



De opkomst van maatschappelijke moderatie

29 Januari 2026 | Marleen Stikker, Danny Lämmerhirt, Isadora Dullaert, Mercedes Leipoldt

waag  futurelab

Samenvatting

De overheid zet zich in om illegale content op het internet tegen te gaan. Het Project Online Content Moderatie (PrOCOM) geeft hier een praktische invulling aan. In dat kader verkent Waag Futurelab (hierna: Waag) in dit onderzoeksrapport welke rollen en verantwoordelijkheden maatschappelijke organisaties kunnen innemen in het modereren van sociale media, specifiek in de context van het federatieve internet ofwel het 'Fediverse'. Het rapport introduceert de term "maatschappelijke moderatie". Dit definiëren wij als 'de zeggenschap van maatschappelijke organisaties over het beleid (regelgeving) en de uitvoer (moderatiewerk) van online content moderatie op sociale media.'

Onderzoeksdoelen en aanpak

Voor dit onderzoek zijn twee focusgroepen georganiseerd en een aantal interviews afgenomen. De eerste focusgroep bracht vertegenwoordigers van maatschappelijke organisaties bijeen. Met hen is in beeld gebracht welke rol maatschappelijke organisaties voor zichzelf zien en wat de voorwaarden zijn om een rol te kunnen vervullen. De tweede focusgroep bestond uit moderators die al op het Fediverse actief zijn, met name voor Mastodon. Met hen is in kaart gebracht tegen welke vraagstukken zij aanlopen.

Resultaten

Uit dit onderzoek blijkt dat maatschappelijke organisaties moderatie een belangrijke maatschappelijke taak vinden, maar dat er op dit moment onvoldoende middelen en ondersteuning bestaan om die verantwoordelijkheid te kunnen of willen nemen. Maatschappelijke organisaties zien hierbij nog niet altijd een actieve rol voor zichzelf. Wel geven zij aan de mogelijkheden van het Fediverse te willen verkennen en te willen onderzoeken hoe zij moderatietaken zouden kunnen uitvoeren in relatie tot hun publieke functie. Enkele organisaties hosten al een Fediverse-instantie en willen dit uitbouwen mits de voorwaarden beter zijn. Een aantal ziet vooral een rol als dienst aanbieder, bijvoorbeeld door het aanbieden van kennis en trainingen of als adviseur in het meedenken over regelgeving rond moderatie. Uit de gesprekken met moderators die al actief zijn blijkt dat men behoefte heeft aan kennisuitwisseling, juridische ondersteuning en betere moderatietools.

Aanbevelingen

Gezien het maatschappelijk belang is het raadzaam om maatschappelijke moderatie te ondersteunen. Op basis van dit onderzoek doen we de volgende aanbevelingen:

- Ontwikkel een 'Programma Maatschappelijke Moderatie' voor de ondersteuning van maatschappelijke organisaties bij moderatie op het gebied van juridische en operationele vraagstukken, onderlinge samenwerking, trainingen en professionalisering en publieksprogrammering.
- Creëer ruimte voor experiment in 'living labs' waar maatschappelijke organisaties ervaring kunnen opdoen met het Fediverse en ondersteun technologische innovatie op het gebied van applicaties, moderatie tools en interface design.
- Geef als overheid het goede voorbeeld door zelf federatieve sociale media te gebruiken.

1. Inleiding

Online contentmoderatie is een belangrijk maatschappelijk vraagstuk. Goede moderatie gaat de verspreiding van ongewenste en illegale content op online platformen tegen en vraagt om een structurele aanpak.

Moderatie is niet neutraal. Het is gebaseerd op een subjectieve interpretatie van waarden zoals vrijheid van meningsuiting, veiligheid of feitelijkheid. De meest gebruikte sociale media platforms zijn in handen van Amerikaanse en Chinese big tech bedrijven. Deze bedrijven bepalen de regels voor moderatie op hun platforms. De zorgen over de wijze waarop big tech de verantwoordelijkheid voor online contentmoderatie nemen, spitst zich toe op het toepassen van verslavende en misleidende algoritmen, het verdienmodel dat privacy schaadt en op politieke manipulatie en beïnvloeding. Europa probeert met wetgeving invloed uit te oefenen. Er blijkt echter dat big tech-bedrijven deze wetten niet voldoende implementeren. Integendeel. Tech-bedrijven investeren steeds minder in hun moderatie infrastructuur¹ en spreken zich expliciet uit tegen Europese regelgeving.

De big tech platforms zijn weliswaar dominant, maar alternatieve sociale netwerken zijn in opkomst. Veel van deze alternatieven zijn federatief. Dat houdt in dat deze netwerken decentraal georganiseerd zijn en met elkaar kunnen 'praten'. Maatschappelijke organisaties kunnen instanties binnen deze netwerken opzetten en zo deel van het netwerk hosten. Gebruikers kunnen de sociale netwerken en host kiezen die bij hun behoeftes passen, en tegelijkertijd in contact blijven met gebruikers van andere netwerken. Met de opkomst van federatieve sociale media zoals Mastodon en Pixelfed zien we voorbeelden waarbij de verantwoordelijkheid voor het moderatiebeleid en de uitvoer ervan bij een diversiteit aan maatschappelijke organisaties komt te liggen. Dit biedt kansen voor het terugwinnen van strategische autonomie.

In dit rapport verkent Waag Futurelab in opdracht van het Ministerie van Justitie en Veiligheid wat het federatieve internet of het 'Fediverse' kan betekenen voor online moderatie en welke rollen en verantwoordelijkheden maatschappelijke organisaties hierbij kunnen. **We definiëren maatschappelijke moderatie als de zeggenschap van maatschappelijke organisaties over het beleid (regelgeving) en de uitvoer (moderatiewerk) van online content moderatie.** Onder maatschappelijke organisaties verstaan we stichtingen, verenigingen of coöperaties met een maatschappelijke, non-profit doelstelling. Daartoe behoren o.a. bibliotheken, culturele instellingen, NGO's, (lokale) publieke omroepen en vrijwilligersorganisaties.

¹ Grote sociale mediabedrijven hebben in de afgelopen jaren duidelijk minder in zogenaamde "Trust and Safety" teams geïnvesteerd, wat ervoor zorgt dat ze ook minder moderatiewerk uitvoeren. Zie: Rachel Moran, Elizabeth; Schafer, Joseph; Bayar, Mert; and Starbird, Kate (2025): The End of Trust and Safety?: Examining the Future of Content Moderation and Upheavals in Professional Online Safety Efforts. <https://dl.acm.org/doi/10.1145/3706598.3713662>

Er zijn verschillende redenen waarom maatschappelijke organisaties een belangrijke rol kunnen spelen. Ten eerste dragen ze een institutionele verantwoordelijkheid voor het publieke maatschappelijke domein. Ze spelen een rol in het garanderen van de pluriformiteit van media, de vrije toegang tot informatie, de productie en distributie van content en het organiseren van het publieke debat. Sommige organisaties, zoals bibliotheken en omroepen, hebben wettelijke taken, bijvoorbeeld op het gebied van digitale inclusie en participatie in de informatiesamenleving.²

Een tweede reden is dat de mogelijkheden voor maatschappelijke moderatie enorm zijn toegenomen door de ontwikkeling van het Fediverse. Het is nu mogelijk om zelf een instantie te starten en medeverantwoordelijkheid te dragen. De architectuur van federatieve sociale media maakt het mogelijk dat beslissingen over moderatie en sociale interacties in de handen van meerdere partijen liggen. Maatschappelijke organisaties hebben op deze platforms zeggenschap over zowel het beleid als de uitvoer.

Om meer inzicht te krijgen in de motivatie en mogelijke rollen van maatschappelijke organisaties binnen het Fediverse, zijn de volgende vragen voorgelegd aan moderators en vertegenwoordigers van maatschappelijke organisaties:

- Wat motiveert maatschappelijke organisaties om federatieve sociale media te gebruiken?
- Wat voor rollen zien ze voor zichzelf met betrekking tot moderatie?
- Wat is ervoor nodig om federatieve sociale media te modereren?

Het rapport begint met een kort overzicht van de onderdelen van online moderatie en hoe deze georganiseerd zijn op federatieve sociale media. Daarna omschrijven we de methode en resultaten van het onderzoek en komen we tot een aantal aanbevelingen.

² In de Openbare Bibliotheken convenant staat een focus op drie actuele maatschappelijke opgaven waarbij opgave 2 zich richt op Participatie in de (informatie)samenleving – Digitale inclusie, democratie en burgerschap. De groei van digitalisering in onze samenleving zorgt ervoor dat we de verantwoordelijkheden van bibliotheken breder kunnen interpreteren. Zie: Ministerie van Onderwijs, Cultuur en Wetenschap, Vereniging van Nederlandse Gemeenten & Koninklijke Bibliotheek. (2023). *Vernieuwing Bibliotheekconvenant 2024–2027*. <https://www.kb.nl/organisatie/beleid/bibliotheekconvenant>

2. Moderatie: een korte geschiedenis

Het vraagstuk van moderatie is zo oud als het internet. Eind jaren zeventig ontstond Usenet, een verzameling van tienduizenden nieuwsgroepen. Ieder van die nieuwsgroepen werd gemodereerd door een of meerdere personen. Het was vanzelfsprekend dat digitale gemeenschappen zelf de moderatieregels bepaalden, zoals over het verwijderen van schadelijke content of het bepalen wie wel of niet deel kon nemen.³ Bij de opening van De Digitale Stad in januari 1994, een sociaal media platform *avant la lettre*, werden meerdere nieuwsgroepen gelanceerd en moderators aangesteld.

Pas met de centralisatie en exploitatie van sociale media platforms in het begin van deze eeuw kwam de organisatie van moderatie - en daardoor sociale interacties - steeds meer bij platformen en hun eigenaars te liggen. Volgens de Amerikaanse mediawetenschapper Tarleton Gillespie zorgde dit ervoor dat deze platformen 'the custodians of the web' werden: zij bepalen welke gesprekken en welk gedrag wel en niet aanvaardbaar is op het web⁴. Contentmoderatie is daarom, volgens Gillespie, de voornaamste dienst die platformen aanbieden. In de praktijk betekent dit dat commerciële bedrijven (en dus een kleine groep eigenaren van deze platformen) de moderatieregels bepalen, en zo een grote invloed uitoefenen op de online publieke sfeer. Daarnaast heeft deze situatie ook impact op de mensen die de moderatie moeten uitvoeren: contentmoderatie is niet eenvoudig; blootstelling aan schadelijke content kan leiden tot mentale gezondheidsproblemen als angst, slaapproblemen of zelfs trauma.⁵ Vaak worden deze taken uitbesteed aan landen waar de lonen laag liggen. Veelal zijn het contentmoderatoren in het mondiale Zuiden, die de schadelijke content uit het mondiale Noorden moeten 'opruimen'.⁶ Daarnaast hebben veel platforms niet voldoende moderators met kennis voor cultuur of taal om efficiënt moderatiekeuzes te maken.⁷

Europees beleid over onlinemoderatie voor 'Very Large Online Platforms' (de Digital Services Act, of DSA) zorgt ervoor dat platformeigenaars en overheden een centrale rol in moderatie spelen om beleidsregels voor platformen te bepalen.⁸ Voor maatschappelijke organisaties biedt deze wetgeving specifieke mogelijkheden om invloed op moderatie uit te oefenen. Maatschappelijke actoren zoals mensenrechtenorganisaties, academici en journalisten spelen bijvoorbeeld een rol als raadsleden voor veiligheid (Trust & Safety Councils) om moderatieregels te bepalen, of kunnen schadelijke content melden als "trusted flaggers". Uit onderzoek⁹ blijkt echter dat de invloed van maatschappelijke actoren op moderatie vaak klein blijft. Platformeigenaren en overheden beslissen bijvoorbeeld welke maatschappelijke organisaties lid van een veiligheidsraad worden en wat er met schadelijke content gebeurt.¹⁰ Hetzelfde geldt voor initiatieven die eindgebruikers meer invloed op moderatie willen geven zoals flagging mechanismen of

³ Usenet Historical Collection (n.d.). Zie: <https://archive.org/details/usenethistorical>.

⁴ Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation and hidden decisions that shape social media*.

⁵ Haime, Z., Kennedy, L. en Biddle, L. (2025) 'Experiences of Moderation, Moderators, and Moderating by Online Users Who Engage with Self-Harm and Suicide Content', *Digital Society* 4, 8. Zie: <https://link.springer.com/article/10.1007/s44206-025-00166-x#citeas>

⁶ Ahmad, Sana, and Martin Krzywdzinski. "Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains." In *The MIT Press eBooks*, 77–96, 2022. <https://doi.org/10.7551/mitpress/13835.003.0008>.

⁷ Amnesty International (2025). A Thousand Cuts. Technology-facilitated gender-based violence against Poland's LGBTI community on X. Zie: <https://www.amnesty.org/en/documents/eur37/0098/2025/en/>

⁸ Project Online Content Moderatie (PrOCOM) (2022). Zie: https://hetccv.nl/app/uploads/2023/08/20221108_Projectplan_PrOCOM.pdf

⁹ Badouard, R. en Bellon, A. (2025). Introduction to the special issue on content moderation on digital platforms. Zie: <https://policyreview.info/articles/analysis/introduction-content-moderation-digital-platforms#editorial>

¹⁰ Caplan, R. (2023). Networked platform governance: The construction of the democratic platform. Zie: <https://ijoc.org/index.php/ijoc/article/view/20035>

personaliseringsmechanismen voor een algoritmische feed. Voor maatschappelijke organisaties is het daarom belangrijk dat ze meer middelen voor directe inspraak krijgen zonder afhankelijk te zijn van platformeigenaren en overheden.

In dit stuk bepleiten wij dat maatschappelijke organisaties een grotere rol moeten innemen in online contentmoderatie. Daarbij kijken we zowel naar de online contentmoderatie in het huidige sociale medialandschap als naar het Fediverse. Daarvoor brengen we in kaart op welke niveaus moderatie plaats vindt en welke taken zijn te onderscheiden in online content moderatie.

2.1 Verschillende niveaus en taken bij moderatie

Om de rol van maatschappelijke organisaties bij moderatie te kunnen bepalen is het belangrijk dat de mogelijkheden en beperkingen van verschillende moderatiemiddelen bekend zijn. Welke niveaus van moderatie bestaan er en welke taken zijn te onderscheiden?

Moderatieniveau

Huidige sociale media combineren een kerninfrastructuur, de fundamentele platformlaag, met technologieën die groepsinteractie mogelijk maken. Facebook en Reddit bieden bijvoorbeeld een kerninfrastructuur voor gebruikers om accounts aan te maken (inclusief ID-management) en functionaliteiten voor groepscommunicatie (Subreddits en Facebook Groups).

Volgens platformwetenschappers is de moderatie van sociale media daarom op te delen in drie niveaus (afbeelding 1): het 'top level', 'middle level' en 'bottom level'. Zoals afbeelding 1 laat zien, wordt content moderatie verricht door meerdere actoren, op verschillende niveaus. De concrete moderatie-instrumenten en wettelijke plichten verschillen per niveau. Daarnaast kunnen de actoren die meewerken aan moderatie, en hoe deze met elkaar samenwerken, per platform verschillen.

Om dit concreter te maken, geeft onderstaande tabel de definities van verschillende soorten moderatie en de moderatiemiddelen per niveau weer.

Niveau	Definitie	Moderatiemiddelen
Top level Platform moderatie (platform/systeem-level governance)	Platformmoderatie is het geheel van regels en normen die voor een heel platform, alle communities en eindgebruikers gelden.	Geautomatiseerde moderatie-instrumenten (inclusief diensten ter bestrijding van illegale inhoud); Algoritmische modellen voor aanbevelingen (vaak gepersonaliseerd op gebruikers); Platform-overkoepelende community guidelines en terms of service, inclusief regels en functionaliteiten voor inhoud en identiteit van gebruikers.
Middle level Community moderatie (groep-level governance)	Community moderatie draait om de regels tussen eindgebruikers binnen een bepaalde groep. Deze groep kan verschillende vormen aannemen (zie de voorbeelden hiernaast). Community moderatie bemiddelt tussen eindgebruikers, platformen en andere communities.	Middle-level media om groepen te vormen: Custom feeds (Bluesky), Community servers (Mastodon), Groups (Facebook), Subreddits (Reddit), Channels (Twitch and YouTube), Language Editions (Wikipedia), etc. Community regels
Bottom level Moderatie door de eindgebruiker (user-level governance)	Eindgebruikers kunnen zelf (in meer of minder grote mate) de zichtbaarheid van content beïnvloeden, door het aanpassen van voorkeuren van algoritmische feeds of het melden van inhoud (flagging).	User interface functionaliteiten om de zichtbaarheid van inhoud aan te passen: "melding", "inhoud niet relevant"

Afbeelding 1: Definitie van moderatieniveaus

Moderatietaken voor maatschappelijke organisaties

Naast de verschillende niveaus, is het belangrijk om onderscheid te maken tussen moderatiebeleid en moderatie-uitvoer. Op grote platformen als X en Facebook wordt het beleid namelijk bepaald door het hoger management, terwijl de dagelijkse uitvoer van dit beleid bij moderatiemedewerkers of vrijwilligers ligt.

Met andere woorden, er zijn dus niet alleen verschillende niveaus waarop moderatie plaats vindt binnen platformen, er is ook een duidelijk verschil tussen het maken en het uitvoeren van beleid. Dit is belangrijk omdat maatschappelijke organisaties de drie niveaus en het verschil tussen beleid en uitvoer kunnen gebruiken om invloed aan te wenden. Door zich bewust te zijn van verschillende taken, kunnen maatschappelijke organisaties bijvoorbeeld gericht lobbyen voor wetgeving over moderatie (beleid), en zich ook inzetten als (trusted) flagger om illegale content aan te merken.

Onderstaande tabel geeft weer wat maatschappelijke organisaties kunnen doen:

Moderatiebeleid: top level	Uitvoer van moderatie: top level
Lobbyen voor bestaande en toekomstige wetgeving Definitie en supervisie van overkoepelende regels, bijvoorbeeld door deelname aan een “Raad van Trust & Safety” of een “Mod Council” Defederatie van instances (voor decentrale sociale media op ActivityPub) Evaluatie van de effectiviteit van geautomatiseerde moderatietools	(Trusted) flagging om illegale content zoals CSAM te melden Verwijderen van illegale inhoud zoals CSAM
Moderatiebeleid: middle level	Uitvoer van moderatie: middle level
Definiëren van regels in een groep (Instance (Mastodon), Custom Feed (Bluesky), Group (Facebook), Subreddit (Reddit)) Faciliteren van participatieve processen voor de vorming van regels met de community Gebruik van groepspecifieke moderatie-instrumenten en processen (bijvoorbeeld publieke logs van moderatiekeuzes)	Infomateriaal voor nieuwe groepsleden beschikbaar maken Melden, blokken, verbannen van ongewenste inhoud en gebruikers

Afbeelding 2: Niveaus en taken voor moderatie

De rol van maatschappelijke organisaties voor contentmoderatie

Samenvattend zien wij dat maatschappelijke organisaties op verschillende manieren invloed kunnen uitoefenen op moderatie. Ze kunnen bijvoorbeeld kiezen om deel te nemen aan een formeel proces voor de definitie of evaluatie van overkoepelende regels voor een bepaald platform, of ze kunnen ervoor kiezen een eigen community op te bouwen. Maatschappelijke organisaties kunnen verschillende redenen hebben om voor een bepaalde manier van moderatie te kiezen. Sommige organisaties willen ervoor zorgen dat een heel platform veilig is, geen illegale inhoud verspreid of ervoor zorgt dat fundamentele en mensenrechten gewaarborgd zijn. Andere organisaties willen misschien een veilige onlineomgeving voor een groep bieden en ervoor zorgen dat deze groep veilig met andere groepen of een heel netwerk kan communiceren (zie bijlage 2 voor voorbeelden van relevante maatschappelijke organisaties).

Hierbij moeten we benadrukken dat er duidelijke verschillen zijn tussen verantwoordelijkheden, rechten, en machtsposities tussen moderatieniveaus (eindgebruikers, communities en platformen).

Voor het top-level is bekend dat maatschappelijke organisaties niet voldoende inspraak krijgen als ze deelnemen aan adviesorganen zoals een raad voor veiligheid. Het is ook onduidelijk op basis van welke criteria maatschappelijke organisatie raadslid of trusted flagger worden. Het is niet transparant wat voor feedback ze op moderatie geven en wat er met hun feedback gebeurt. Daarnaast hebben maatschappelijke organisaties, onderzoekers en experts vaak onvoldoende toegang tot platformdata om de effectiviteit van moderatiebeleid en uitvoer te beoordelen. Uit de literatuur blijkt dat meerdere zaken belangrijk zijn om maatschappelijke moderatie op top-level te

steunen: er moeten regels zijn om feedback van organisaties ook daadwerkelijk naar beleid te vertalen, het moet inzichtelijk zijn welke organisaties worden gekozen als leden van een raad voor moderatie en als trusted flaggers en waarom. Bovenal moet onafhankelijk onderzoek naar moderatiemaatregelen worden gesteund door toegang tot platformdata.

Voor het middle-level is bekend dat communities zelf op verschillende manieren moderatie kunnen beïnvloeden en daarbij vaak met top-level en user-level moderatie rekening moeten houden. Op YouTube kanalen kunnen het YouTube platform en de eigenaar van een kanaal besluiten welke comments op videos worden geblokt. Zogenaamde 'Mod Councils' op Reddit maken het mogelijk voor moderators om commentaar te geven op Reddit's community guidelines, maar Reddit is vervolgens niet verplicht om dit feedback vervolgens te implementeren. En op Facebook Groups kunnen administratoren besluiten om gebruikers te blokken, maar hebben de eindgebruikers geen invloed op wie er administrator van een groep is.

Om de machtsverhoudingen rond moderatie te verbeteren pleiten we voor een decentrale aanpak voor maatschappelijke moderatie. Dit betekent dat maatschappelijke organisaties gebruik kunnen maken van moderatiemiddelen op verschillende niveaus, om tussen een heel netwerk van gebruikers, bepaalde groepen, en eindgebruikers te kunnen bemiddelen. Sociale mediaplatformen in het Fediverse hebben een decentrale architectuur die community governance op een middle-level bevordert. Dat biedt maatschappelijke organisaties specifieke mogelijkheden om hun eigen gemeenschappen te beheren, onafhankelijk van platformeigenaren op een top-level. Daarom verkennen we in de volgende sectie de rol van het Fediverse als een decentrale infrastructuur voor sociale media.

2b. Het Fediverse

Het Fediverse bestaat uit genetwerkte sociale media, die zich al meer dan tien jaar ontwikkelen en die veel wind in de zeilen hebben gekregen door het ongemak met de big tech. Het belangrijkste kenmerk van het Fediverse is dat het decentraal wordt georganiseerd.

Wat zijn gefedereerde netwerken?

Gefedereerde netwerken hebben een gedecentraliseerde serverstructuur en kunnen gegevens met elkaar uitwisselen omdat ze interoperabel zijn.

Gebruikersgegevens staan niet op één centrale locatie, maar zijn opgeslagen bij een van de gefedereerde servers waar de gebruiker staat geregistreerd.

Gebruikers kunnen met gebruikers op andere servers communiceren. Hierdoor kunnen gebruikers van gefedereerde sociale media zich aanmelden bij één aanbieder en toch verbonden blijven met andere platforms en servers. Dit voorkomt user lock-in: gebruikers kunnen makkelijker tussen servers, platformen en diensten wisselen en hun volgers meenemen. Gefedereerde netwerken worden momenteel voornamelijk gebruikt voor sociale netwerken (mogelijk gemaakt door het ActivityPub-protocol) en chatsoftware (mogelijk gemaakt door de Matrix- en XMPP-protocollen), maar ze kunnen breder worden toegepast voor andere diensten.

Het Fediverse is een familie van (sociale) netwerken die met elkaar kunnen 'praten' doordat ze hetzelfde ActivityPub protocol gebruiken. Bekende netwerken zijn Mastodon, PeerTube en Pixelfed, respectievelijk vervangers voor X, Youtube en Instagram. Het aantal netwerken dat gebruik maakt van het ActivityPub protocol groeit. Naast het ActivityPub protocol zijn er ook het AT Protocol, bekend van Bluesky, en Matrix met toepassingen als Element en PubHubs. Hoewel het AT Protocol in potentie een federatief protocol is, is er op het moment slechts een aanbieder (Bluesky), waardoor de toepassing van het protocol momenteel gecentraliseerd is. Waag is betrokken bij het onderzoek of het mogelijk is om een Eurosky instantie te starten die federeert met Bluesky, maar onder Europees moderatietoezicht staat.

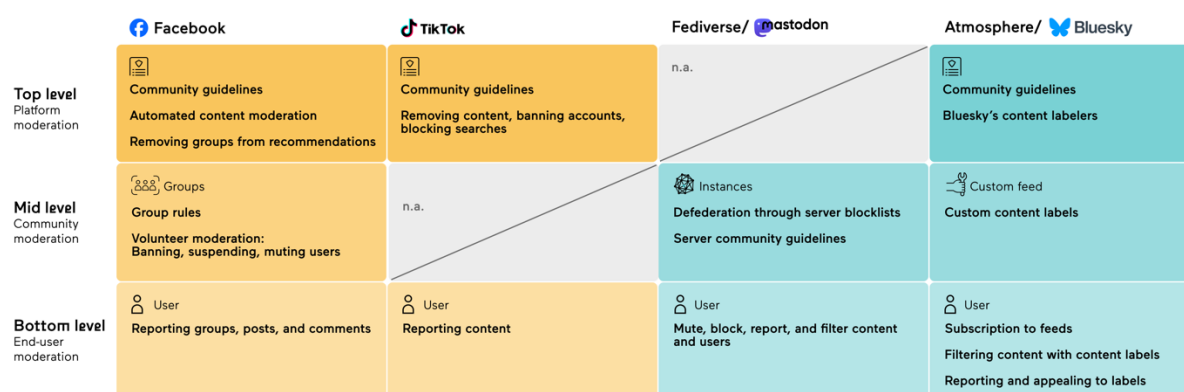
Federatieve protocollen maken het mogelijk dat iedereen een deel van het federatieve netwerk kan hosten. Er zijn wereldwijd meer dan 10.000 Mastodon-servers¹¹ die door verschillende organisaties worden beheerd. Op dit moment heeft het Fediverse meer dan 17 miljoen gebruikers, Bluesky heeft er bijna 38 miljoen en het Matrix-protocol zelfs meer dan 115 miljoen.¹² Afbeelding 3 visualiseert het verschil tussen Mastodon en Instagram. Hier zien we dat mensen zich bij diverse aanbieders, instanties geheten, kunnen aanmelden om met het Mastodon-netwerk te communiceren. Elke Mastodon-instantie heeft specifieke kenmerken. Gebruikers kunnen dus kiezen wat bij hen past. Er zijn ook diverse apps beschikbaar om Mastodon te gebruiken. Bluesky volgt een andere aanpak. Het is op het AT Protocol opgebouwd en biedt gemeenschappen de mogelijkheid om Custom Feeds te bouwen. Deze Custom Feeds zijn algoritmische News Feeds die op basis van verschillende criteria zijn ontworpen. BlackSky¹³ is een Amerikaans project om een Custom Feed te bouwen die relevant is voor zwarte gebruikers en aan de community regels

¹¹ Fediverse servers (2025). Zie: <https://mastodon.help/instances/en>

¹² Federated Works (2024) 'Fediverse & social web statistics'. Zie: <https://federated.works/page/collaboration/statistics>.; Bluesky user count (2025). Zie: <https://bsky-users.theo.io/>.; The Stack (2023) 'Matrix hits 115 million users, as founders look to add OpenID Connect, enterprise features'. Zie: <https://www.thestack.technology/matrix-protocol-users-2023/>

¹³ <https://blog.rudyfraser.com/blacksky-part-one/>

van deze groep voldoet. Sociale media zoals Mastodon en Bluesky bieden daarom verschillende middelen voor maatschappelijke organisaties om zelf community moderatie op een middle level uit te voeren en te bepalen wat voor content mag worden verspreid en welke regels voor content en sociale interacties gelden. Dit is een fundamenteel verschil met platforms zoals TikTok, waar slechts één aanbieder is en gebruikers zich op één centrale plek aanmelden om toegang tot een algoritmische feed te krijgen. Afbeelding 3 biedt een overzicht van de verschillende moderatieniveaus van vier platformen. De afbeelding maakt duidelijk dat sommige platformen niet over alle niveaus beschikken en dat moderatieniveaus ook met elkaar kunnen overlappen. Zo worden bijvoorbeeld Facebook Groups zowel door Facebook's eigen moderatieteam gemodereerd als door groepsadministratoren en gebruikers.



Afbeelding 3: Moderatieniveaus en -actie van verschillende platformen

Vergeleken met platformen zoals Facebook en TikTok biedt het Fediverse ook een andere gebruikerservaring. Gebruikers kunnen van aanbieder wisselen zonder hun volgers te verliezen. De volgers zijn gebonden aan de persoon, niet aan de aanbieder. Wat hier beschreven staat voor Mastodon geldt ook voor de andere familieleden van het Fediverse, zoals Pixelfed en PeerTube. Gebruikers hebben ook meer invloed op hun newsfeeds. Zo maakt Mastodon bijvoorbeeld geen gebruik van algoritmische feeds. De posts verschijnen in chronologische volgorde. Ook is het op Mastodon niet mogelijk om een bericht te citeren en te voorzien van eigen commentaar. Dit zijn bewuste keuzes om de publieke sfeer te verbeteren. Er is geen verdienmodel op basis van algoritmen of ongewenst gedrag (bijvoorbeeld posts belachelijk maken en zo niet bijdragen aan een structureel debat). Het blijkt wel dat nieuwe leden aan de cultuur op het Fediverse moeten wennen. Omdat ze grote platformen gewend zijn, melden veel gebruikers zich bijvoorbeeld automatisch aan bij grootste instanties.¹⁴

Wat het Fediverse bijzonder maakt, is dat een Mastodon-gebruiker ook een account op PeerTube of Pixelfed kan volgen en vice versa. Qua opzet is het Fediverse daarom vergelijkbaar met email, waar gebruikers van verschillende email providers toch met elkaar kunnen communiceren.

2c. Moderatie op het Fediverse

Doordat er niet een centrale aanbieder is, maar er duizenden aanbieders zijn, is moderatie op het Fediverse fundamenteel anders geregeld. Elke aanbieder stelt eigen moderatieregels op, in een zogenaamde Code of Conduct. Daarnaast heeft elke aanbieder eigen moderators. Dit betekent dat de verantwoordelijkheid van het modereren wordt verdeeld over vele partijen en dat deze partijen actief het beleid en de uitvoer van moderatie kunnen vormgeven. De structuur van online moderatie op het Fediverse verschilt van andere platforms daarom zowel wat betreft beleid als uitvoer. Het beleid wordt per aanbieder opgesteld en stelt hen in staat om zelf regels op te stellen. De uitvoer van moderatie is de verantwoordelijkheid van de aanbieder.

In de praktijk zijn er twee soorten moderatiehandelingen die moderators verrichten: 1) moderatie van de instantie zelf en 2) moderatie van de relatie tot andere instanties in het Fediverse.

1. Op de server zelf wordt gemodereerd aan de hand van de regels opgesteld in de Code of Conduct. Leden van de server kunnen posts rapporteren, de moderator beoordeelt dit en achterhaalt of de gerapporteerde content inderdaad in strijd is met het beleid. Moderatoren kunnen vervolgens verschillende acties verrichten, zoals de post verwijderen of een gebruiker (en daardoor diens posts) als 'gevoelig' markeren. In extremere gevallen kan de moderator er ook voor kiezen om een account tijdelijk te bevriezen, 'limiteren' (account is verborgen voor alle gebruikers van een server, behalve volgers) of schorsen (praktisch verwijderen).¹⁵
2. Omdat servers in het Fediverse met elkaar in verbinding staan, kan content ook gefedereerd binnenkomen. Dit betekent dat schadelijke berichten ook via andere kanalen op de feed van een server terecht kunnen komen. Moderatoren hebben daarom ook de mogelijkheid om volledige servers te modereren. Ze kunnen bijvoorbeeld ook een server 'limiteren' (wat op hetzelfde neerkomt als alle accounts op die server limiteren), of een complete server schorsen (de server wordt geblokkeerd en alle banden met die server worden beëindigd). Dit wordt ook wel 'defedereren' genoemd. Er zijn bestaande 'blocklists' met servers die zich schuldig maken aan het verspreiden van schadelijke of illegale content.¹⁶ Een bekend voorbeeld van een gedefedereerde instantie is Gab, een extreemrechts netwerk met nazisympathieën. Doordat deze instantie door veruit de meeste andere instanties geblokkeerd is, is Gab geïsoleerd.¹⁷

Omdat de praktijk van moderatie anders is opgezet op het Fediverse dan bij gecentraliseerde netwerken, komen daar ook andere rollen bij kijken. In deze beleidsbrief onderzoeken we welke rollen maatschappelijke organisaties kunnen en willen innemen, binnen de context van het Fediverse.

¹⁵ NB. Mastodon is hier als voorbeeld gebruikt, op andere platformen in het Fediverse kunnen er andere conventies gelden. https://docs.joinmastodon.org/admin/moderation/?utm_source=chatgpt.com#individual-moderation

¹⁶ Zie bijvoorbeeld <https://github.com/gardenfence/blocklist>

¹⁷ Robertson, A. "How The Biggest Decentralized Social Network Is Dealing With Its Nazi Problem." *The Verge*, July 12, 2019. <https://www.theverge.com/2019/7/12/20691957/mastodon-decentralized-social-network-gab-migration-fediverse-app-blocking>.

3. Methode

Voor dit onderzoek zijn gesprekken met maatschappelijke organisaties en met moderatoren gevoerd. Om context te bieden aan het empirische gedeelte van dit onderzoek, hebben we een literatuurstudie naar online contentmoderatie op het Fediverse uitgevoerd. Dit stelde ons in staat om interviewvragen op te stellen en achtergrondinformatie te verzamelen. In een eerste interview hebben wij vier moderatoren geraadpleegd. Drie van hen modereren op het Fediverse, een van hen modereert voor Kanker.nl. Dit groepsinterview diende ertoe inzicht te krijgen in huidige ervaringen om te modereren op het Fediverse en in huidige online communities. Ook was deze informatie input voor een interview met maatschappelijke organisaties. In dit gesprek hebben we met vijf maatschappelijke organisaties geïnventariseerd welke rollen zij willen en kunnen innemen in online moderatie. Ter aanscherping van het onderzoek hebben wij zes individuele interviews uitgevoerd met organisaties die niet in staat waren om aanwezig te zijn bij het groepsgesprek. De organisaties die we betrokken hebben zijn actief in verschillende domeinen, waaronder mediapartijen als Vers Breton, humanitaire organisaties als Amnesty International, en organisaties met een kritische blik op technologie en de online publieke sfeer, zoals Setup en Movisie. Bijlage 1 geeft weer welke organisaties we gesproken hebben.

4. Bevindingen

Om in kaart te kunnen brengen welke rollen maatschappelijke organisaties kunnen spelen in het Fediverse, hebben we twee groepen geïnterviewd. Allereerst zijn we in gesprek gegaan met moderators die op dit moment actief zijn. Dit geeft inzicht in het dagelijks werk van moderatie en de uitdagingen waar moderators tegenaan lopen. De bevindingen hebben we gebruikt in gesprekken met maatschappelijke organisaties die een rol zouden kunnen of willen vervullen bij maatschappelijke moderatie.

4a. Uitdagingen voor moderators

We hebben vier moderators bijeengebracht. Het zijn mensen die al enige jaren deze rol vervullen. De schaalgrootte van het netwerk varieerde van Fosstodon (8.700 gebruikers) tot een kleinere Lemmy server (1.200 gebruikers). In het groepsgesprek werden uitdagingen naar voren gebracht die betrekking hebben op de praktische uitvoering van het moderatiewerk en uitdagingen die samenhangen met de mentale belasting die dit werk met zich meebrengt. Hieronder worden de uitdagingen kort toegelicht en aangevuld met informatie uit het literatuuronderzoek. We doen ook steeds aanbevelingen hoe het betreffende vraagstuk geadresseerd kan worden.

Gebrek aan kennisdeling

Op dit moment bestaat er in Nederland geen platform, opleiding of kennisbank waar kennis over moderatie kan worden gedeeld. Omdat inzichten, tools en 'best practices' nauwelijks worden overgedragen, moeten moderators het wiel steeds zelf uitvinden. Dit heeft een negatief effect op de efficiëntie van het moderatiewerk. Het maakt moderators ook onnodig kwetsbaar.

Hoge werkdruk en complexe beslissingen

Moderators krijgen regelmatig met (ethisch) complexe kwesties te maken. Zij worden bijvoorbeeld gevraagd om beslissingen te nemen over gevoelige of politiek controversiële onderwerpen, zoals desinformatie rondom internationale conflicten. Keuzes over moderatie kunnen lastig zijn, omdat het vereist een moreel standpunt in te nemen.¹⁸ Daarnaast is het modereren tijdrovend wanneer gerapporteerde onderwerpen niet terugkomen in de Code of Conduct van de server waar de moderator actief is, omdat dit veel onderzoek vereist.

Naast politiek controversiële kwesties kan ook de omvang van de gerapporteerde content veel werk opleveren. Het is mogelijk om een hele instantie aan te merken als problematisch. Als dat gebeurt, betekent dit tijdrovend onderzoek om uit te zoeken of het terecht zou zijn die instantie te blokkeren voor alle gebruikers. Dit kan tot hoge werkdruk leiden. Het is belangrijk om hierbij in acht te nemen dat instanties meestal volledig afhankelijk zijn van donaties, wat betekent dat moderators zelden gecompenseerd worden voor hun werk.¹⁹

Hoewel platforms als Mastodon beschikken over goede rapportagesystemen, kunnen deze niet alles opvangen. Aanvallen vinden bijvoorbeeld vaak plaats via meerdere accounts, waardoor het achterhalen van de specifieke context soms veel onderzoek vereist. Tools om dit proces te

¹⁸ Spencer-Smith, C. and Tomaz, T. (2025) 'Labour Pains: Content Moderation Challenges in Mastodon Growth.' *Internet Policy Review*. Zie: <https://doi.org/10.14763/2025.1.1831>.

¹⁹ Mansoux, A., & Abbing, R. R. (2020). 'Seven theses on fediverse and the becoming of FLOSS'. In K. Gansing & I. Luchs (Eds.), *The eternal network: The ends and beginnings of network culture*. Zie: <https://networkcultures.org/blog/publication/the-eternal-network/>

versnellen zouden de werkdruk verlichten. Zoals eerder genoemd, worden er in de literatuur ook andere tools aangestipt, zoals de behoefte aan geautomatiseerde tools om sommige handelingen die nu handmatig worden uitgevoerd te versnellen.²⁰

Uit de literatuur blijkt dat de toestroom van een groot aantal nieuwe leden ook tot hoge werkdruk kan leiden.²¹ Sommige onderzoekers wijzen er daarom op dat een te grote moderatielast kan leiden tot een gebrek aan efficiëntie, wat ten koste kan gaan van de veiligheid en inclusiviteit van de server.²²

Gerichte aanvallen met illegale content

Moderatoren zijn regelmatig het doelwit van online aanvallen. Kwaadwillenden kunnen bijvoorbeeld bewust illegale content zoals kindermisbruikmateriaal (CSAM) plaatsen, met het doel om deze vervolgens zelf te rapporteren. Dit vereist snel handelen van moderatoren, die dit veelal zelf moeten oplossen. De literatuur wijst erop dat er nog niet voldoende tools zijn voor het onderscheppen van CSAM op het Fediverse. Bestaande tools zijn namelijk ontwikkeld voor gecentraliseerde systemen en kunnen nog niet effectief ingezet worden op gedecentraliseerde netwerken.²³ Daarnaast kan illegale content in het Fediverse ook 'gefedereerd' binnenkomen: ongewenste content van een andere instantie kan ook bij de leden van hun eigen instantie binnenkomen.²⁴ Dit betekent dat moderatoren niet alleen alert moeten zijn op posts van hun eigen leden, maar ook op ongewenste content van buitenaf.

Naast de vraag hoe illegale content effectief onderschept en verwijderd kan worden, speelt het vraagstuk van juridische aansprakelijkheid. De moderatoren waren in sommige gevallen bezorgd dat zij persoonlijk aansprakelijk gesteld zouden kunnen worden voor illegale content die tijdens een aanval op hun server geplaatst wordt. Een mogelijke oplossing is het oprichten van een stichting om persoonlijke aansprakelijkheid te voorkomen.

Intimidatie en burn-out

Naast gerichte aanvallen, vindt er online intimidatie van moderatoren plaats. Dit gaat bijvoorbeeld om bedreigingen, doxing (het openbaar maken van privégegevens) en het benaderen van moderatoren via hun privé e-mailadres. Vijandige bejegening en andere vormen van intimidatie worden vaak als zware mentale belasting ervaren. Wanneer moderatoren niet voldoende rust kunnen nemen of onprettige ervaringen niet met collega's kunnen delen, kan dit tot burn-out leiden. Moderatoren benadrukken daarom het belang van een team waarin over moeilijke gebeurtenissen en ervaringen gesproken kan worden.

²⁰ Zhang, Z. et al. (2024) 'Trouble in Paradise? Understanding Mastodon Admin's Motivations, Experiences, and Challenges Running Decentralised Social Media.' *Proceedings of the ACM on Human-Computer Interaction* 8. Zie: <https://doi.org/10.1145/3687059>.

²¹ Ermoshina, K. and Musiani, F. (2025). 'Safer Spaces by Design? Federated Socio-technical Architectures in Content Moderation.' *Internet Policy Review*. Zie: <https://doi.org/10.14763/2025.1.1827>. Spencer-Smith, C. and Tomaz, T. (2025) 'Labour Pains: Content Moderation Challenges in Mastodon Growth.' *Internet Policy Review*. Zie: <https://doi.org/10.14763/2025.1.1831>.

²² Zhang, Z. et al. (2024) 'Trouble in Paradise? Understanding Mastodon Admin's Motivations, Experiences, and Challenges Running Decentralised Social Media.' *Proceedings of the ACM on Human-Computer Interaction* 8. Zie: <https://doi.org/10.1145/3687059>.

²³ Thiel, D. en DiResta, R. (2023). 'Child safety on federated social media', *Stanford Internet Observatory*. Zie: <https://stacks.stanford.edu/file/druid:vb515nd6874/20230724-fediverse-csam-report.pdf>

²⁴ Agarwal et al. (2024). 'Decentralised Moderation for Interoperable Social Networks: A Conversation-based Approach for Pleroma and the Fediverse', *Proceedings of the Eighteenth International AAAI Conference on Web and Social Media*. Zie: <https://ojs.aaai.org/index.php/ICWSM/article/view/31293/33453>

Aanbeveling

Moderatoren vervullen een belangrijke functie in het maatschappelijk debat en behoren daar voldoende in ondersteund te worden. Om moderatie zowel effectiever als toegankelijker te maken voor nieuwe moderatoren is het essentieel om een centrale plek voor kennisuitwisseling te realiseren. Ook een publieksprogramma met workshops, podcasts, cursusaanbod en publiek debat kan daaraan bijdragen.

Gezien de expertise en vaardigheden die moderatoren nodig hebben, kunnen we spreken van een verantwoordelijke functie waar een financiële compensatie tegenover moet staan. Door het als een volwaardige functie te honoreren, hoeft het moderatiewerk niet bovenop andere werkzaamheden te komen. Daarnaast is er behoefte aan juridische ondersteuning ter voorkoming van persoonlijke aansprakelijkheid.

Om burn-out te voorkomen is samenwerking in teamverband aanbevolen. Daarnaast is er behoefte aan een plek om (onprettige) ervaringen te kunnen delen en aan trainingen over het omgaan met onveilige en intimiderende situaties. Ook is er behoefte aan betere moderatie tools, zodat moderatoren hun werk op een efficiëntere manier uit kunnen voeren.

4b. Rol van maatschappelijke organisaties

Met maatschappelijke organisaties zijn we in gesprek gegaan over de rol die zij zouden willen of kunnen nemen bij maatschappelijke moderatie. Deze gesprekspartners zijn in te delen in twee categorieën: 1) organisaties die zich bezighouden met het digitale domein, maar (nog) geen rol of verantwoordelijkheden voor zichzelf zien wat betreft content moderatie en 2) organisaties die wel een rol/verantwoordelijkheid voor zichzelf zien vooral op het gebied van bewustwording en educatie.

Zoals eerder genoemd betekent maatschappelijke moderatie dat maatschappelijke organisaties zeggenschap over moderatiebeleid en uitvoer krijgen. Waag definieert hiervoor de volgende rollen. We leggen eerst deze rollen uit en bespreken daarna hoe maatschappelijke organisaties hier tegenaan kijken.

1. **Aanbieder van een instantie (zelf-hosting):** organisatie die zelf of met anderen een instantie opzet om bijvoorbeeld een community van lezers, bezoekers, of andere groepen een plek te geven. Dit omvat ook het bepalen van moderatieregels voor deze groep en voor de federatie met andere servers.
2. **Moderator:** organisatie die actief meedoet bij de uitvoer van moderatie. Deze rol vereist niet dwingend dat de organisatie ook een instantie zelf host.
3. **Adviseur rond moderatiebeleid:** organisatie die advies geeft op Europese en nationale wet- en regelgeving rond onlinemoderatie en op moderatieregels op bepaalde platformen en instanties.
4. **Gebruiker met eigen contentstrategie voor het Fediverse:** organisatie die het Fediverse gebruikt om eigen content te verspreiden. Dit kan in de rol van een host zijn of in de rol van een gebruiker.
5. **Dienstaanbieder voor gebruikers, moderatoren en hosts:** organisatie die diensten aanbiedt en ondersteunt in het opbouwen van kennis, vaardigheden of IT-diensten.

Maatschappelijke organisaties tussen aanbieder en verkennende gebruiker

Uit de gesprekken bleek dat voor de meeste deelnemers het Fediverse nog onbekend terrein is. Dit maakt het lastig om een keuze te maken over de rol die zij daar het beste in kunnen spelen. Daarom zien onze gesprekspartners zich vooral in een rol als verkennende gebruiker van het Fediverse. Waag heeft zelf een dubbele rol als een maatschappelijke organisatie die ook zelf modereert op haar eigen instantie waag.social. Enkele deelnemers zijn gebruiker met een actieve contentstrategie. Of het hosten een optie is hangt af van de verkenning op een aantal gebieden:

Kosten, baten en businessmodellen

Het is voor de organisaties (nog) niet altijd duidelijk wat de kosten en baten zijn van maatschappelijke moderatie. De problemen met de platforms van big tech worden onderkend. Het is echter de vraag wie die rol dan moet gaan innemen. De stap om zelf een aanbieder van een sociaalmedianetwerk te worden is een strategische afweging die ook financiële en organisatorische consequenties heeft. De vraag is hoe die rol gefinancierd moet worden en of deze rol binnen de missie van de organisatie past. Enkele organisaties gaven aan dat ze momenteel niet voldoende capaciteit hebben om een moderatierol op zich te nemen. Hieruit blijkt dat er momenteel niet voldoende capaciteit is of niet wordt vrijgemaakt binnen organisaties om moderatietaken te overnemen. Voor maatschappelijke organisaties is daarom ook de vraag naar het businessmodel van belang.

Herdefiniëren van bereik

Het bereiken van een doelgroep is een belangrijke afweging bij de keuze voor het gebruik van een platform. Een strategie voor het Fediverse zal altijd in samenhang zijn met de vraag hoe het bereik daar vergroot kan worden. Bestaande voorbeelden uit de praktijk laten zien dat maatschappelijke organisaties kunnen leren van hoe bijvoorbeeld online-marketeers en mediabedrijven het Fediverse gebruiken voor hun communicatie. Uit deze ervaringen blijkt dat het Fediverse geschikt is voor intensievere discussies en bindingen met een groep geïnteresseerde mensen, in plaats van voor het virale verspreiden van inhoud.²⁵ Een andere punt is dat het Fediverse geen algoritmen biedt die posts kunstmatig op de voorgrond kunnen brengen. Daarnaast wordt bereik op een andere, minder uitgebreide manier gemeten dan op mainstream sociale media, wat onder andere belangrijk is voor communicatie- en marketingafdelingen. Mediabedrijven ontwikkelen manieren om wel te kunnen meten welke invloed een aanwezigheid op het Fediverse heeft op clicks. Dit vraagt om het investeren in en ontwikkelen van nieuwe kennis bij communicatiemedewerkers en marketeers. Dit is ook een onderzoeks- en ontwikkelingsdomein waar op dit moment nauwelijks of geen fondsen actief zijn.

Hoe om te gaan met 'neutraliteit'

Uit de gesprekken met maatschappelijke organisaties komt ook de vraag naar voren wat maatschappelijke moderatie betekent voor de wens van organisaties om neutraal te zijn. Maatschappelijke organisaties zijn terughoudend om zich te positioneren in politiek gevoelige debatten. Dat kwamen we vooral bij mediapartijen en mensenrechtenorganisaties tegen. Er is namelijk een grijs gebied, waar een post niet de wet of moderatieregels overtreedt, maar ook niet bijdraagt aan een gezond online ecosysteem. Zoals de ervaring van de moderators leert, vereist

²⁵ Public Spaces (2023). 'De Mastodon marketeer: in gesprek met Wendy Hoenkamp over ethisch online opereren'. Zie: <https://publicspaces.net/2023/11/07/de-mastodon-marketeer-in-gesprek-met-wendy-hoenkamp-over-ethisch-online-opereren/>. Flipboard (2025) 'Lessons on the Road to Reviving Journalism via the Fediverse'. Zie: <https://medium.com/fedi-curious/lessons-on-the-road-to-reviving-journalism-via-the-fediverse-01b3748cacc5>

moderatie keuzes te maken. Dat kan botsen met de wens om een neutrale positie te behouden, maar is wel nodig om te zorgen dat het onlinegesprek geen schadelijke elementen bevat.

Associaties met ongewenste inhoud

De moderators zeiden daarnaast dat maatschappelijke organisaties moeten nadenken over de associatie die ze opbouwen met gebruikers en ongewenste inhoud. Volgens hen speelt de vraag hoe kan worden omgegaan met de associatie of relatie die ontstaat wanneer gebruikers lid worden van een bepaalde instantie. Sommige maatschappelijke organisaties kunnen bijvoorbeeld content plaatsen over hun activiteiten, maar niet geassocieerd willen worden met de informatie die leden via hun server posten.

Maatschappelijke organisaties als moderators?

De organisaties bevestigen de behoefte aan actieve moderatie. Uit dit onderzoek blijkt dat er verder moet verkend worden of en hoe maatschappelijke organisaties dit op zich kunnen nemen. Dat vereist een verdere verkenning van hoe deze moderatietaken aansluiten bij het dagelijks werk van maatschappelijke organisaties met verschillende missies, doelgroepen, competenties en kernactiviteiten. Deze bevindingen wijzen op de noodzaak om draagvlak en capaciteit voor deze taken binnen het maatschappelijk veld te stimuleren. Enkele organisaties, waaronder Amnesty International Nederland en Waag, geven aan dat ze reeds actief contentmoderatie op zich nemen.

Maatschappelijke organisaties als adviseur voor moderatiebeleid?

Sommige van de bevraagde organisaties zien vooral een adviserende rol rond regelgeving voor moderatie. Partijen zoals Sharing Perspectives Foundation en Defend Democracy geven aan dat ze kunnen helpen om mee te denken over de regels en kaders waarbinnen moderatie kan plaatsvinden, zonder zelf de uitvoerende partij te zijn.

Maatschappelijke organisaties als dienstenaanbieder?

Naast de verkennende rol zien de maatschappelijke organisaties vooral een rol als dienstenaanbieder van kennis en trainingen. Met betrekking tot moderatie denken de organisaties dat ze een bijdrage kunnen leveren door kennisopbouw en door het bevorderen van bewustwording. Organisaties zoals de Koninklijke Bibliotheek, Setup, Waag, Mediawijsheid.nl en Noorderlicht zien een rol voor zichzelf om bij te dragen aan bewustwording en training van deelnemers in het herkennen en omgaan met online haat en discriminatie en misinformatie. Soms omvat dit ook trainingen aan gebruikers geven zodat zij onderling afspraken kunnen maken over de juiste omgangsvormen en moderatie. De verantwoordelijkheid voor moderatie ligt volgens sommige van deze partijen bij andere maatschappelijke organisaties en de burgers zelf.

Samenvatting: Wat is er nodig om maatschappelijke organisaties bij moderatie te betrekken?

Maatschappelijke organisaties hebben de volgende dingen nodig om actief bij te dragen aan onlinemoderatie:

- Verkenning van kosten en baten, businessmodellen en use cases: Maatschappelijke organisaties hebben behoeften aan kennis over het Fediverse, hoe het kan worden ingezet om doelgroepen te bereiken, en hoe de aanwezigheid op het Fediverse gekoppeld kan worden aan verdienmodellen.

- Discussies over hoe organisaties zelf moderatieregels kunnen bepalen en moderatiekeuzes kunnen nemen (neutraliteit): Maatschappelijke organisaties kunnen het doel hebben om zich neutraal op te stellen in publieke debatten. Een nog verder te onderzoeken vraag is daarom welke positie deze organisaties willen of kunnen innemen in het modereren van content.
- Investeren in capaciteit: Maatschappelijke organisaties geven aan niet voldoende tijd en middelen te hebben voor de uitvoer van moderatie. Ondersteuning door derde partijen kan maatschappelijke organisaties in staat stellen om te modereren.

5. Aanbevelingen

Op basis van het onderzoek en de gesprekken met moderatoren en maatschappelijke organisaties komen we tot de volgende aanbevelingen.

Programma Maatschappelijke Moderatie

Moderatoren dragen een verantwoordelijkheid voor een maatschappelijke taak. Om hun taak naar behoren te kunnen uitoefenen is er behoefte aan ondersteuning op een aantal gebieden. In een Programma Maatschappelijke Moderatie kunnen onderstaande behoeften worden ondergebracht. Het programma kan worden uitgevoerd door een consortium van partijen of door een nieuw op te richten expertisecentrum. Een aantal geïnterviewde organisaties hebben aangegeven hier een rol voor zichzelf te zien. Wij adviseren dat het programma de volgende onderdelen meeneemt:

- **Juridisch advies:** Er is behoefte aan (juridische) ondersteuning en kennis, om een strategie te ontwikkelen die enerzijds snelle en effectieve reactie op aanvallen mogelijk maakt (volgens de bevrageden zou een zogeheten ‘threat model’²⁶ nuttig zijn) en anderzijds bijdraagt aan het voorkomen van persoonlijke aansprakelijkheid wanneer er illegale content op de server verschijnt.
- **Ontwikkelen van moderatietools:** Er is behoefte aan betere moderatietools, zodat moderatoren hun werk op een efficiëntere manier uit kunnen voeren. Ook moet betrokkenheid van gebruikers gestimuleerd worden. Zij spelen zelf een belangrijke rol in het handhaven van de cultuur en veiligheid van een instantie en kunnen bijdragen door schadelijke inhoud actief te melden. Hierbij is het volgens de bevrageden ook belangrijk mensen te betrekken die slachtoffer van online geweld worden en een behoefte aan veiligere omgevingen hebben.
- **Opbouwen van ‘communities of practice’ voor kennisdeling:** Om burn-out te voorkomen, is samenwerking in teamverband aanbevolen. Daarnaast is er behoefte aan een plek voor moderatoren om (onprettige) ervaringen te kunnen delen en aan trainingen over het omgaan met onveilige en intimiderende situaties. Om moderatie zowel effectiever als toegankelijker te maken voor nieuwe moderatoren, is het essentieel om een centrale plek voor kennisuitwisseling te realiseren, zoals een kennisbank of een adviesorganisatie zoals IFTAS²⁷ in de Verenigde Staten.
- **Financiële ondersteuning:** Om los te komen van big tech’s sociale media en om het gebruik van federatieve sociale media te stimuleren is financiële steun van de overheid cruciaal. Op dit moment is het Fediverse vrijwel volledig afhankelijk van donaties en is moderatie vrijwilligerswerk. Het is aan te bevelen om een speciaal fonds op te zetten voor maatschappelijke moderatie waar maatschappelijke organisaties subsidie kunnen aanvragen. Dit kan helpen om personeel aan te nemen dat een actieve moderatierol vervult en om moderatiewerk te professionaliseren, maar ook om het platform te kunnen onderhouden.
- **Publieksprogramma’s voor bredere maatschappelijke betrokkenheid en kennis over het onderwerp:** er ontbreekt nog steeds kennis van de mogelijkheden van federatieve sociale media en hoe online omgevingen worden gemodereerd. Het is daarom belangrijk om de verschillende partijen die verantwoording dragen voor moderatie en goed gedrag (individuele burgers/online gebruikers en maatschappelijke organisaties) te informeren via publiekelijk

²⁶ Een threat model is een proces wat mogelijke aanvallen en hun methoden identificeert en oplossingen vindt om deze af te wenden.

²⁷ IFTAS (2025). ‘Frequently asked questions’. Zie: <https://connect.iftas.org/faq/>

toegankelijke informatieprogramma's. Denk aan moderatiewerkshops, events, podcasts, online tutorials of hackathons.

- Betrokkenheid bij **beleidsontwikkeling** en goede afstemming met verantwoordelijke overheden en diensten. Toekomstig beleid moet rekening houden met hoe het Fediverse aansluit bij bestaande wettelijke kaders zoals de Digital Services Act. Maatschappelijke organisaties moeten worden betrokken bij deze beleidsontwikkeling. Maatschappelijke organisaties kunnen bijvoorbeeld bijdragen aan de vormgeving van moderatieregels die volgens hen nodig zijn en ze kunnen feedback geven over de uitvoerbaarheid van beleid. Omdat maatschappelijke organisaties hun eigen belangen inbrengen en geprivilegeerde rollen toegewezen krijgen moet inzichtelijk zijn welke organisaties worden gekozen als leden van een beleidsorgaan of als trusted flaggers en waarom.
- **Effectieve zeggenschap** bij beleidsontwikkeling is belangrijk. Er moeten duidelijke regels zijn hoe een bepaald beleidsorgaan (zoals een raad voor veiligheid) feedback van organisaties naar beleid vertaalt. Deze regels moeten publiek toegankelijk zijn. Daarnaast is belangrijk dat een diverse set van organisaties deelneemt aan een beleidsorgaan, omdat moderatie vaak culturele gevoeligheid en expertkennis vereist. Het is vooral belangrijk dat maatschappelijke organisaties die kwetsbare groepen vertegenwoordigen een stem in platformmoderatie krijgen.
- Moderatie vereist daarnaast **toegang tot platformdata**, zodat maatschappelijke organisaties onafhankelijk onderzoek naar de effectiviteit van moderatiemaatregelen kunnen doen.

Living labs

Een aantal organisaties is zelf host van Mastodon en draagt daarmee reeds verantwoordelijkheid als aanbieder. Voor de meeste organisaties is het een strategische vraag welke rol ze willen en kunnen spelen. Maatschappelijke organisaties zien het belang van het 'vermaatschappelijken' van de verantwoordelijkheid voor sociale media, maar hebben onvoldoende zicht op wat die verantwoordelijkheid behelst. Er is behoefte aan een programma dat experimenteerterruimte biedt voor diverse vormen van content, interactie en publieksbereik, en dat daarnaast in beeld brengt welke consequenties dat heeft voor organisatie en bedrijfsmodellen. Een programma in de vorm van een living lab past daar goed bij. Living labs bieden de ruimte aan praktijkgebonden onderzoek en ontwikkeling en maken het mogelijk om goed in te spelen op de dynamische omgeving waarin sociale media zich bewegen.

Fonds voor maatschappelijke moderatie

Op dit moment is het Fediverse vrijwel volledig afhankelijk van donaties. Het is aan te bevelen om een fonds op te zetten voor maatschappelijke moderatie waar maatschappelijke organisaties subsidie kunnen aanvragen als zij deze maatschappelijke taak uitoefenen. Dit biedt de mogelijkheid om te professionaliseren en het platform te kunnen onderhouden.

Voorbeeldfunctie overheid

De overheid kan het goede voorbeeld geven door zelf actief aan deze ontwikkeling bij te dragen. Dat kan door zelf actief op het Fediverse aanwezig te zijn en communicatie die nu via big tech platforms wordt verspreid ook actief te verspreiden via het Fediverse. Daarbij kunnen verwijzingen naar het Fediverse standaard in publiekscampagnes en websites worden meegenomen. Ook kan een equivalent van de te besteden middelen aan het adverteren op big tech platforms ingezet worden voor communicatie via het Fediverse.

Conclusie

Deze verkenning heeft maatschappelijke moderatie in beeld gebracht, een urgent vraagstuk dat nog zeer onderbelicht is geweest. Dit is een eerste stap. Verder onderzoek is noodzakelijk, zeker aangezien de problematiek rondom de dominantie van big tech bovenaan de maatschappelijke en politieke agenda zal blijven staan. Met maatschappelijke moderatie en het federatieve internet breiden we de mogelijkheden uit voor het terugwinnen van strategische autonomie in het digitale domein. Het is van strategisch belang om de verantwoordelijkheid daarvoor beter in beeld te brengen en deze verkenning voort te zetten. Bovenstaande aanbevelingen bieden richting. Concreet kan implementatie van de aanbevelingen worden vormgegeven door het aanstellen van een kwartiermaker of het ondersteunen van een alliantie van maatschappelijke organisaties (zogenaamde 'frontrunners') die de uitdaging van maatschappelijke moderatie aan willen gaan.

Bronnenlijst

Agarwal et al. (2024). 'Decentralised Moderation for Interoperable Social Networks: A Conversation-based Approach for Pleroma and the Fediverse', Proceedings of the Eighteenth International AAAI Conference on Web and Social Media. Zie:

<https://ojs.aaai.org/index.php/ICWSM/article/view/31293/33453>

Ahmad, Sana, and Martin Krzywdzinski. "Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains." In The MIT Press eBooks, 77–96, 2022.

<https://doi.org/10.7551/mitpress/13835.003.0008>.

Amnesty International (2025). A Thousand Cuts. Technology-facilitated gender-based violence against Poland's LGBTI community on X. Zie:

<https://www.amnesty.org/en/documents/eur37/0098/2025/en/>

Badouard, R. en Bellon, A. (2025). Introduction to the special issue on content moderation on digital platforms. Zie: <https://policyreview.info/articles/analysis/introduction-content-moderation-digital-platforms#editorial>

Bluesky user count (2025). Zie: <https://bsky-users.theo.io/>.

Caplan, R. (2023). Networked platform governance: The construction of the democratic platform.

Zie: <https://ijoc.org/index.php/ijoc/article/view/20035>

Ermoshina, K. and Musiani, F (2025). 'Safer Spaces by Design? Federated Socio-technical Architectures in Content Moderation.' Internet Policy Review. Zie:

<https://doi.org/10.14763/2025.1.1827>.

FediDB. (2015) 'Fediverse servers'. Zie: <https://fedidb.com/servers>

Federated Works (2024). 'Fediverse & Social Web Statistics'. Zie:

<https://federated.works/page/collaboration/statistics>

Flipboard (2025) 'Lessons on the Road to Reviving Journalism via the Fediverse'. Zie:

<https://medium.com/fedi-curious/lessons-on-the-road-to-reviving-journalism-via-the-fediverse-01b3748cacc5>

Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation and hidden decisions that shape social media.

Haime, Z., Kennedy, L. en Biddle, L. (2025) 'Experiences of Moderation, Moderators, and Moderating by Online Users Who Engage with Self-Harm and Suicide Content', Digital Society 4,

8. Zie: <https://link.springer.com/article/10.1007/s44206-025-00166-x#citeas>

IFTAS (2025). 'Frequently asked questions'. Zie: <https://connect.iftas.org/faq/>

Mansoux, A., & Abbing, R. R. (2020). 'Seven theses on fediverse and the becoming of FLOSS'. In K. Gansing & I. Luchs (Eds.), The eternal network: The ends and becomings of network culture. Zie: <https://networkcultures.org/blog/publication/the-eternal-network/>

Ministerie van Onderwijs, Cultuur en Wetenschap, Vereniging van Nederlandse Gemeenten & Koninklijke Bibliotheek. (2023). Vernieuwing Bibliotheekconvenant 2024–2027. <https://www.kb.nl/organisatie/beleid/bibliotheekconvenant>

Moran, R. E., Schafer, J., Bayar, M. and Starbird, K. (2025) 'The End of Trust and Safety?: Examining the Future of Content Moderation and Upheavals in Professional Online Safety Efforts', The End of Trust and Safety?: Examining the Future of Content Moderation and Upheavals in Professional Online Safety Efforts. Zie: <https://dl.acm.org/doi/10.1145/3706598.3713662>

Robertson, A. "How The Biggest Decentralized Social Network Is Dealing With Its Nazi Problem." The Verge, July 12, 2019. <https://www.theverge.com/2019/7/12/20691957/mastodon-decentralized-social-network-gab-migration-fediverse-app-blocking>.

Spencer-Smith, C. and Tomaz, T. (2025) 'Labour Pains: Content Moderation Challenges in Mastodon Growth.' Internet Policy Review. Zie: <https://doi.org/10.14763/2025.1.1831>.

The Stack (2023) 'Matrix hits 115 million users, as founders look to add OpenID Connect, enterprise features'. Zie: <https://www.thestack.technology/matrix-protocol-users-2023/>

Thiel, D. en DiResta, R. (2023). 'Child safety on federated social media', Stanford Internet Observatory. Zie: <https://stacks.stanford.edu/file/druid:vb515nd6874/20230724-fediverse-csam-report.pdf>

Project Online Content Moderatie (PrOCOM) (2022). Zie: https://hetccv.nl/app/uploads/2023/08/20221108_Projectplan_PrOCOM.pdf

Public Spaces (2023). 'De Mastodon marketeer: in gesprek met Wendy Hoenkamp over ethisch online opereren'. Zie: <https://publicspaces.net/2023/11/07/de-mastodon-marketeer-in-gesprek-met-wendy-hoenkamp-over-ethisch-online-opereren/>.

Usenet Historical Collection (n.d.). Zie: <https://archive.org/details/usenethistorical>.

Ward, S., Davis, J. L., Mackenzie, A, and Jones, P. K. (2025) 'Moral Entrepreneurs of the Mastodon Migration.' Symbolic Interaction. Zie: <https://doi.org/10.1002/symb.1244>.

Zhang, Z. et al. (2024) 'Trouble in Paradise? Understanding Mastodon Admin's Motivations, Experiences, and Challenges Running Decentralised Social Media.' Proceedings of the ACM on Human-Computer Interaction 8. Zie: <https://doi.org/10.1145/3687059>.

Bijlage 1: Lijst van gesprekspartners

	Maatschappelijk organisatie / sociale media	Rol van de bevroagde medewerker
1	Sharing perspectives foundation	Executive Director
2	Mediawijzer/Mediawijsheid.nl	
3	Versbeton	Journalist
4	Defend Democracy	Oprichter en Executive Director
5	Cybersoek	Projectmanager
6	Setup	Algemeen directeur
7	Kanker.nl	Community manager (moderator)
8	Movisie	Programmableider Kennisplatform Inclusief Samenleven (KIS)
9	Amnesty International	Community manager voor aansturing van het content team
10	Noorderlicht	Artistiek directeur
11	Koninklijke Bibliotheek	Directieteamlid Lezen en Digitale Geletterdheid
12	Waag Futurelab Mastodon server	Moderator
13	Mastodon server (Fosstodon)	Moderator
14	Fediverse servers (Lemmy en Mastodon servers)	Moderator

Bijlage 2: Relevante organisaties ter bescherming van fundamentele en mensenrechten

Het is belangrijk dat een diverse groep maatschappelijke organisaties een stem in platformmoderatie krijgen. Onderzoek laat zien dat bepaalde groepen opmerkelijk veel last van onlinebedreiging krijgen. Dat betreft jongeren, mensen met andere seksuele geaardheid of geslachtsidentiteit en mensen met migratieachtergrond. Het is belangrijk dat organisaties meedoen bij moderatie die bekend zijn met de problemen van specifieke kwetsbare groepen mensen. Daarnaast zijn er organisaties die zich inzetten voor de bescherming van mensenrechten, digitale rechten, en een gezonde publieke debattencultuur. Om effectief aan moderatie bij te dragen is het belangrijk dat deze organisaties diverse vormen van steun ontvangen (zie onze aanbevelingen op de pagina's 20 tot 22). Hierbij noemen we enkele voorbeelden van organisaties die relevante rollen voor moderatie kunnen spelen:

Belangenverenigingen voor het adresseren van discriminatie:

- Stichting Stop Pesten Nu
- De Goede Zaak
- Het Kinderrechtencollectief
- Nederlands Jeugdinstituut
- Monitor Politieke Ontwikkeling en Meldpunt Islamofobie
- Transgender Netwerk Nederland
- Federatie COC Nederland

Mensenrechtenorganisaties actief in het digitale domein:

- Amnesty International Nederland
- Autoriteit Persoonsgegevens
- Bits of Freedom
- Oxfam Novib

Verbeteren publieke debat in onze democratie:

- PublicSpaces (coalitie van publieke organisaties die zich bezighouden met de publieke digitaal domein)
- Koninklijke Bibliotheek
- ProDemos
- Stichting Democratie en Media
- Defend Democracy

Beheren van online omgevingen (zoals online fora) voor specifieke groepen burgers:

- Kanker.nl en Vereniging BovenJan