

**Hoofdvraag:** Op welke manier kan de implementatie voor een monitorings overzicht van de WOO-verzoeken in Nederland geïmplementeerd worden?

**Inleiding:** Beschrijving dat dit in Nederland nog niet aanwezig is maar wel in de UK.

**Mogelijke deelvraag 1:** Welke data uit het overzicht in de UK is ook in Nederland beschikbaar: Is dit te extraheren? Is er een overzicht te vinden (tot 2019 wel)? Is er al data aanwezig in de Woogles data?

**Andere mogelijke deelvragen:**

- Op welke manier is dit overzicht geïmplementeerd?
- Kan dit overzicht van toegevoegde waarde zijn op het "aanhaak platform?"
- Welke toegevoegde waarden heeft een overzicht zoals degene in de UK.

**Analyse open overheid:** Het lijkt erop alsof alle "besluiten van WOO-verzoeken hier terug zijn te vinden". Het aantal WOO besluiten lijkt best wel te kloppen met de data die verzameld is in "<https://zoek.officielebekendmakingen.nl/kst-28362-21.html>" indien gedupliceerd worden verwijderd. Bijvoorbeeld in 2018 waren er blijkbaar 42 WOB verzoeken vanuit de landbouw, natuur en voedselkwaliteit sector. Als ik dit onderzoek vind ik er meer dan 200, maar het ziet er naar uit dat er heel veel duplicaten zijn van 1 enkel besluit waren namelijk vier verschillende instanties te vinden op de eerste pagina. Echter 1x was de titel wel iets anders (rekening mee houden).

Voor het vergaren van de data omtrent Wob-Woo verzoeken analyseer ik de documenten die beschikbaar zijn op de website Open.Overheid.nl. Specifiek de besluiten op de verzoeken. Alle Wob/Woo verzoeken staan netjes genoteerd als "besluit op Woo-verzoek" met hier achter de titel van het verzoek en als er extra omschrijving nodig is staat dit achter een dash (-). Vandaar dat de titels van Woo-verzoeken op te delen zijn in twee delen. De data die wordt verzameld komt uit de omschrijving. Bij het analyseren van verschillende pdf's op open.overheid werd namelijk al snel duidelijk dat er vrijwel geen overeenkomsten zijn tussen alle PDF's. Sommige artikelen hebben netjes een mooi colofon waar makkelijk data uit gehaald zou kunnen worden indien elk artikel dit zou toepassen. Echter zijn er genoeg artikelen waar helemaal geen data te vinden is. Dus is data-extractie van de artikelen zelf vrijwel nutteloos. Het werd al snel duidelijk dat de data uit de omschrijvingen wel bruikbaar was want deze was overal een tabel met 5 of 6 dezelfde gegevens. Waarvan het documentsoort natuurlijk altijd hetzelfde is dus dit is totaal niet interessant. Thema is heel vaak onbekend, dus het is nog maar de vraag of hier iets mee mogelijk is. Onderwerp (de 6e gegeven) ontbreekt vaak helemaal uit de omschrijving als deze ontbreekt wordt er NaN verzameld. Het hierdoor is echter de vraag of deze kolom ook bruikbaar zal zijn. In de Thema en Onderwerp kolommen kan soms ook een lijst door middel van bullet points staan. Hier wordt in de analyse een python lijst van gemaakt, maar het is dus de vraag of dit wel bruikbaar zal zijn.

Dat houdt eigenlijk maar drie bruikbare soorten data over. Namelijk de verantwoordelijke, hier staat het ministerie dat verantwoordelijk is voor de behandeling van het Woo-verzoek en

de publicatie en document datum. Dit kan bijvoorbeeld gebruikt worden om te achterhalen in welk kwartaal het besluit is gemaakt. Verder worden uiteraard ook de titels van de besluiten verzameld en opgesplitst in de twee delen die eerder al zijn beschreven.

### Hoe heb ik deze informatie verzameld?

Door middel van een python script met Selenium. De url die als input wordt gebruikt is de URL voor alle Wob/Woo verzoeken vanaf 2019 per jaar. Ik heb gekozen voor per jaar zodat ik een csv file kan maken met de data per jaar en zodat ik een checkpoint heb per jaar mocht er toch iets mis gaan in de data van 2020 dan heb ik in elk geval de data van 2019 al verzameld.

### Problemen met de data + oplossingen?:

Duplicates: Van besluiten kunnen meerdere publicaties verschijnen op dit platform. Echter omdat de titel dus gesplitst wordt door middel van de dash zal de naam van het verzoek dus voor de dash staan en de naam van de specifieke publicatie achter de dash. Op deze manier kunnen duplicates dus achterhaald worden op basis van naam voor de dash. Dit maakt dan ook deel uit van de beantwoording van de eerste opwarmklus bij het downloaden van de eerste 50 resultaten. Als er een door spaties omgeven dash aanwezig is weet je dus dat het document tot een onderdeel is van een ander document. Zoals net ook al vermeld was de lijst met bulletpoints een probleem en deze werd omgezet naar python list. Dit had wellicht met selenium gedaan kunnen worden, maar ik heb gekozen om zinnen te extraheren met regex. Werkt ook in elk geval.

**Na het verzamelen van 2019:** Na het verzamelen van de data van 2019 werd het duidelijk dat nog lang niet alle data gestructureerd is, en dat ook het begin van de namen van artikelen nog lang niet altijd hetzelfde is. Veruit in de meeste gevallen wordt hier geopend met het woord besluit of een synoniem als beslissing, besluiten of deelbesluiten, of gewoon letterlijk met Wob-verzoek. Maar ook het beginwoord bijlage(n) komt vaak terug. Verder zijn er nog een aantal vaker voorkomende beginwoorden.

```
[('Besluit', 2053), ('Wob-verzoek', 124), ('Beslissing', 78), ('Bijlagen', 69), ('Bijlage', 38), ('Wob-besluit', 33), ('Deelbesluit', 28), ('Tweede', 26), ('Eerste', 22), ('1e', 18), ('Herzien', 16), ('2e', 16), ('Besluiten', 13), ('3e', 12), ('Bezwaar', 11), ('Wijzigingsbesluit', 9), ('Derde', 9), ('Documenten', 7), ('Hernieuwde', 6), ('Eeste', 4), ('Herziene', 4), ('Wob-verzoeken', 3), ('Twee', 3), ('Herstelbesluit', 1), ('Wob', 1), ('Inventarislijst', 1), ('JenV', 1), ('Nader', 1), ('Uitvoering', 1), ('Te', 1)]
```

Verdere analyse laat zien dat er 1478 unieke titels voorkomen tussen de documenten in 2019. Bij het eerste deel voor de dash (de "verzoeknamen") zijn dit er 993. En aangezien er nog minimale verwerking van data is gedaan zou het best kunnen dat er nog wel flink aantal extra documente geschraapt zou kunnen worden.

Ik heb verder ook het idee dat Wooverheid/Woogole ook met deze data werkt, maar alsnog vind ik het wel leuk om hier zelf een beetje mee aan de slag te gaan. En is dus zeker een leuke opwarmklus.

Dit artikel geeft eigenlijk een vergelijkbaar overzicht als het britse form dus wellicht is hier te lezen waar deze data te verkrijgen is. Echter kan ik op [rijksoverheid.nl](http://rijksoverheid.nl) waar ze deze data

hebben verkregen geen data vinden over deadlines van verzoeken, dus geen idee waar tegenwoordig zulk soort data verkrijgbaar is.

<https://openstate.eu/wp-content/uploads/sites/14/2022/01/Ondraaglijk-traag-280122-def.pdf>

**Analyse Britse Data:** Hierin probeer ik te achterhalen welke data uit de Britse data ook in Nederland terug te vinden is. Analyseren van de grafieken op de site want er staan wel heel veel verschillende parameters in de csv file. Meer analyse volgt nog.