.5
inventarislijst los beschikbaar	24.1
stukken los beschikbaar	34.5
stukken in zip beschikbaar	10.3

En hiermee sluit het onderzoek naar de Woo dossiers bij de verschillende bestuursorganen.

## 7.2 Bereikbaarheidsgegevens en Woo-index

We hebben gekeken hoe goed de bereikbaarheidsgegevens van de bestuursorganen beschikbaar zijn op de nieuwe versie van de “Staatsalmanak”, [organisaties.overheid.nl](https://organisaties.overheid.nl). De data is opgehaald uit de dagelijkse dump die gepubliceerd wordt in [dit XML bestand](#). Daarnaast hebben we voor alle waterschappen bekeken hoe de bestuursinformatie en bereikbaarheidsgegevens op hun eigen site was weergegeven.

### 7.2.1 Vertegenwoordiging van de bestuursorganen

De eerste tabel geeft aan hoeveel van de bestuursorganen die een Tooi-code hebben, daadwerkelijk zijn terug te vinden in de staatsalmanak. We zien dat dit bij de vertegenwoordigende bestuursorganen en de ministeries prima in orde is, maar dat dit percentage bij de overige bestuursorganen een stuk lager ligt.

	Aantal in Woo_index	Aantal orgs	Percentage in Woo-index
Gemeenten (gm)	342	342	100
Samenwerkingsorganen (so)	99	356	28
Overige bestuursorganen (oorg)	55	140	39
Zelfstandige bestuursorganen (zb)	25	157	16
Waterschappen (ws)	21	21	100
Ministeries (mnre)	12	12	100
Provincies (pv)	10	12	83

**Tabel Aantal organisaties in Woo-index (N=565) en aantal organisaties met een tooi-code (N=1.043), uitgesplitst per type bestuursorgaan. Meting gedaan op 2023-10-17.**

## 7.2.2 Medewerkers en functies

Bereikbaarheidsgegevens gaan natuurlijk over de bereikbaarheid van iemand of iets. Dat kan dus een specifiek persoon, een bepaalde functie, of een (deel-)organisatie binnen het bestuursorgaan zijn. We tellen hier hoeveel medewerkers er in de almanak vermeld staan, en tonen dat per soort bestuursorgaan.

We tellen ook het aantal verschillende namen van functies zoals die voorkomen per organisatie, zoals hier bijvoorbeeld voor [Gemeente Almere](#) die 7 functies heeft. We zeggen bewust *namen van functies* en niet *functies* omdat die benamingen niet gestandaardiseerd zijn en we dus erg veel *verschillende manieren in de almanak zien om naar dezelfde functie te verwijzen*. Dit maakt het *gebruik en hergebruik* van de data in de staatsalmanak lastig. Zo kan men bijvoorbeeld niet met een simpele “database vraag” als “geef me alle commissarissen van de Koning” de lijst met actuele commissarissen krijgen.

Eén functie heeft een speciale status binnen de Woo gekregen, die van Woo-contact persoon. Dus is ook geteld hoe vaak die al in de almanak staat.

	gm	mnre	oorg	pv	so	ws	zb
Medewerkers	14187	736	563	349	256	51	32
Organisaties	342	12	140	12	356	21	157
Organisaties met Medewerkers	342	12	55	10	99	21	25
Organisaties zonder Medewerkers	0	0	85	2	257	0	132
Fractie Organisaties zonder Medewerkers	0	0	0.61	0.17	0.72	0	0.84
Gemiddeld aantal Medewerkers per Organisatie	41	61	10	34	2	2	1
Fractie Organisaties met een Woo-contactpersoon	0.26	0.25	0.09	0.7	0.74	0.43	0.92
Aantal Functies	9	331	116	43	64	6	5
Fractie Functies dat 1x Voorkomt	0	0.85	0.76	0.49	0.56	0.33	0.4
Fractie Functies dat 1x Voorkomt na Normalisatie	0	0.39	0.49	0.35	0.55	0.33	0.4

**Tabel: Overzicht<sup>1</sup> van medewerkers, functies en Woo-contact persoon van bestuursorganen die zijn te vinden op [organisaties.overheid.nl](https://organisaties.overheid.nl), gegroepeerd per type bestuursorgaan.**

<sup>1</sup> De rijen hebben de volgende betekenis: **Medewerkers:** Telt het aantal medewerkers die per organisatietype beschikbaar zijn op [organisaties.overheid.nl](https://organisaties.overheid.nl). Bij de de TOOI-soort gm is dit dus het totale aantal van alle medewerkers van alle gemeenten. **Organisaties:** Telt het aantal organisaties wat binnen elke TOOI-soort beschikbaar zijn. Zo heeft de TOOI-soort pv (Provincies) logischerwijs 12 verschillende organisaties. **Organisaties met Medewerkers:** Dit is het aantal organisaties die ook echt vindbare medewerkers hebben op [organisaties.overheid.nl](https://organisaties.overheid.nl). **Organisaties zonder Medewerkers:** En dit zijn de organisaties die geen vindbare medewerkers hebben op [organisaties.overheid.nl](https://organisaties.overheid.nl). **Fractie Organisaties zonder Medewerkers:** Het aantal organisaties zonder medewerkers gedeeld door het totaal aantal organisaties. **Medewerkers per Organisatie** Dit is het gemiddeld aantal medewerkers per organisatie met medewerkers. Dit wordt berekend door het totaal aantal medewerkers te delen door het aantal organisaties met medewerkers. **Fractie Organisaties met een Woo-contactpersoon:** Een nieuwe ontwikkeling op [organisaties.overheid.nl](https://organisaties.overheid.nl) is de toevoeging van de woo-contactpersoon functie. Hier is de fractie van het aantal organisaties met medewerkers dat ook een woo-contactpersoon heeft. **Aantal Functies:** telt het aantal verschillende functie namen zoals die voorkomen per organisatie, zoals hier bijvoorbeeld voor [Gemeente Almere](#) die 7 functies heeft. **Fractie Functies dat 1x Voorkomt:** Omdat het aantal verschillende functies zo groot is tussen de organisatietypen is gekeken hoe vaak een functietitel maar 1 keer over alle organisaties voorkomt. Dit kan bijvoorbeeld betekenen dat een functie daadwerkelijk maar 1x voorkomt (minister-president) of dat er een bijvoorbeeld sprake is van een spellingsvariatie (bijvoorbeeld Commissaris van de Koning of cvdk). **Fractie Functies dat 1x Voorkomt na Normalisatie:** Door normalisatie, in dit geval het eerste woord en zonder verschillen in hoofdletters, kunnen vergelijkbare functies gegroepeerd worden en zo zouden er minder functies slechts 1 keer voor moeten komen. Bijvoorbeeld alle verschillende ministers en staatssecretarissen worden nu genormaliseerd naar minister en staatssecretaris.

## 7.2.3 Informatie per medewerker

Ten slotte kijken we per type bestuursorgaan welke eigenschappen aanwezig zijn voor hun medewerkers. We geven dat als percentage. De namen die we hier gebruiken komen uit twee zogenaamde *namespaces*, de *Friend of a Friend (foaf)* en de *Freedom of Information (foi)* namespace. De betekenis spreekt redelijk voor zich en kan ook afgeleid worden uit dit voorbeeld.

eigenschap	gm	mnre	oorg	pv	so	ws	zb
Voornaam	0	0.31	0.15	0.11	0.45	0.06	0.19
Achternaam	0.99	0.94	0.99	0.98	0.71	0.84	0.19
Aanspreektitel	0.99	0.73	0.58	0.27	0.27	0.43	0.19
Initialen	0.99	0.94	0.99	0.98	0.71	0.84	0.19
Telefoonnummer	0.97	0.36	0.04	0.22	0.05	0.14	0.28
Politieke partij <sup>2</sup>	0.95	0.04	0	0.43	0.03	0	0
Bezoekadres	1	0.2	0.03	0.07	0.08	0.16	0.38
Postadres (postbus)	0.99	0.27	0.23	0.07	0.03	0.1	0.12
datum indiensttreding	0	0.21	0.54	0.08	0.7	0.61	0.72

**Tabel: Per type bestuursorgaan, het percentage van alle medewerkers waarbij het attribuut daadwerkelijk is ingevuld.**

## 7.2.4 Woo index

De *verwijsindex* uit het BIT advies over PLOOI, vorig jaar ook beloofd door de minister komt nu rap van de grond, en is in te zien op [organisaties.overheid.nl](https://organisaties.overheid.nl). We geven hier een overzicht van de stand van zaken eind September 2023.

Er zijn 402 bestuursorganen die minstens een deel van de Woo-index hebben ingevuld. Er zijn 17 informaticategorieën in te vullen. In totaal zijn er 1.672 links naar plaatsen op het web met Woo-informatie ingevoerd. Dit is een beetje een overtekening want meer dan 1.000 links zijn al automatisch ingevuld, voor 3 categorieën, die al centraal gepubliceerd worden.

organisatietype	count	mean	std	min	25%	50%	75%	max
Gemeenten (gm)	341	4	2	3	3	3	3	16
Ministeries (mnre)	10	7	3	5	5	5	7	12
oorg	3	4	2	2	3	4	5	6
Provincies (pv)	12	6	4	3	3	4	8	13
Waterschappen (ws)	21	6	4	3	3	3	6	16
Zelfstandige bestuursorganen (zb)	15	4	3	2	2	4	6	11

**Tabel: per organisatietype, de verdeling van het aantal ingevulde informaticategorieën. De *mean* van 7 bij het type mnre zegt dus dat elk van de 10 (zie de *count*) ministeries gemiddeld bij 7 van de informaticategorieën een link heeft opgegeven.**

De volgende tabel toont het aantal ingevulde links in de Woo index per categorie en per type bestuursorgaan. Wat opvalt is dat het aantal links hoger kan zijn dan het aantal bestuursorganen. Bijvoorbeeld 14 links naar beschikkingen voor 12 provincies. Dit komt omdat men per informaticategorie meerdere hyperlinks (met een toelichting) kan invullen.

<sup>2</sup> De naam van de politieke partij waarvan de medewerker lid is. Natuurlijk geldt dit voornamelijk voor de gemeenten, provincies en ministeries. Waterschappen zou men wellicht dit ook verwachten, maar helaas staan de gekozen leden niet in het register.



	gm	mnre	oorg	pv	ws	zb	totaal
Adviezen	6	10	0	0	1	1	18
Agenda's en besluitenlijsten bestuurscolleges	64	0	0	7	9	1	81
Bereikbaarheidsgegevens	55	3	3	4	7	13	85
Beschikkingen	345	10	0	14	22	3	394
Bij vertegenwoordigende organen ingekomen stukken	6	0	0	2	1	0	9
Convenanten	13	10	0	1	11	1	36
Jaarplannen en jaarverslagen	27	0	1	4	6	11	49
Klachtoordelen	13	0	0	3	2	2	20
Onderzoeksrapporten	10	3	0	1	1	3	18
Ontwerpen van wet- en regelgeving met adviesaanvraag	3	1	0	0	0	1	5
Organisatie en werkwijze	34	2	1	1	7	11	56
Overige besluiten van algemene strekking	343	10	1	12	21	2	389
Subsidieverplichtingen anders dan met beschikking	2	0	0	1	1	0	4
Vergaderstukken Staten-Generaal	0	0	1	0	0	0	1
Vergaderstukken decentrale overheden	40	0	0	3	3	2	48
Wetten en algemeen verbindende voorschriften	341	10	2	12	21	4	390
Woo-verzoeken en -besluiten	37	8	3	6	5	10	69

**Tabel: Per organisatietype en per Woo-categorie het aantal bestuursorganen die een link voor die categorie hebben ingevuld in de Woo-index.**

## 7.2.5 Conclusies

De combinatie van de oude Staatsalmanak en de nieuwe Woo-index op [organisaties.overheid.nl](https://organisaties.overheid.nl) is erg waardevol en wordt snel beter gevuld. Er is gekozen om de verantwoordelijkheid voor het invullen en bijhouden decentraal te laten. Daarnaast is er ook erg veel vrijheid in de manier van invullen en de te gebruiken terminologie. Dit zorgt voor een grote verscheidenheid aan zowel de volledigheid van de gegevens als de wijze waarop ze gecodeerd zijn. Dat maakt vooral hergebruik een stuk minder makkelijk. Het lijkt dat dit –naast een handige plek om iets specifiek terug te vinden– een belangrijk doel was. Dit maken we op uit het uitgebreide archief van de website met ook talloze verschillende (gedateerde) versies van de almanak.

Een paar opvallende punten zijn

- De bereikbaarheidsgegevens van de gemeentes komen uit een databank beheerd door de VNG. Dit is goed te zien in de hoge mate van volledigheid en uniformiteit van die gegevens.
- Bij de waterschappen is er een vergelijkbaar helder uniform systeem. Jammer genoeg ontbreken hier alle gegevens van de raadsleden. Die vinden we wel allemaal op de eigen websites.
- Bij de provincies lijkt iedereen het op een eigen manier te doen. Statenleden zijn bijvoorbeeld niet centraal bekend, maar wel decentraal op alle websites van de provincies. Die worden wel centraal bijgehouden door Woog, maar dat gaat via het breekbare *scraping*, zie bijvoorbeeld de lijst van [Zuid Holland](#).
- Een andere centraal bijgehouden categorie in de Woo-index, de beschikkingen, is voor de gemeentes, waterschappen, ministeries en provincies geheel gevuld. Die verwijzen allemaal naar dezelfde centrale plek, onder beheer van het Ministerie BZK. Ook hier zien we weer dat centrale sturing bij dit soort zaken loont.

Al met al is de Woo-index een waardevolle toevoeging aan het arsenaal aan overheidswebsites die iets doen met de Woo.

---

## 7.2.6 Waterschappen apart belicht

Waar de gegevens van de bestuurders voor de waterschappen in de Staatsalmanak beperkt blijven tot de dijkgraaf en een enkele andere bestuurder, zijn die op hun eigen websites in de meeste gevallen zeer volledig en aantrekkelijk weergegeven. De tabel hieronder geeft aan wat voor eigenschappen we dan tegenkomen.

Een voorbeeld dat eruit sprong was de pagina met bestuursinformatie van [waterschap de Dommel](#). Dit is gebaseerd op het systeem van *ibabs*, wat veel waterschappen gebruiken. Maar zelfs met zo'n zelfde systeem zie je toch dat bestuursorganen dat op een eigen wijze kunnen invullen. Wat hier opvalt is dat het lijkt alsof de bestuursleden zelf hun tekst hebben mogen invoeren, zie bijvoorbeeld de pagina van [Veerle Slegers](#). Ook opvallend is dat elke persoonlijke pagina eindigt met het hele stemgedrag, per onderwerp, en waarbij de persoonlijke stem is afgezet tegen die van de meerderheid. Bij waterschap [Drents Overijsselse Delta](#) valt op dat er per persoon een mooi overzicht is van activiteiten en van de toezeggingen die gehouden zijn in de vergaderingen.

Eigenschap per bestuurder	Aantal
partij	20
foto	20
email	21
functie in bestuur	17
nevenfuncties	14
Eigen pagina met eigen url	14
tel nummer	6
Sociale media links	3
Woonplaats	2
Bestuurlijke overleggen	1
Stemgedrag	1
geslacht	1
Activiteiten	1
Sprekersfragmenten	1

**Tabel: Hoeveel waterschappen deze eigenschap per bestuurder op hun site noemen (N=21).**

## 7.3 Convenanten

### 7.3.1 Bekeken data

We hebben de convenanten van de ministeries op [open.overheid.nl](#) (397 in oktober 2023), en van alle bestuursorganen die eind september hun links naar hun convenanten op de [Woo-index](#) hadden staan bekeken.

### 7.3.2 Ministeries

Alle ministeries publiceren hun Woo-dossiers op identieke wijze via [open.overheid.nl](#) maar ook via [www.rijksoverheid.nl](#). Wij hebben de dossiers uit het eerste portaal bekeken.

---

## Waar publiceren de ministeries hun convenanten?

Convenanten gesloten met een ministerie vinden we op minstens 3 plaatsen. Het is onduidelijk hoeveel overlap hier tussen zit. Zoeken op documenten met “convenant” in de titel levert weer andere hits op dan het beperken tot documenten met “convenant” als waarde van het facet rubriek of documentsoort.

Een hit op “convenant” in de titel levert vaak geen convenant zelf op, maar een bericht erover, bijvoorbeeld in de Staatscourant. Natuurlijk zit daar geen hyperlink bij naar het Convenant zelf.... Googlen helpt dan vaak wel, zo vonden we met Googlen op [Convenant+LSI+2017](#) het convenant waar we over lazen op <https://zoek.officielebekendmakingen.nl/stcrt-2023-26141.html>. Google vond dit convenant als eerste hit bij de [belastingdienst](#).

Bij zoeken op 2023-10-09 op de volgende 3 plekken vonden we heel verschillende aantallen convenanten:

1. [www.rijksoverheid.nl](http://www.rijksoverheid.nl): 133 hits met restrictie op document type
2. [open.overheid.nl](http://open.overheid.nl): 714 hits met restrictie op document type
  - Maar weer andere hits als je zoekt op hits met de term *convenant* in de titel van het document
  - De derde hit gaat over het *Convenant Mainport en Groen*. Maar de hit geeft alleen een brief. Even Googlen op “Convenant Mainport en Groen” levert al snel het convenant zelf op: op weer een andere plaats: <https://zoek.officielebekendmakingen.nl/stcrt-1996-235-p6-SC7570.pdf>, terwijl de eerste twee ondertekenaars toch echt ministers zijn
3. [zoek.officielebekendmakingen.nl](https://zoek.officielebekendmakingen.nl): 1092 hits met restrictie op document type. en 943 bij beperken tot alleen van ministeries 1. zoeken op “convenant” in de title levert maar 995 hits op. 2. En maar 806 met de beperking tot convenanten van de ministeries

De Woo index verwijst voor het ministerie van Defensie naar [zoek.officielebekendmakingen.nl](https://zoek.officielebekendmakingen.nl). Daar vinden we er 9. Op [open.overheid.nl](http://open.overheid.nl) leidt de restrictie tot het Ministerie van Defensie en documenten van type “convenant” tot 16 hits. En op [rijksoverheid.nl](http://rijksoverheid.nl) vinden we er geen één met deze beperking.

## De convenanten op [open.overheid.nl](http://open.overheid.nl)

Half oktober 2023 staan er op [open.overheid.nl](http://open.overheid.nl) 397 convenanten als volgt verdeeld over de ministeries:

	aantal
Ministerie Van Algemene Zaken	2
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	102
Ministerie Van Buitenlandse Zaken	22
Ministerie Van Defensie	16
Ministerie Van Economische Zaken En Klimaat	49
Ministerie Van Financiën	8
Ministerie Van Infrastructuur En Waterstaat	42
Ministerie Van Justitie En Veiligheid	14
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	28
Ministerie Van Onderwijs, Cultuur En Wetenschap	57
Ministerie Van Sociale Zaken En Werkgelegenheid	7
Ministerie Van Volksgezondheid, Welzijn En Sport	50

**Tabel: aantal convenanten per Ministerie op [open.overheid.nl](http://open.overheid.nl).**

En als volgt verdeeld over de jaren:

jaar	aantal
2008	1
2009	1
2010	18
2011	32
2012	48
2013	35
2014	38
2015	23
2016	20
2017	6
2018	38
2019	61
2020	22
2021	30
2022	16
2023	8

**Tabel: aantal convenanten per jaar op [open.overheid.nl](https://open.overheid.nl).**

### De kwaliteit van de gepubliceerde convenanten

Bij deze 397 convenanten horen 622 PDF documenten bestaande uit in totaal 8.658 paginas. We kijken weer naar de machineleesbaarheid van die stukken. We zien dat 20% van de paginas en 21% van alle bestanden helemaal geen enkel machine leesbaar karakter bevat. Uitgesplitst over de departementen zijn die percentages als volgt:

Departement	Percentage PDF zonder tekst
Ministerie Van Algemene Zaken	0
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	26.4
Ministerie Van Buitenlandse Zaken	22.7
Ministerie Van Defensie	25
Ministerie Van Economische Zaken En Klimaat	15.7
Ministerie Van Financiën	0
Ministerie Van Infrastructuur En Waterstaat	16.7
Ministerie Van Justitie En Veiligheid	21.4
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	7.1
Ministerie Van Onderwijs, Cultuur En Wetenschap	12.3
Ministerie Van Sociale Zaken En Werkgelegenheid	28.6
Ministerie Van Volksgezondheid, Welzijn En Sport	18.4

**Tabel: Percentage PDF bestanden behorend bij een convenant zonder machine leesbare tekst, per ministerie.**

---

## Ingescande documenten

Hoe vaak is een convenant nou **niet digitaal**? Dit meten we weer op door te tellen hoeveel van de pagina's in een convenant volledig worden ingenomen door een afbeelding (dit is dan dus een afbeelding van de tekst die op die pagina van het convenant staat). Dat zijn 30% van de 8.658 pagina's. Hieronder weer de uitsplitsing per ministerie. De enorme verschillen vallen erg op.

Departement	%
Ministerie Van Algemene Zaken	0
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	36
Ministerie Van Buitenlandse Zaken	26
Ministerie Van Defensie	59
Ministerie Van Economische Zaken En Klimaat	27
Ministerie Van Financiën	6
Ministerie Van Infrastructuur En Waterstaat	4
Ministerie Van Justitie En Veiligheid	21
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	79
Ministerie Van Onderwijs, Cultuur En Wetenschap	24
Ministerie Van Sociale Zaken En Werkgelegenheid	23
Ministerie Van Volksgezondheid, Welzijn En Sport	11

**Tabel: Het percentage bladzijden van convenanten per ministerie die een scan zijn.**

Ingescande documenten zijn ook onnodig groot: de gemiddelde *filesize* in Bytes van files zonder plaatjes is 36 KiloByte per pagina en van files waarbij elke pagina een plaatje is met 325 Kb bijna 10 keer zo groot.

## Conclusie

Het aantal convenanten dat uit een scan bestaat is nog heel erg hoog, en in feite totaal onnodig, omdat convenanten vaak vlak voor publicatie gemaakt worden.

## Aanbevelingen

1. Publiceer een convenant als een *digital born* PDF bestand. Is het nodig om de handtekeningen van de ondertekenaars ook in het gepubliceerde convenant te hebben, voeg die dan in als plaatje. Het omgekeerde –van alle tekst in het convenant een niet machine leesbaar plaatje maken– komt nog veel te veel voor.
2. Publiceer de stukken consequent op één vaste plek.
3. Publiceer convenanten als *linked data*, dus met minstens de datum en plaats van ondertekening, de ondertekenaars en de organisaties die ze vertegenwoordigen, en de loopdata van de afspraken als gestandaardiseerde metadata.

### 7.3.3 Bestuursorganen uit de Woo index

In de Woo index hadden eind september 10 ministeries en 18 andere bestuursorganen een verwijzing naar een pagina met hun convenanten ingevuld (12 gemeentes, 4 waterschappen, 1 provincie en 1 ZBO, het CBR). Daarvan waren er 7 nog slechts een “placeholder”, want daar stonden (nog) geen convenanten, 2 hadden er meer dan 10, en de rest tussen de 1 en 3.

Vijf gemeentes verwezen heel handig naar een zoekvraag op <https://zoek.officielebekendmakingen.nl/>, beperkt tot hun gemeente en convenanten. Vrijwel altijd verwijzen links ook echt naar de tekst van het convenant. Slechts 1 keer naar een aankondiging van de ondertekening in de Staatscourant.

Als we kijken naar de kwaliteit van de convenanten met artikel 2.4 lid 3 van de Woo in gedachten vinden we het volgende

- Convenanten op [officielebekendmakingen.nl](http://officielebekendmakingen.nl) zijn goed te lezen, beschikbaar in toegankelijk HTML en ook als PDF. Metadata is heel basaal.
- De presentatie en metadatering bij de BAR-gemeentes is weer heel mooi in orde. Een stralend voorbeeld voor de rest van de gemeentes, zie bijvoorbeeld [dit convenant](#). Dit komt omdat de BAR-gemeentes zijn aangesloten bij de openpub beweging en gebruikmaken van het openpub model van een convenant. Dat is in essentie een Google-form achtig invul formulier met vragen voor elk metadata veld, heel vergelijkbaar met wat ze ook gebruiken voor hun Woo-dossiers. Blijkbaar nodigt het hebben van zo'n formuliertje ook uit tot het invullen van de vragen (ook al zijn ze niet verplicht).
- De PDFs zijn jammer genoeg toch weer heel vaak scans, ook al is er niets weggelakt. Een slecht voorbeeld is een [convenant gesloten door het CBR](#). De handtekeningen op de eerste bladzijde en wat persoonsgegevens op de tweede zijn weggelakt, verder lijkt er niks meer gelakt. Toch is het hele document ingescand, zonder OCR, en vaak zeer slecht leesbaar, zie het screenshot hieronder. Het lijkt weggelakt, maar op andere plekken gebruikte men gewoon zwarte lak. Waarschijnlijk is het een lichter gekleurd font.
- De 4 waterschappen gebruikten het veld voor de convenanten in de Woo-index om aan te geven dat ze bij een gemeenschappelijke regeling zaten. Dat is natuurlijk wel een afspraak, maar dit lijkt niet de bedoeling. Ook een zoektocht op het internet per waterschap leverde heel weinig op: 1 keer een nu nog lege pagina, 1 waterschap met een lijst deelnames aan gemeenschappelijke regelingen, en 1 waterschap met een echte lijst met links naar convenanten waar ze aan deelnamen (het waren er maar 2, maar toch).



Figuur: Screenshot van een convenant gesloten door het CBR.

---

## 7.4 Beschikkingen

De beschikkingen (Woo categorie 2k) worden al geruime tijd via [officielebekendmakingen.nl](https://officielebekendmakingen.nl) gepubliceerd. Dit gebeurt via organen als de Staatscourant en het Waterschapsblad. Eind November 2023 stonden er daar 2.8 miljoen beschikkingen, als volgt verdeeld over de organen:

Blad	Aantal beschikkingen
Blad gemeenschappelijke regeling	733
Gemeentebblad	2571329
Provinciaal blad	66787
Staatscourant	53652
Waterschapsblad	110054

De eerste beschikkingen komen uit 1995. Alle bestuursorganen hebben keurig een verwijzing naar hun respectievelijke pagina met beschikkingen in de Woo-index staan. Die is er namelijk automatisch ingezet. Artikel 3.3.a beschrijft de verplichte metadata voor beschikkingen. Hier is op dit moment nog geen sprake van.

De technische kwaliteit van de stukken in de bladen uit bovengenoemde tabel is zeer goed. De beschikkingen zoals die in de bladen verschijnen worden gepubliceerd in meestal 4 formaten: prima toegankelijke HTML, en netjes opgemaakte PDF. Daarnaast wordt de bron gepubliceerd in 2 afzonderlijke XML bestanden, 1 voor de metadata, en 1 voor de tekst. Tot slot wordt het ook in ODT formaat vrijgegeven, een formaat dat ingelezen kan worden in Word. Dus de beschikkingen zijn **goed machine leesbaar**, en ook **goed toegankelijk voor visueel beperkten** (via de HTML versie).

Voor de rest van ons onderzoek naar de beschikkingen hebben we ons beperkt tot beschikkingen van de waterschappen. Omdat alle beschikkingen al zo'n lange tijd via hetzelfde kanaal openbaar gemaakt worden geloven we dat dit een representatief beeld schetst.

### 7.4.1 De inhoud van de beschikkingen

Artikel 3.3.a van de Woo geeft heel goed de semantische kern van een beschikking weer. Het zou inderdaad voortreffelijk zijn als de beschikkingen op deze gestructureerde uniforme manier machine lees- en verwerkbaar beschikbaar komen. Op dit moment is dat dus nog niet het geval. Nu bestaat een beschikking uit 4 onderdelen:

1. de officiële beschikking zoals gepubliceerd in bijvoorbeeld het waterschapsblad
2. metadata van deze publicatie
3. externe bijlages hierbij
4. een geografische aanduiding

#### De officiële beschikking

Heel vaak is de officiële beschikking erg kort (mediaan van 167 woorden (N=61.003)), en vaak wordt daar ook in aangekondigd dat de stukken een bepaalde periode ter inzage liggen of per email of telefonisch kunnen worden opgevraagd. Hieronder een voorbeeld uit [deze bron](#), waarbij we het deel over bezwaar maken (wat steeds terugkomt) hebben weggelaten.

De burgemeester van Rotterdam heeft besloten een exploitatievergunning voor de exploitatie van een coffeeshop te verlenen voor:

Adres: Witte de Withstraat 92A 3012BT Rotterdam Geldig vanaf:20-11-2023 Geldig tot en met:20-11-2024 Kenmerk: 0599ESUITE6196202023 Datum verzending besluit:20-11-2023

Voor meer informatie kunt u op werkdagen van 9.00 tot 17.00 uur contact opnemen met team Horecaverunningen, tel. 010 - 267 25 00 of per mail [horecaverunningen@rotterdam.nl](mailto:horecaverunningen@rotterdam.nl).

---

## De metadata

De metadata wordt handig als een apart XML bestand vrijgegeven. Hieronder een voorbeeld uit [deze bron](#). We zien dat er 4 *namespaces* gebruikt worden: de internationaal gestandaardiseerde Dublin Core attributen onder namespace DC en DCTERMS en twee zelf ontwikkelde onder OVERHEID en OVERHEIDop. Wat opvalt is dat er géén gebruik wordt gemaakt van de TOOI waardelijsten die bestaan voor bestuursorganen en categorieën. Dit maakt interoperabiliteit weer lastiger, omdat dan gekoppeld moet worden op strings als “Rotterdam” en “Openbare orde en veiligheid | Organisatie en beleid”.

```
<metadata_gegevens>
<metadata name="OVERHEID.organisationType" scheme="OVERHEID.Organisatietype" content=
  ↪"gemeente"/>
<metadata name="OVERHEIDop.publicationName" scheme="" content="Gemeentebled"/>
<metadata name="OVERHEIDop.doctype" scheme="" content="Officiële Publicaties, versie
  ↪1.1"/>
<metadata name="OVERHEIDop.configuratie" scheme="" content="https://repository.
  ↪officiële-overheidspublicaties.nl/MasterConfiguraties/MC-DRP-BeschikkingAanvraag-
  ↪3Pas-ZM/2.53/xml/MC-DRP-BeschikkingAanvraag-3Pas-ZM.xml"/>
<metadata name="DC.creator" scheme="OVERHEID.Gemeente" content="Rotterdam"/>
<metadata name="DC.type" scheme="OVERHEID.Informatietype" content="officiële
  ↪publicatie"/>
<metadata name="DCTERMS.publisher" scheme="OVERHEID.Gemeente" content="Rotterdam"/>
<metadata name="OVERHEID.authority" scheme="OVERHEID.Gemeente" content="Rotterdam"/>
<metadata name="OVERHEID.category" scheme="OVERHEID.TaxonomieBeleidsagendaDecentraal"
  ↪content="Openbare orde en veiligheid | Organisatie en beleid"/>
<metadata name="DC.type" scheme="OVERHEIDop.Rubriek" content="exploitatievergunning"/>
<metadata name="OVERHEIDop.referentienummer" scheme="" content="0599ESUITE6196202023"/
  ↪>
<metadata name="DCTERMS.abstract" scheme="" content="Coffeeshop Witte de With-
  ↪Exploitatievergunning coffeeshop"/>
<metadata name="DCTERMS.language" scheme="DCTERMS.RFC4646" content="nl"/>
<metadata name="OVERHEIDop.gebiedsmarkering" scheme="OVERHEIDop.locatietype" content=
  ↪"Punt">
<metadata name="OVERHEIDop.geometrieLabel" scheme="" content="Gebiedsmarkering"/>
<metadata name="OVERHEIDop.geometrie" scheme="" content="POINT(92282 436684)"/>
</metadata>
<metadata name="DC.title" scheme="" content="De burgemeester van Rotterdam heeft
  ↪besloten een exploitatievergunning voor de exploitatie van een coffeeshop te
  ↪verlenen voor:"/>
<metadata name="DCTERMS.available" scheme="DCTERMS.W3CDTF" content="2023-11-22"/>
<metadata name="OVERHEIDop.jaargang" scheme="DCTERMS.W3CDTF" content="2023"/>
<metadata name="OVERHEIDop.publicationIssue" scheme="" content="497792"/>
<metadata name="DC.identifier" scheme="OVERHEIDop.GmbID" content="gmb-2023-497792"/>
</metadata_gegevens>
```

## De externe bijlagen

Externe bijlagen bij een beschikking zijn enorm gevarieerd. Dat kan de gehele beschikking zelf zijn, kaartjes, fotos, en andere stukken. Juist hier zit natuurlijk de informatie die relevant is voor een burger die niet direct bij de beschikking betrokken is maar er allicht wel door wordt geraakt. Het is dan ook wenselijk dat al die externe bronnen handig bij de officiële beschikking openbaar gemaakt worden. We zien bij de waterschappen dat er een flinke variatie zit in deze manier van doen. Onderstaande tabel geeft dit handig weer. De waterschappen Limburg en Amstel, Gooi en Vecht zijn uitschieters in de rijkheid van hun gepubliceerde beschikkingen. Waterschap AA en Maas publiceert nooit een externe bijlage.



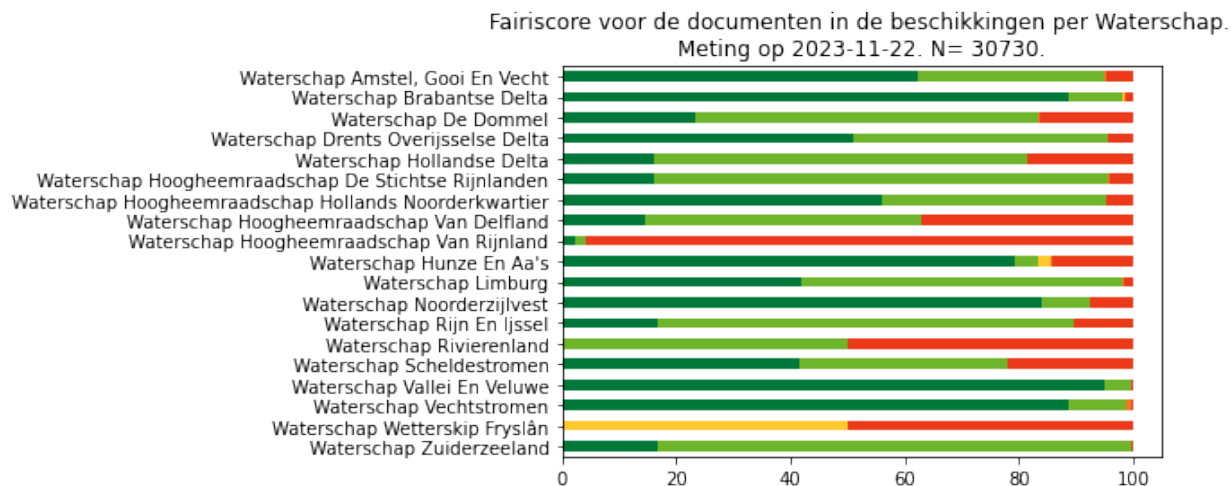
dc_publisher_name	Aantal Beschikkingen	Gemiddeld aantal docs per beschikking	Percentage beschikkingen bestaande uit 1 document	mediaan woordlengte beschikking in Waterschapscourant
Waterschap Aa En Maas	398	1	100	485
Waterschap Amstel, Gooi En Vecht	758	4.3	7.1	443
Waterschap Brabantse Delta	12470	2.1	28.1	683
Waterschap De Dommel	212	2.4	29.2	289
Waterschap Drents Overijsselse Delta	854	1.1	98	330
Waterschap Hollandse Delta	2557	1	98.1	326
Waterschap Hoogheemraadschap De Stichtse Rijnlanden	3627	1.4	76.6	224
Waterschap Hoogheemraadschap Hollands Noorderkwartier	15678	1	99.9	137
Waterschap Hoogheemraadschap Van Delfland	2546	2.1	3.2	359
Waterschap Hoogheemraadschap Van Rijnland	3944	1.7	27.4	287
Waterschap Hoogheemraadschap Van Schieland En De Krimpenerwaard	1376	1	100	155
Waterschap Hunze En Aa's	386	2	1	178
Waterschap Limburg	344	4.6	0.6	515
Waterschap Noorderzijlvest	1323	2.7	1.1	105
Waterschap Rijn En Ijssel	453	1.5	99.3	323
Waterschap Rivierenland	10359	1	100	153
Waterschap Scheldestromen	178	1.2	82	280
Waterschap Vallei En Veluwe	1014	1.8	17.1	436
Waterschap Vechtstromen	1312	2.2	31.6	171
Waterschap Wetterskip Fryslân	854	1	99.9	344
Waterschap Zuiderzeeland	359	2	1.1	310

## De Geo informatie

Iets meer dan 15% van de beschikkingen heeft een geo-aanduiding. Vanaf 2021 lijkt dit standaard ingevoerd en wordt dat vrijwel 100%. Dat kan op allerlei verschillende wijzes, maar het vaakst voorkomend is als 2 coördinaten waarmee een punt op een kaart kan worden aangegeven. In dat geval wordt dat punt op de kaart ook getoond, zoals [bij deze beschikking](#). Soms is er ook een compleet adres, en soms ook een gebied, ook netjes met coördinaten weergegeven. Dit is erg handig, want het maakt *geografisch zoeken* mogelijk (“geef me alle beschikkingen afgegeven in 2023 in een straal van 10km rond mijn huis”).

## 7.4.2 De machine leesbaarheid

Zoals gezegd is de machine leesbaarheid van de stukken in de Staatscourant en de overige bladen uitstekend. Maar zelfs voor de externe bijlagen is die vaak ook prima in orde. We zien hier een variatie aan bestandsformaten. Wat een goed teken is, omdat we dan dus vaak met het originele bestand te maken hebben. Bij een spreadsheet of iets uit een GIS systeem is dit natuurlijk veel waardevoller. Onderstaande grafiek geeft de verdelingen van de fairiscore voor de externe bestanden per waterschap. We zien ook hier nog wel wat variatie, maar over het algemeen is het best goed. We moeten hierbij aantekenen dat de fairiscore meting is afgesteld op PDFs met vrijwel alleen tekst. Het kan in sommige gevallen dus zijn dat de score lager uitvalt dan feitelijk juist is (bijvoorbeeld bij een kaartje met daarin handgeschreven informatie of een pijl dat een hele bladzijde beslaat met verder geen tekst; dit wordt dan gezien als een niet geOCRde scan en krijgt dus score E).



## 7.4.3 Conclusies en Adviezen

De beschikkingen vormen een Woo-categorie die er positief uitspringt wat betreft machine leesbaarheid, vindbaarheid en toegankelijkheid. Het feit dat de beschikkingen via een centraal orgaan worden verwerkt en ook centraal worden aangeboden zorgt voor hoge kwaliteit en goede vindbaarheid. Natuurlijk is er nog wel wat te wensen. Onze adviezen zijn dan ook:

1. Vermijd duplicatie van hetzelfde. Veel waterschappen hebben een lange pagina gemaakt met al hun (vaak duizenden) beschikkingen erop. Dit is niet meer van deze tijd. Mensen zoeken gewoon met Google. Die lijsten verwijzen dan naar de officiële plek waar alle beschikkingen staan. Daar staan ze duidelijk en netjes. Besteed de energie liever aan het verbeteren van de kwaliteit van de gepubliceerde beschikkingen zelf.
2. Maak het bijvoegen van externe documenten die de beschikking echt beschrijven de standaard optie.
3. En zorg dat die documenten FAIR zijn. Dus liefst in het originele machine verwerkbaar formaat, en als het PDFs zijn, geen scans, maar digitale documenten.
4. Maak bijvoegen van de rijke geo-informatie ook de standaard optie.
5. Gebruik de TOOI codes in de metadata voor het bestuursorgaan en de categorie en/of thema.
6. Zorg dat de PDF met de beschikking voldoet aan de WCAG norm.
7. Tot slot, maak werk van Artikel 3.3 lid a van de Woo, en zorg voor een mooie manier om die rijke metadata die daarin vereist wordt ook echt als FAIR metadata beschikbaar te maken.

---

## 7.5 Beleid en bestuurstukken

We hebben gekeken naar de beleid en bestuurstukken van de lagere overheden, de gemeenten, de provincies en de waterschappen. De gemeenten en een aantal provincies maken hun stukken beschikbaar via [OpenRaadsinformatie](#) en via eigen kanalen. De waterschappen alleen nog via eigen kanalen.

We bespreken hier onze bevindingen wat betreft de gemeenten en [OpenRaadsinformatie](#), en de waterschappen. Aangezien de provincies niet in de API van [OpenRaadsinformatie](#) te vinden waren hebben we die buiten beschouwing gelaten.

Wij hanteren een breed begrip van beleid en bestuurstukken: verslagen van raads- en commissie-vergaderingen, alsmede agendas en besluitenlijsten, plus alle stukken die op die vergaderingen besproken worden. Een mooi voorbeeld van een helder opgezette website met dit soort informatie is die van [Almelo](#). Dit voorbeeld van een [Politiek Beraad](#) geeft de complexiteit van zo'n vergadering weer. We hebben hier een sterk aan elkaar geknoopt geheel bestaande uit de [agenda](#), de videotulen van de hele avond, en per agendapunt een titel, korte samenvatting, namen van belangrijke spelers op dat punt en de bijbehorende stukken. Wat in dit voorbeeld opvalt is dat de “stukken achteraf”, bijvoorbeeld de gedane toezeggingen of genomen besluiten niet gekoppeld zijn aan deze gebeurtenis.

### 7.5.1 Gemeenten

De VNG en OpenState hebben in een gezamenlijk project een soort eigen PLOOI voor de beleid en bestuurstukken van de gemeenten en de provincies ontwikkeld: [OpenRaadsinformatie](#). Deze website wordt automatisch gevuld vanuit de zaaksystemen van de bestuursorganen, en dan min of meer in een gemeenschappelijk schema geplaatst. Informatie is zowel beschikbaar via de website als via een API. Via de website is er slechts weinig metadata per document beschikbaar, via de API is dat flink meer, maar valt de standaardisatie weg.

De zogeheten “rapporten” (dit zijn eigenlijk dossiers met wat metadata en één of meer bijbehorende stukken) van alle gemeenten uit de API zijn verzameld. Dit waren 107 gemeenten. Dit zijn in totaal 158.378 rapporten met 251.185 verschillende documenten. Samen hebben deze 1.601.812 pagina's. Het blijkt dat er in totaal 262 verschillende termen worden gebruikt om deze rapporten te classificeren. Hieronder de top daarvan, met het aantal rapporten dat zo geklassificeerd is. De grote verscheidenheid valt op. Voor gebruik vanuit één bestuursorgaan geeft dit natuurlijk niets. Voor een overzichts website als PLOOI of Woogie en ook Openraadsinformatie is dit natuurlijk erg onhandig.

```
'Ingekomen stukken': 18395, 'Moties': 13918, 'Raadsbrieven': 10888, 'Raadsvragen': 9578, 'Schriftelijke vragen': 8251, 'Toezeggingen': 7733, 'Brieven aan de raad': 7465, 'Export': 6634, 'Raadsinformatiebrieven': 4602, 'Amendementen': 4358, 'RSS': 4036, 'Besluitenlijst B&W': 3141, 'Collegebrieven': 2970, 'Informatieve Stukken College': 1777, "Informatienota's": 1663, 'Post aan de raad': 1503, 'Collegebesluiten ter informatie': 1430, 'Verslagen Raad': 1238, 'Lijst Ingekomen Stukken': 1204, "Memo's": 1148, 'Besluitenlijsten college': 1108, 'Schriftelijke vragen en antwoorden': 1083, 'Ingekomen stukken - College': 1080, 'Uitnodigingen': 1054, 'Artikel 43 vragen': 1012, 'Technische vragen': 1005, 'Besluitenlijsten B&W': 996, 'Griffie-info documenten': 994, 'Ingekomen stukken Cluster Sociaal': 987, 'Artikel 44 vragen': 940, 'Raadsvragen Schriftelijke': 892, 'Raadsinfo': 876, 'Notities ter info': 872, 'Ingekomen stukken raad': 866, 'Raadsvragen en antwoorden': 864, 'Verslagen': 843, 'Raadsbesluiten': 810, 'export': 802, 'B&W besluiten': 776, 'Schriftelijke vragen art. 41': 729, "Raadsmemo's": 720, 'Collegebesluiten': 646, 'Toezeggingen Export': 645, 'Verslagen M&S': 576, 'Besluitenlijsten': 565, 'Openbare besluitenlijsten college': 533, 'Ingekomen stukken Cluster Bestuur': 487, 'Stukkenstroom': 481, 'Commissie informatiebrieven': 453, 'Schriftelijke vragen (artikel 41 RvO)': 435, 'Verslagen S&R': 406
```

**Figuur: Top meest gebruikte termen samen met het aantal keer dat ze gebruikt zijn in de openraadsinformatie API om rapporteren te classificeren.**

---

In de API wordt ook de machineleesbare tekst van de stukken meegegeven. Het valt op dat die vaak van prima kwaliteit is, maar dat we ook best veel gekke, half leesbare, met OCR fouten doorspekte stukken tekst tegenkomen. Naast de niet gestandaardiseerde metadata is hier dus ook de kwaliteit van de machineleesbare tekst een probleem.

Om te kijken of de metadata niet makkelijk uniformer en rijker zou kunnen hebben we twee gemeentes, Hoorn en Utrecht, die beiden het ibabs bestuurlijkeinformatie systeem gebruiken dieper geanalyseerd. Het idee was dat allicht alle gemeentes met hetzelfde achterliggende systeem ook dezelfde terminologie zouden gebruiken. Jammer genoeg bleek dit, althans voor ibabs, een misvatting. Hoorn gebruikte 8 typen om hun documenten te klassificeren, en Utrecht 14. Slechts twee daarvan waren gedeeld: de *besluitenlijst B&W*, en *RSS*. Daarnaast hebben de documenten zelf ook weer metadata. Een motie heeft bijvoorbeeld indieners. Ook om die metadata aan te geven gebruiken deze twee gemeentes weer hun eigen terminologie. Hoorn heeft er 24 en Utrecht 18. Hiervan komen er 8 overeen: Gerelateerde items, Portefeuillehouder, Afgedaan, Deadline, Omschrijving, Onderwerp, Datum beantwoording en Opmerking. Dit inkijsje laat zien dat hier een wereld te winnen is.

## 7.5.2 Waterschappen

Van de 21 waterschappen gebruiken er 11 het *bestuurlijkeinformatie* systeem van ibabs, en 2 een notubiz systeem. Zoals in onderstaande tabel te zien is zijn de basisonderdelen overal wel aanwezig, maar verschillen de waterschappen nog best in de uitvoering. Ook de inrichting van de publicatielaag bovenop zo'n ibabs systeem kan op veel manieren gebeuren, en dat zien we ook.

Eigenschap	
Videotulen per vergadering	21
Agendas per vergadering	21
Elke vergadering eigen url	21
Alle agendas eigen URL	14
Vergadering als meerdere files	12
Staan ze handig bij elkaar? Makkelijk te downloaden?	12
Besluitenlijsten per vergadering	9
Elke besluitenlijst eigen url	5
Sprekers aangegeven met naam en tijdsaanduiding	1

**Tabel: Eigenschappen van de beleids en bestuurstukken, hoe vaak komen ze voor? (N=21)**

### Aanbevelingen

1. Het deel voorafgaand aan een vergadering (de agenda, en de te bespreken stukken) is prima in orde, maar het deel achteraf (de besluitenlijst) kan in veel gevallen duidelijker gepresenteerd worden.
2. Videotulen hebben als grote voordeel dat ze meteen beschikbaar zijn, tegen lage kosten. Het nadeel is dan weer dat er veel minder makkelijk in te zoeken is dan in uitgeschreven notulen of *Handelingen*. Dit kan voor een groot deel opgelost worden door automatische spraakherkenning toe te passen, en het zo ontstane transcript op te lijnen met de video. Daarbij komt zoeken met Control F, maar dan in een video voor iedereen beschikbaar. Ook kan het automatische transcript beschikbaar gesteld worden. Maurice Silvério heeft zo'n systeem ontwikkeld voor de gemeente Buren en beschreven in [zijn sctiptie](#).

---

### 7.5.3 Conclusie

De weerslag van de democratische processen in de lagere overheden gaat vaak heel goed op lokaal niveau. Maar om dit samen te brengen in één overkoepelend systeem waarin alles op dezelfde uniforme en rijke wijze beschikbaar is, zal nog een enorme klus zijn. Het openraadsinformatie initiatief laat zien dat het technisch mogelijk is alle stukken automatisch bij elkaar te brengen, en zo een soort PLOOI-achtig platform mogelijk te maken. Dat zou echter veel rijker kunnen als de metadata gestandaardiseerd zou zijn. De niet uniforme metadata zorgt dat de API van openraadsinformatie zowel voor *hergebruik* als voor het *koppelen van informatiebronnen* nog niet ideaal is.

Hier is nog een wereld te winnen!

## 7.6 Blijvende beschikbaarheid

Woogel verzamelt vanaf de herfst 2022 Woo-dossiers om die via haar zoekstelsel vindbaar en beschikbaar te maken. De hits in Woogel *verwijzen* naar de originele plek waar de stukken gevonden zijn, net als we gewend zijn bij *Google*. Het blijkt dat stukken toch nog best vaak “verdwijnen”; in ieder geval krijgt de gebruiker dan een zogenaamde 404 foutmelding: het gevraagde document is op de opgegeven URL niet beschikbaar. Dit is bijvoorbeeld het geval voor [deze URL](#). Het stuk lijkt verplaatst naar [open.overheid.nl](https://open.overheid.nl)<sup>1</sup>.

Hoe goed beschikbaar zijn stukken vrijgegeven onder de Woo nog binnen een jaar?

Om de vraag te beantwoorden hebben we voor een aantal Woo informatiecategorieën alle bron-urls uit Woogel verzameld, en die één voor één opgevraagd, en het resultaat opgetekend. We rapporteren puntsgewijs.

- **Beschikkingen** Van de 2.304 links naar (PDFs met) beschikkingen hadden er maar 5 een foutmelding. Deze kwamen van 4 aanbieders, de Kansspelautoriteit, de AFM en 2 waterschappen. De beschikkingen van de waterschappen staan allemaal op <https://zoek.officielebekendmakingen.nl/>.
- **Beslisnotas** Alle 8.092 URLs naar beslisnotas werkten nog steeds. Die staan allemaal op [open.overheid.nl](https://open.overheid.nl) en zijn afkomstig van ministeries. Een voorbeeld is <https://open.overheid.nl/documenten/121dcb64-75e6-4991-b5f8-0b603fd4b52b/file>.
- **Stukken op [rijksoverheid.nl](https://www.rijksoverheid.nl)** 19% van de 2.119 stukken in Woogel die zijn opgehaald van en verwijzen naar [rijksoverheid.nl](https://www.rijksoverheid.nl) geven een HTTP foutcode (en zijn daar dus niet meer op te halen of te bekijken). Een voorbeeld van zo'n URL (die ten tijden van schrijven wel op te halen was) is <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/wob-verzoeken/2022/02/18/besluit-wob-verzoek-representativiteitstoets-2010-bpf-houtwerkende-industrie/2021-0970-besluit.pdf>.
- **Besluiten op Woo-verzoeken** We hebben 7.669 stukken in Woogel die zo geclassificeerd zijn. Dit zijn altijd besluiten, en vaak is daar als bijlage nog een inventarislijst en soms ook de vrijgegeven stukken aangehecht. Een flink deel daarvan kon niet meer gevonden worden, zie onderstaande tabel, uitgesplitst naar domein/aanbieder. Hierbij moet aangetekend worden dat voorlopers zoals de gemeentes Amsterdam, Utrecht en Rotterdam hier relatief slecht vanaf komen, maar dat komt allicht omdat ze juist voorlopers waren. Positief vallen de openpub gemeentes op, allemaal foutloos.

---

<sup>1</sup> Zie <https://www.rijksoverheid.nl/documenten/woo-besluiten/2022/09/23/besluit-op-wob-woo-verzoek-over-dierenverhuurbedrijven>. In deze sectie beantwoorden we de volgende vraag:

site	#correct	#met foutmelding	percentage fout
https://www.utrecht.nl	272	750	73.4
https://wob.amsterdam.nl	258	315	55
https://www.rijksoverheid.nl	324	132	28.9
https://www.provinciegroningen.nl	14	2	12.5
https://rotterdambis.notubiz.nl	318	32	9.1
https://cms.dordrecht.nl	57	5	8.1
https://wobcovid19.rijksoverheid.nl	89	4	4.3
https://open.overheid.nl	3773	19	0.5
https://files.tweedekamer.nl	28	0	0
https://openpub.barendrecht.nl	6	0	0
https://openpub.ridderkerk.nl	9	0	0
https://www.flevoland.nl	13	0	0
http://www.zeeland.nl	21	0	0
https://www.zeeland.nl	25	0	0
https://www.provincie.drenthe.nl	58	0	0
https://www.provincie-utrecht.nl	43	0	0
https://www.zuid-holland.nl	43	0	0
https://wo.denhaag.nl	557	0	0
https://geol.arnhem.nl	81	0	0
https://openpub.gemeentehw.nl	94	0	0
https://www.noord-holland.nl	138	0	0
https://www.politie.nl	183	0	0
https://openpub.albrandswaard.nl	6	0	0
<b>totaal</b>	<b>6410</b>	<b>1259</b>	<b>16.4</b>

**Tabel: HTTP foutmeldingen op URLs in Woogle die verwijzen naar besluiten op Woo-verzoeken.**

## 7.6.1 Conclusie

Veel stukken zijn binnen maximaal een jaar niet meer beschikbaar op de plek waar ze eerst stonden. Dat betekent dat als men naar zo'n stuk verwijst, in een email, een rechterlijke uitspraak, een wetenschappelijk artikel, of wat dan ook, die verwijzing niets meer waard is. Het stuk is feitelijk weg, en zal opnieuw gevonden moeten worden.

Er zijn natuurlijk allerlei redenen om stukken te verplaatsen, een URL is een *adres*, geen *unieke identifier URI*. Maar het internet heeft daar technieken voor, namelijk een *redirect*, wat gewoon betekent dat een gebruiker met de oude URL doorgestuurd wordt naar de nieuwe locatie. Als een stuk echt verwijderd (vernietigd misschien wel) is, dan kan er gekozen worden om die boodschap te melden. Bij [rijksoverheid.nl](https://www.rijksoverheid.nl) lijkt het alsof men de boodschap wil afgeven dat de schuld bij de gebruiker ligt bij een 404 foutmelding. Die gebruiker krijgt namelijk dit te zien:

De pagina die u wilde zien of het bestand dat u wilde bekijken is niet gevonden. Controleer de URL.

De Woo gaat voor een groot deel over de betrouwbaarheid van de overheid. Fouten maken is helemaal niet erg, en hoeft betrouwbaarheid ook niet te schaden. De manier waarop met fouten wordt omgegaan kan dat wel doen.

---

## Voorstellen tot andere werkwijzes

---

In Hoofdstuk 3 hebben we naast het onderzoek naar de stand van zaken ook een aantal snel door te voeren technische aanbevelingen beloofd. Die zetten we in dit hoofdstuk op een rijtje. De meeste zijn *work in progress* waar het Woogle team op dit moment samen met voorlopende bestuursorganen aan werkt. Actuele informatie over de activiteiten van dit team zijn te vinden op het [wooverheid blog](#). Hier vind men ook contactgegevens.

Onze voorstellen zijn in 3 groepen verdeeld. De eerste verbeteren primair de vindbaarheid, de tweede de toegankelijkheid, en de derde eigenlijk alle vier de FAIR principes.

### 8.1 Maak echt werk van die machine leesbaarheid

Dit is de makkelijkst uitvoerbare groep, met een enorm groot effect op de waarde van de openbaar gemaakte stukken: ze worden echt goed (terug)vindbaar. Alles wat we hier beschrijven valt onder het **vermijden van de SCAN en OCR**.

#### 8.1.1 Anonimiseer *by design*

Heel veel stukken, waaronder Woo besluiten, worden gemaakt vanuit een sjabloon. Dit is veelal een MS Word bestand met lege velden die ingevuld worden, stukken tekst die geselecteerd of juist verwijderd worden, met daarnaast natuurlijk ruimte om het stuk aan de omstandigheid aan te passen. Tijdens dit proces kan direct in Word aangegeven worden wat weggelakt moet worden en onder welke grond. Juist bij een besluit, een beslisnota of een convenant zien we maar erg weinig tekst-redactie en veelal is dit volkomen gestandaardiseerd<sup>1</sup>. Op dit moment doet wooverheid onderzoek naar deze vorm van anonimiseren in samenwerking met een aantal lagere bestuursorganen.

---

<sup>1</sup> In een Woo besluit is dit meestal alleen de naam en andere persoonsgegevens van de ontvanger en de opsteller. Heel soms zijn er ook persoonsgegevens af te leiden uit de verwoording van het verzoek, zodat hier ook wat weggelakt moet worden. Dit gaat om nooit meer dan één paragraaf. Het inzetten van slimme laksoftware bespaart hier dus vrijwel niets.

---

## 8.1.2 Maak in Word, lak digitaal

Voor stukken zoals een besluit of een convenant waarvan de bron in MS Word beschikbaar is, maar waar zaken in weggelakt moeten worden bevelen we de volgende werkwijze aan als het niet direct in MS Word kan: 1) bewaar het bestand in MS Word als PDF en kies dan “best for electronic publishing and accessibility”; 2) doe het weghalen van de tekst in Adobe Acrobat Pro, volgens de door Adobe [gegeven richtlijnen](#).

## 8.1.3 Lak digitaal

Is een document alleen beschikbaar als PDF, en is dit een niet gescande PDF (zoals waarschijnlijk de overgrote meerderheid van opgevraagde documenten), doe het lakken dan digitaal. Dat kan bijvoorbeeld in Adobe Acrobat Pro, maar ook in andere laksoftware die ook meehelpt bij het vinden van te lakken stukjes (zoals persoonsgegevens). Dit [Woo-besluit](#) van het Ministerie van VWS dat Zylab gebruikt voor het lakken is een voorbeeld van een document waaruit alle mensleesbare tekst weggehaald is. Er is geen enkel machine leesbaar karakter overgebleven. Een onwenselijke situatie die ook niet voldoet aan artikel 2.4.3 van de Woo.

## 8.1.4 Repareer al openbaar gemaakte niet of slecht leesbare PDFs

Dit rapport laat zien hoeveel van de stukken totaal niet of slecht machine leesbaar zijn. Dat is voor een groot deel heel makkelijk op te lossen door de OCR opnieuw met kwalitatief hoogstaande software uit te voeren. Dat kan zelfs gratis, met de open source [Tesseract OCR software](#), die bijvoorbeeld ook is ingezet om de Woo stukken vrijgegeven na Covid 19 gerelateerde Woo-verzoeken goed machine leesbaar te maken<sup>2</sup>.

## 8.2 Produceer toegankelijke documenten volgens de WCAG norm.

Het is in de regel niet leuk om nadat een document eindelijk (en natuurlijk net op tijd) inhoudelijk af is nog eens te gaan zorgen dat het toegankelijk is voor mensen met een beperking en voldoet aan de WCAG normen. Maar het moet, en voor een groot deel kan het ook best makkelijk, mits meegenomen in de gebruikte werkwijze<sup>3</sup>.

### 8.2.1 Neem de tips en tricks mee in de gehanteerde werkwijze

Een hoop van de fouten die [pdfchecker.nl](#) aangeeft zijn heel goed te voorkomen door een (vaak éénmalige) verandering in de werkwijze. Zo is de taal waarin het document gesteld is en metadata als de auteur en de titel bijvoorbeeld al in te stellen in het sjabloon, en worden deze heel veel voorkomende fouten dus vermeden *by design*. Op [pdfchecker.nl](#) en de website [digitoegankelijk.nl](#) van het ministerie BZK staan goede en vaak makkelijk uitvoerbare tips. Wooverheid werkt samen met de provincie Zuid Holland en de BAR gemeentes aan dit soort sjablonen.

---

<sup>2</sup> Alle vrijgegeven stukken op de nieuwe [OpenVWS](#) site voor Corona gerelateerde Woo-dossiers zijn (opnieuw of voor het eerst) geOCRed met behulp van de Tesseract software. Ook worden standaard alle documenten op Woogel geOCRed met Tesseract. Dit zorgt dat heel veel stukken die niet vindbaar zijn op [open.overheid.nl](#), en zelfs niet via Google, wel via Woogel terug te vinden zijn.

<sup>3</sup> WCAG staat voor *Web Content Accessibility Guidelines* en is ontwikkeld door het World Wide Web Consortium (W3C). De [WCAG Wikipedia pagina](#) vormt een goed beginpunt om de WCAG te bestuderen. Het is voor de overheid al sinds 2008 verplicht om websites toegankelijk te maken. Per 1 juli 2018 is deze verplichting overgegaan in een wettelijke verplichting. Deze wettelijke verplichting is vastgelegd in het [Tijdelijk besluit digitale toegankelijkheid overheid](#).



---

## 8.2.2 Maak een HTML versie erbij

De overheid maakt haar documenten op tal van plekken openbaar, waaronder de website [officielebekendmakingen.nl](http://officielebekendmakingen.nl). Hier wordt vrijwel elk document, naast het vaak officiële PDF stuk, ook in een goed toegankelijke HTML versie getoond. Het is vele malen makkelijker om een HTML versie te maken die aan de WCAG eisen voldoet dan een PDF. Veel mensen met een beperking geven sowieso de voorkeur aan HTML, dus dit is een eenvoudige oplossing met hoog rendement. Vanuit MS Word kan men exporteren naar HTML, en er bestaan ook sterke transformatie programmas zoals [pandoc](#). Woogle en wooverheid bieden voor elk PDF document een *preview* versie in HTML aan, waarin geprobeerd wordt zoveel mogelijk van zowel de inhoud als de vorm (opmaak als kopjes, paragrafen, opsommingen, hyperlinks, nadruk door middel van schuin of vet-gedrukt, etc) te behouden in het omzetten van (een heel vaak ingescande) PDF naar HTML. Daarmee komen we op de volgende technische aanbeveling:

## 8.2.3 Repareer al bestaande PDFs zo veel mogelijk

De vorige tips gingen uit van een document wiens bronbestand in handen was. Als dat niet zo is, en het stuk is alleen beschikbaar als PDF, dan kan de PDF alleen nog gerepareerd worden. Dit kan prima. Wees hierbij realistisch en doel- en rendements-gericht. Als voor een fractie van de kosten al 80% van de WCAG fouten geheel automatisch opgelost kunnen worden, doe dat dan liever voor heel veel documenten dan slechts een beperkt aantal documenten handmatig en daardoor dus erg duur 100% goed (*WCAG compliant*) te krijgen. Woogle werkt volgens dit kostenefficiënte principe en repareert niet alleen de PDF maar maakt in één handeling door ook een zo goed mogelijk toegankelijke HTML versie

Heel veel van de stukken heeft helemaal niet zo'n lastige opmaak. Die kunnen voor een groot deel goed automatisch gerepareerd worden zodat ze prima kunnen worden voorgelezen.

## 8.3 Hergebruik en publiceer slim

### 8.3.1 Maak het eigen interne systeem publiek toegankelijk

Niet voor niets zijn informatiehuishouding en openbaarmaking samengebracht in één adviescollege. Een slim bestuursorgaan maakt gebruik van hun interacties. Zo kunnen gemeenten die het ibabs vergadersysteem gebruiken dit op allerlei manieren “openzetten voor iedereen” (via de `<bestuursorgaan>.bestuurlijkeinformatie.nl` websites)<sup>4</sup>. We kunnen verwachten dat iedereen die professioneel bezig is met de Woo-documenten de makkelijk toegankelijke open versie gaat gebruiken. Er is dus direct een prikkel om dat portaal zo goed en zo open mogelijk te krijgen. Als gratis bij-effect wordt de burger dan (bijna) net zo goed bedient als de ambtenaren.

### 8.3.2 Lokaal is een *view* van een centraal systeem

Lokale bestuursorganen als gemeentes en waterschappen zijn vaak trots op hun website en willen die er speciaal en uniek uit laten zien. Dit zien we terug in al de verschillende versies van het [bestuurlijkeinformatie.nl](http://bestuurlijkeinformatie.nl) systeem van ibabs, en vergelijkbaar voor Notubiz en de Openwebconcept gemeentes. De kracht achter die sites is dat ze eigenlijk een gelocaliseerde *view* zijn van een enorm database systeem dat ook de gegevens van tientallen andere bestuursorganen bevat. Met deze opzet worden twee vliegen in één klap geslagen. Ten eerste blijft de centrale database actueel, volledig en wordt gecontroleerd op fouten en omissies. Ten tweede is dit enorm kostenefficiënt: het meeste verkeer gaat lokaal, maar er is maar één infrastructuur nodig. Daarnaast kan dit draaien op een *commons* gedachte: verbeteringen doorgevoerd voor en betaald door één of een paar organen komen ook meteen gratis beschikbaar voor de anderen.

De [Woogle](#) database is al volledig klaar voor deze opzet: alle interactie kan gelocaliseerd worden tot een selectie van bestuursorganen en kan lopen via APIs. In feite is het maken van een op één bestuursorgaan toegesneden en totaal zelf

---

<sup>4</sup> De gemeente Almelo heeft precies deze “gooi alles maar open” gedachte ver doorgevoerd op <http://almelo.bestuurlijkeinformatie.nl/>. Eén van de motivaties was dat men hoopte dat het aantal Woo-verzoeken zo zou afnemen of dat men de verzoeker snel en makkelijk naar de (voor iedereen toegankelijke) plek waar de stukken staan kon verwijzen.

---

vormgegeven Woogle een kwestie van het samenvoegen van de gebruikers-interacties via de API met de huisstijl van het bestuursorgaan. Op het moment van schrijven is Woogle dit aan het uittesten samen met de gemeente Nijmegen. Zij gebruikt nu nog de Woogle huisstijl voor haar [pagina met Woo besluiten](#).

## HOOFDSTUK 9

---

Verder lezen

---

Alle voor dit onderzoek geanalyseerde stukken zijn terug te vinden via de zoekmachine voor Woo documenten [Woogle](#). Woogle indexeert onder andere alle Woo-dossiers na een Woo-verzoek gepubliceerd op [open.overheid.nl](#) en die van de OpenPub gemeenten.

Nieuws over software dat in het kader van dit onderzoek aan het Instituut voor Informatica aan de Universiteit van Amsterdam wordt ontwikkeld is het makkelijkst te volgen via de [wooverheid.nl](#) blog. Er is ook een [aparte pagina met peer-reviewed publicaties](#) over dit onderwerp. Links naar demonstratie versies van de software zijn te vinden onder de *software* tab.

---

Dit onderzoek is uitgevoerd in opdracht van het Adviescollege Openbaarheid en Informatiehuishouding (ACOI) in de herfst van 2023 onder leiding van Maarten Marx, universitair docent aan het Instituut voor Informatica van de Universiteit van Amsterdam.

De volgende mensen hebben meegewerkt aan het onderzoek: Maik Larooij heeft alle opgehaalde bestanden geanalyseerd en geïndexeerd in de Woogel zoekmachine. Maik hield de Woogel database draaiende en zorgde dat ontwikkelde analyse software ook kon draaien op de miljoenen pagina's aan Woo-bestanden die ondertussen in Woogel zitten. Ramon Duursma heeft gewerkt aan het verzamelen van de bestanden en het onderzoek naar de waterschappen. Nazli Aydin heeft de analyse van de zbo's gedaan. Gregory Slager werkte aan een systeem van sjablonen voor Woo-besluiten en beschikkingen. Lars Nelissen, Lars Reddering en Gensi Eugenia werkten aan het repareren van PDFs en het maken van toegankelijke HTML versies. Roy Duineveld en Lennaert Feijtes lieten zien dat \*text redaction\* (het weglakken van bepaalde stukjes tekst) goed en veilig direct in een digitaal PDF bestand kan. Voor het PDF onderzoek was de hulp van Maxwell Bland (voor het PDF redactie) en Andy Robinson (voor het opnieuw opbouwen van een PDF) heel nuttig en waardevol.

Het ACOI heeft het onderzoek financieel mogelijk gemaakt. Anne-Maartje Douqué en Jelle Clemens van het ACOI hebben enorm geholpen met het bewaren van focus en hebben waardevol commentaar gegeven op eerdere versies van dit rapport. Meer indirecte hulp hebben we gekregen van Jaap Kamps, Ruben van Heusden (beide UvA), Laura Ootes (eScience Center), Guido Enthoven (IMI), Marianne de Nooij (Prov ZH) en Joep Meidertsma (Argu).