

Het is niet haalbaar om een volledig overzicht te maken van alle gegevens die de overheid precies allemaal heeft

waarmaken? En tegen welke kosten - niet alleen financieel maar ook qua prioriteit: wat doe je dan niet?"

Een representatie

Meer realisme in ambities kan leiden tot andere keuzes. Informatiehuishouding is niet altijd het meest tot de verbeelding sprekende onderwerp, maar in het verbeteren daarvan

ligt de basis van veel andere ambities. Klievink geeft een voorbeeld. "Neem de ambitie om de burger inzicht te geven in de gegevens die overheid van burgers heeft. Het is niet haalbaar om een volledig overzicht te maken van alle gegevens die de overheid precies allemaal heeft, met alle kopieën ervan. En in welke processen die gegevens precies worden gebruikt binnen al die overheidsorganisaties. Wanneer we burgers inzicht willen geven met een portaal waar die gegevens staan, is dat een representatie, maar niet de representatie. We hebben daar immers niet de informatiehuishouding voor." Dat betekent overigens niet dat je dit niet moeten doen, benadrukt hij. "Je kunt beginnen met de gegevens die je wel hebt. Maar de vraag is wel: is de burger daarmee het beste gediend? Misschien kun je een stap eerder zetten, in de kwaliteit van je informatiehuishouding. Door data die je niet mag of niet zou moeten willen hebben gewoon weg te gooien?"

Gematigdere ambitie

Klievink denkt dat de maatschappelijke coalitie kan bijdragen aan meer realisme. Dat vraagt wel dat de overheid op de een of andere manier meedoet. "Met het maatschappelijk middenveld, de wetenschap, vertegenwoordiging van cliënten en burgergroepen, en vertegenwoordigers van de overheid samen een plan maken. Anders wordt het weer een vraag van buiten aan de overheid. Een gematigdere ambitie en een realistisch plan gaan veel meer bijdragen dan een hele scherpe probleemanalyse - hoe waar die ook kan zijn - als deze alleen van 'buiten' de overheid is."

Hollen en stilstaan

'Hollen én stilstaan: hoe data en digitalisering de overheid veranderen', is de titel van Klievink's oratie. Het 'en' in de titel is belangrijk, benadrukt hij. De overheid moet zich voor datagedreven werken zowel revolutionair (hollen) als evolutionair (stilstaan) verbeteren. En dat bovendien tegelijkertijd. Hij doet onderzoek naar hoe dat gaat bij AI-innovaties. "We zien dat organisaties op een paar manieren innovatie vormgeven. Enigszins gechargeerd: óf je maakt er iemand van - bijvoorbeeld een innovatieafdeling - óf je stimuleert innovatie binnen bestaande afdelingen en zelfs bij individuen." Aparte innovatieafdelingen blijken beter

in staat om grootschalige, geavanceerdere innovaties te bedenken en maken. Toch levert incrementele verbetering in alles wat je doet uiteindelijk meer verandering op. Omdat adoptie in het primaire proces al is ingebakken in die innovatiebenadering. Maar er zit een grens aan hoeveel je daarmee kunt bereiken. De hamvraag: hoe krijg je allebei gedaan? "Daar kunnen pilots en dergelijke bij helpen. Toch zijn het uiteindelijk gewoon keuzes van organisatie en politiek. Is er budget en capaciteit? En we moeten steeds de vraag stellen of we bepaalde innovaties willen, en niet alleen hoe we die gerealiseerd krijgen."

34

iBestuur

Woo-verzoeken: van hoofdpijndossier naar waarde voor de samenleving

Hergebruik van overheidsinformatie zo makkelijk mogelijk maken; dat is waar Maarten Marx, universitair docent aan de Universiteit van Amsterdam, binnen het NWO Access programma aan werkt. Met een eenvoudige standaard, gebaseerd op FAIR data-principes, kunnen Woo-verzoeken veel beter toegankelijk en herbruikbaar worden. Het wenkend perspectief: Woogie, een zoekmachine en verwijzindex voor Woo-dossiers.

seerde metadata over het dossier en de afzonderlijk vrijgegeven documenten. Van de vrijgegeven documenten is 31 procent van de pagina's niet door de computer te lezen en daardoor niet doorzoekbaar.

Software voor Woo-ambtenaar

Voor informatiezoekers is het vanwege deze manco's een lastige zoektocht om relevante gegevens uit de vrijgegeven documenten te halen. Marx schetst waar dat probleem ontstaat. "In de huidige situatie is het veelal zo dat ambtenaren hun Woo-dossiers wel FAIR op hun eigen schijf hebben staan, maar dat er in de laatste publicatiestap iets misgaat", aldus Marx. Om de dossiers ook FAIR te kunnen publiceren moeten de documenten digitaal in (zip)mapjes staan, liefst met een uniforme

In de Wet open overheid (Woo) is het recht geregeld om informatie te vragen over wat de overheid doet. Het hogere doel is het versterken van het vertrouwen in de democratie en het draagvlak voor beleidsmaatregelen. De wet is een belangrijk instrument om informatie op tafel te krijgen voor bijvoorbeeld journalisten en Tweede Kamerleden. Nieuw in de Woo ten opzichte van voorganger Wob (Wet openbaarheid van bestuur) is dat de overheid zelf de plicht heeft om informatie actief openbaar te maken.

"Ik wil eraan bijdragen dat hergebruik van informatie mogelijk en makkelijk wordt. Door te zorgen dat informatie vindbaar, toegankelijk, makkelijk te verwerken, en herbruikbaar is", aldus Marx, lid van de maatschappelijke coalitie Over Informatie Gesproken. Dat verklaart ook zijn aandacht voor Woo-dossiers, die momenteel niet altijd aan de genoemde criteria voldoen. Het team van Marx onderzocht zo'n 1 miljoen pagina's aan door de Nederlandse overheden (gemeenten, provincies, ministeries) gepubliceerde Woo-dossiers. Niet op inhoud, maar op de vier FAIR data-principes: Findable, Accessible, Interoperable, Reusable. De uitkomst is dat de dossiers vaak tekortschieten op deze punten. Woo-dossiers worden nu veelal gepubliceerd zonder computerleesbare en gestandaardiseerde

Woo-dossiers zijn veelal niet door de computer te lezen en daardoor niet doorzoekbaar



Nummer 45, januari 2023

35



inventarislijst. Om dat mogelijk te maken ontwikkelde Marx een open source softwarepakket dat Woo-coördinatoren kunnen gebruiken en dat inzetbaar is voor elke bestuurslaag die Woo-dossiers produceert. "Het werkt heel makkelijk: je voert de losse vrijgegeven stukken in, samen met de inventarislijst en wat basisgegevens over het dossier die ook al in het besluit staan. Wij voorzien alle documenten van unieke *digital object identifiers*, en slaan alles op – lokaal of op een server, afhankelijk van de wens van de Woo-aanbieder – en kunnen zo gebruikers snel en makkelijk naar elk Woo-document verwijzen."

Het enige extra werk is het maken van een overzichtje, vergelijkbaar met de infobox die op vrijwel elke Wikipedia-pagina rechtsboven staat. Een andere mogelijkheid van het aanbieden van de Woo-dossiers is via het zaakstelsel van het bestuursorgaan, heel vergelijkbaar met hoe nu al volkomen automatisch de stukken uit 277 gemeenten en provincies op de website zoek.openraadsinformatie.nl komen. Als iedereen meedoet

zijn straks alle Woo-dossiers vanaf één plek te doorzoeken en te vinden.

Wooglee

De R van reusable was de belangrijkste drijfveer voor Marx en zijn collega Jaap Kamps van UvA Archiefwetenschappen om aan de slag te gaan met Woo-dossiers. "Het initiele doel was alle Woo-verzoeken samenbrengen en beschikbaar maken voor wetenschappelijk onderzoek. Zodat je kunt zoeken en terugvinden welk materiaal er op tafel is gekomen. E-mails, appjes, beschikkingen, et cetera. Maar voordat je zo'n onderzoek kunt doen, moet je eerst het materiaal verzamelen. Zo kwamen wij erachter hoe lastig dat is. Een van de uitdagingen is dat vrijwel alle bestuursorganen de gekke gewoonte hebben om alle vrijgegeven documenten in een Woo-dossier aan elkaar te plakken tot één, vaak enorm PDF-bestand. De grenzen tussen de afzonderlijke bestanden zijn niet eenduidig aangegeven.



Maarten Marx, UvA
BESLUIT NEDERLAND

Dit is voor gebruikers van een zoekmachine lastig. Het is nogal een verschil als de zoekmachine je na een zoekvraag naar een mailtje verwijst of naar een enorm bestand met honderden mails, vergaderstukken, kaarten, en wat al meer. Dan mag je het zelf verder uitzoeken in een PDF-lezer. Als je dan ook nog niet eens met Control F in dat bestand kunt zoeken (wat dus voor bijna een derde van de pagina's het geval is) wordt iets terugvinden een best frustrerende bezigheid. De tekst leesbaar maken kunnen computers nu al best goed. Het weer opknippen is veel lastiger." Afstudeerders van Marx en Kamps probeerden dat met neurale *deep learning* AI-technieken. "Voor een mens is het redelijk gemakkelijk om de grens te vinden van een document, maar voor een computer is dat lastiger. We leren dat aan de computer door duizenden dossiers die wel opgeknipt zijn als voorbeelden te geven. Dit werkt aardig goed, maar zeker nog niet foutloos."

Als *proof of concept*, en om gemeenten en provincies te overtuigen zich aan te sluiten, bouwde de groep van Marx en Kamps een zoekmachine en verwijsindex voor Woo-dossiers: Wooglee. Dat bevat momenteel zo'n 7.000 dossiers en meer dan een miljoen pagina's van allerlei soorten aanbieders. Uithreiding naar de zestien andere in de Woo genoemde informatiecategorieën staat op stapel. Elk document heeft hier een unieke permanente naam, gebaseerd op het bestuursorgaan dat het publiceerde, de soort informatiecategorie en het moment van publiceren. Een voorbeeld van een keurig machineleesbaar dossier vrijgegeven door de gemeente Waalwijk staat op data.wooverheid.nl/gm0867/publications/2022/nl-gm0867.2022.21.4/nl-gm0867.2022.21.4-inhoud.zip.

Informatie als grondstof

Marx wijst graag op het enorme potentieel aan opbrengsten wanneer Woo-dossiers voor hergebruik beschikbaar zijn. "Dit kan bijdragen aan een betere informatievoorziening van de overheid. Ik wil ambtenaren ervan overtuigen dat ze zeer waardevolle informatie in handen hebben en van het

Informatie is als de liefde, je moet er iets mee doen, want anders verliest het z'n waarde

belang dat anderen die informatie kunnen hergebruiken. We groeien als samenleving als de overheid informatie beschikbaar stelt." Marx geeft als voorbeeld Google Translate; dat had nooit bestaan zonder netjes vrijgegeven overheidsinformatie. "Machine learning werkt op basis van voorbeelden. Voor Google Translate waren dat de notulen en wetten van het Europees parlement, simultaan vertaald in alle EU-talen; miljoenen voorbeelden van zinnen in verschillende talen die precies hetzelfde betekenen. Wie had destijds deze geweldige toepassing kunnen voorzien?"

Marx citeert tot slot een uitspraak van zijn promotor: *Informatie is de enige grondstof die groeit in het gebruik*. Waarom dat zo'n goed motto is? "Informatie is als de liefde. Je moet er iets mee doen, want anders verliest het z'n waarde. Anders dan kolen of gas kun je informatie niet opgebruiken; door informatie te combineren ontstaat nieuwe informatie."

API

Eind 2022 komt er een API beschikbaar waarmee zaakstelselbeheerders als Visma, Babs en Notuliz Woo-dossiers direct en automatisch in de Wooglee zoekmachine kunnen plaatsen. Daarmee vervult Wooglee in zekere zin een vergelijkbare rol die de Woo toedacht aan het Platform Open Overheidsinformatie PLOOI. Het onderzoek van Marx en Kamps over de Wet open overheid is te volgen op hun blog wooverheid.nl.

Eind 2022 werd overigens bekend dat op aanbeveling van het Adviescollege ICT-toetsing de stekker uit de huidige versie van PLOOI wordt getrokken. Belangrijkste reden: verkeerde keuzes in de ICT-ontwikkeling.