



Independent Federated Trust and Safety

Better Social Media for All: Supporting content and conduct moderation in the Fediverse

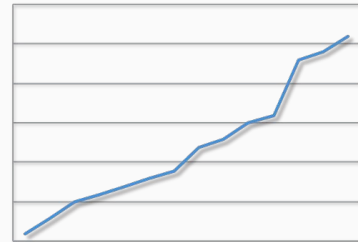
Fediverse Moderation



The Fediverse is growing, but the value derived from context-aware, volunteer moderators is already being tested.

Service providers are a mix of volunteer community leaders, hobbyists and technologists. Trust and safety provides the core of the social network value, but without easy-to-access, evidence-based resources, T&S will continue to degrade.

Service providers and content moderators are searching for help with blocklists, moderation guidance, and complex regulatory issues.

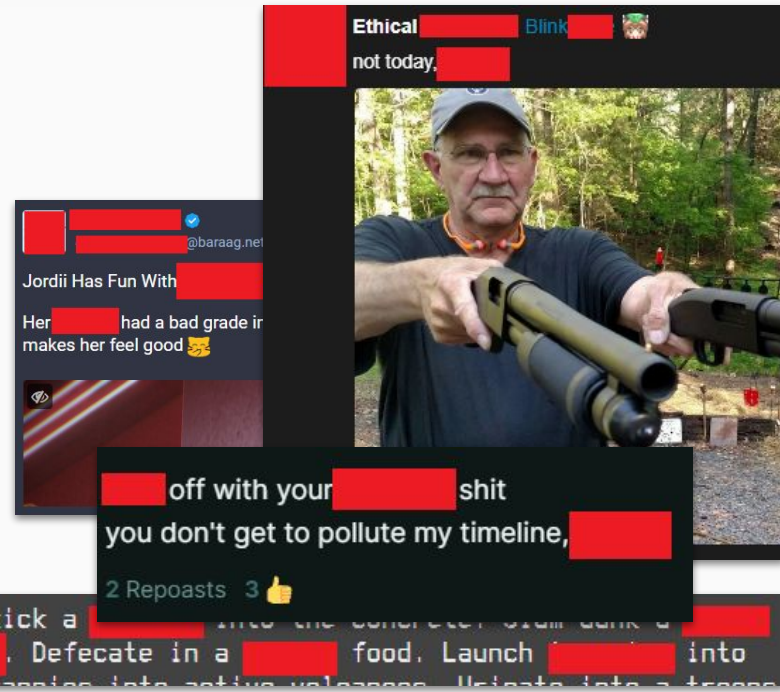


Fediverse Moderation



Marginalised voices and communities can find themselves on poorly resourced or under-informed services that do not have the resources to assure safety online.

Low barrier to entry means service providers can unwittingly expose new members to hate and abuse



Fediverse Moderation



Avoidable service shutdowns create negative outcomes for adoptees

New actors struggle to operate successfully with little regard for the basic requirements of managing UGC

Fragmentation of Codes of Conduct, moderation styles, creates uncertainty for the member base

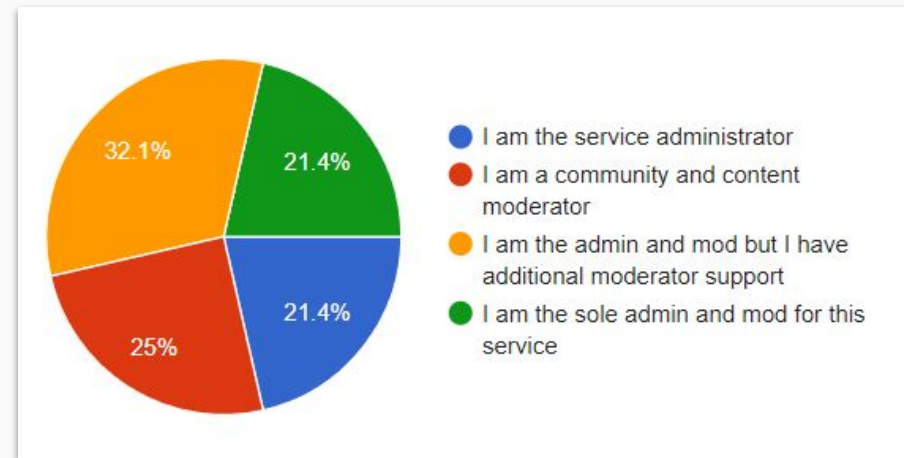
A collage of screenshots from Mastodon and news articles. The top left shows a post by Nathan (@nathan@mastodon.lol) stating 'Mastodon.lol will shut down on today.' and providing a link to migration docs. The top right shows a post by John Proffitt with the headline 'The Raspberry Pi social media incident shows some Mastodon admins need more moderation chops'. The middle section features a news article titled 'Financial Times Sets Up Mastodon Server, Realizes Laws Exist (Which It Was Already Subject To), Pulls Down Mastodon Server'. Below this is a dark-themed article titled 'mastodon.technology Shutdown' dated 'October 6, 2022'. The bottom right shows a post by Matthew Keys dated 'December 13, 2022' with the headline 'LGBT server on Mastodon blocks group for journalists'. The bottom left shows a post from 'Dear mastodon.tech' dated 'December 1, 2022' stating 'I have sad news that I have decided to shut down the mastodon.technology instance. In accordance with the Mastodon Server Covenant, the server will be shut down no earlier than December 1, 2022.'

Fediverse Moderation



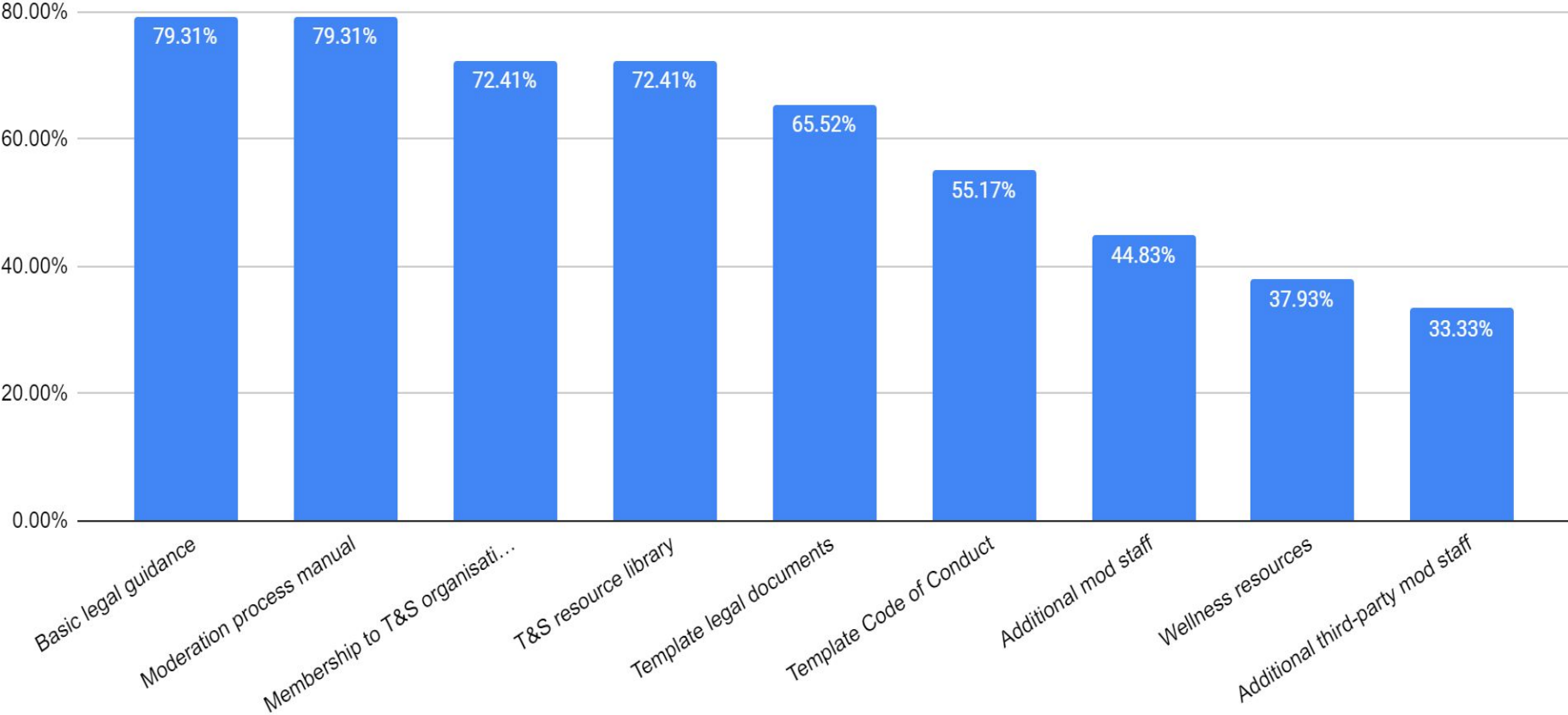
Two independent surveys of service administrators and moderator staff found the majority of respondents have concerns about illegal content, DMCA or other copyright issues, privacy laws, and liability issues.

Survey respondents account for ~5% of all Mastodon usage



[Academic Mastodon Survey, November 2022](https://academicmastodon.org/survey) :: [Digital Humanities @ Stanford](https://digitalhumanities.stanford.edu/) , and <https://drive.google.com/file/d/1wWFz1qSWbuLnmHIX8xGEiWEguru7qoUm/>

Percent of respondents affirming interest in accessing resources



Would you be interested in accessing the following resources?

Proposal



Create a non-profit Trust and Safety entity to support the growing Fediverse:

1. Establish a member-driven Community of Practice for moderators
2. Recruit and train moderators to offer Moderation-as-a-Service
3. Commission and distribute legal and regulatory guidance and templates

Three Pillars



- **Moderation-as-a-service**
 - Provide actionable resources and moderation at scale; from single event to full time moderation
 - Build sustainable pathways to compensating Fediverse moderators
 - Convene protocol, platform and dev community to build federated moderation software
- **Community of Practice**
 - Convene, support, and grow the moderator community
 - Provide communication tools (chat, discussion, wiki) and empanel working groups
 - Create a sliding-scale membership approach to encourage member support from broad base
 - Collect and disseminate academic research, best practice, extant knowledge base
- **Trade Representative**
 - Law enforcement / regulatory reporting / data privacy bodies; insurers; professional development and memberships

Goals



Bolster the moderation community with free resources and best practices for communities serving marginalised and minority voices

Engage the existing volunteer moderation community to ensure long standing principles and social norms are upheld

Provide membership to the existing volunteer moderator community at no cost

Recruit from the community to provide paid opportunities

Critical Need



Entities like the TSPA and DTSP have demonstrated an impactful need for convening and organising practitioners in the private social media sector.

National Information Sharing and Analysis Centers (ISACs) have demonstrated how sectors that encounter problems they cannot solve by themselves have convened under a shared mission with shared resources to cooperate and counter collective threats.

Mission



To serve as a valued resource for trust and safety in a complex social media landscape



Vision



IFTAS is a non-profit organization created to help like minded Fediverse members foster and preserve inclusive, civil discourse for the common good.



Based on Allan, B. (2008). Knowledge creation within a community of practice

Scope



As of April 2023:

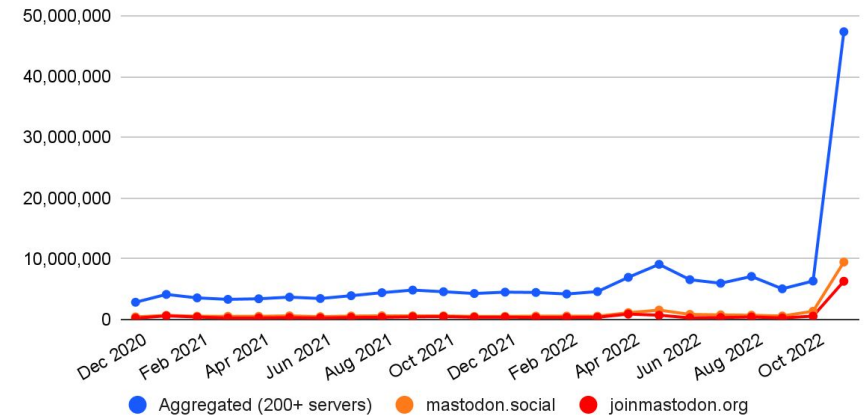
21,000 ActivityPub service providers
(Mastodon, Pixelfed, Peertube etc)

1B user-generated posts (text, image, video, audio)

9M member accounts

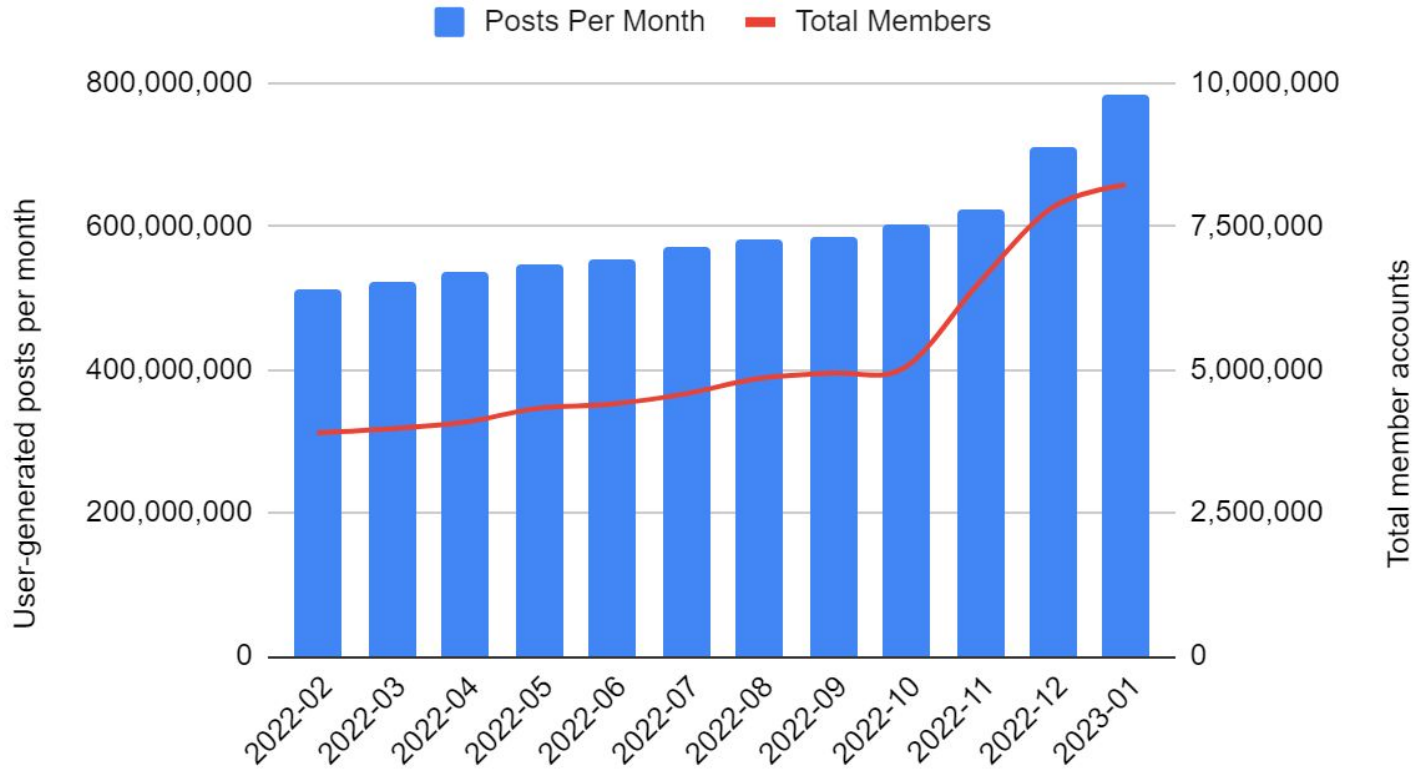
Mastodon jumping off points and big picture

Monthly Visits Desktop & Mobile Web Worldwide

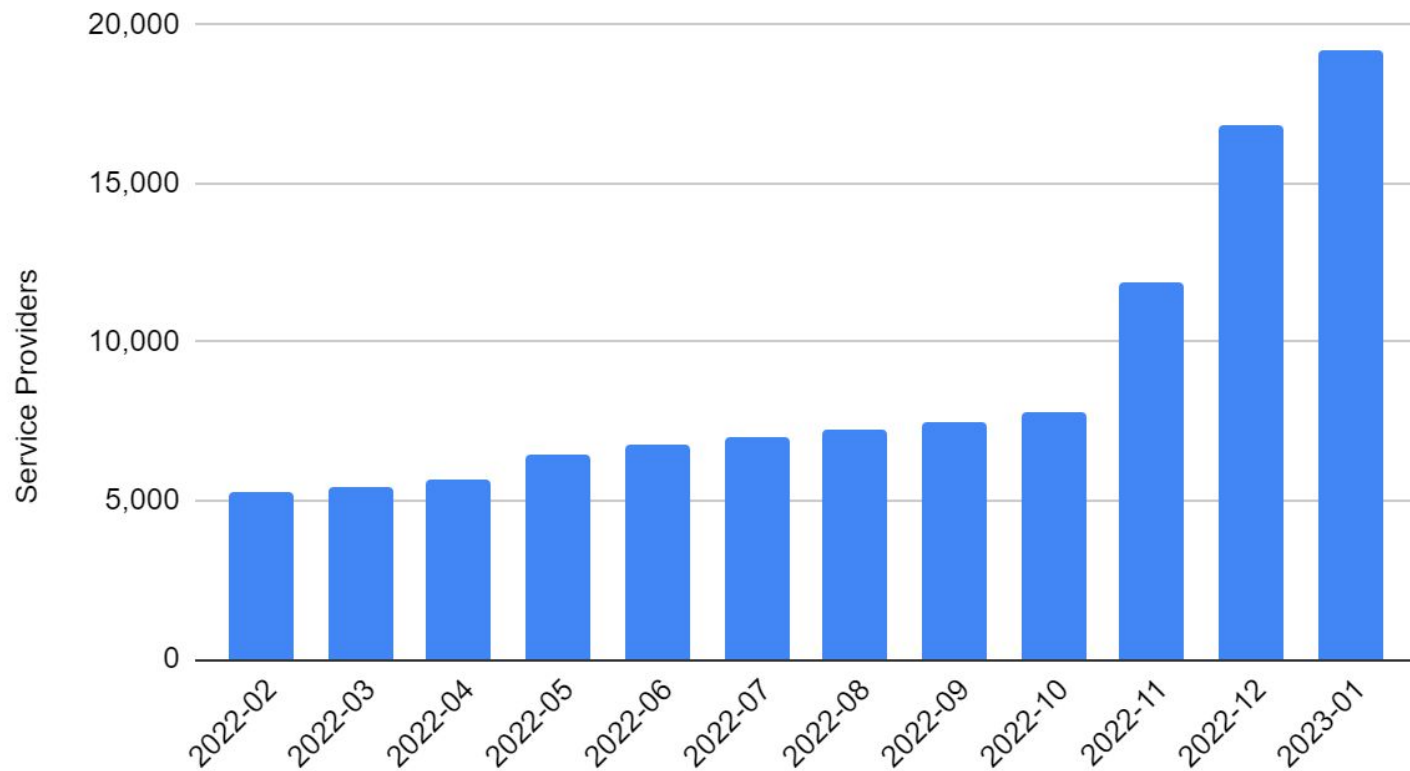


Source: <https://www.similarweb.com/blog/insights/mastodon-growth/>

Posts Per Month and Total Members



ActivityPub Service Providers



Sustainability Model



Initially funded by founding corporate members and non-profit investment, growing to self-sustainability through paid services, membership, and public donations.

Offer paid services to the large and growing service providers, ad hoc support to mid-size, enabling support and paid moderator opportunities.

Offer professional membership on a sliding scale.

Sustainability Model



Paid services to include:

- Moderation-as-a-Service
- First 90 T&S - blocklists, legalese, jurisdiction-aware guidance
- Ad hoc language support (content moderation for unsupported languages)
- Training and wellness
- Group liability insurance
- Funnel for mandatory reporting

Sustainability Model



Membership:

- Digital community of practice
- Wiki, forums and chat, directory of resources
- Pret-a-porter CoC, ToS, moderator agreements
- Updates on regulatory changes
- Access to self-guided training, wellness and resilience resources
- TSPA membership

Roadmap



- Gather support from IFTAS working group
- Gather support from active moderator community
- Recruit initial moderator team that can cover all timezones
- Offer basic moderation support to large servers in need of TZ coverage
- Commission basic legal guidance by major jurisdiction
- Investigate group liability insurance (US, CA, UK, EU, AU, NZ)
- Build directory of Fediverse-aware legal community

Team



- [Jaz-Michael King](#) - Digital communities and non-profit tech
- [Yoel Roth](#) - former head Twitter Trust and Safety, Fellow UC Berkeley
- [Shauna Gordon-McKeon](#) - Non-profit governance and community engagement
- [Grace L Williams](#) - Writer, Journalist
- [Damien Mar Chong](#) - Non-profit finance and business operations
- With support from advisors including J. Nathan Matias (CAT Lab @ Cornell), August Bournique (Bournique Law), David Slifka (BlueM)

Collaborate on moderation topics: Follow then mention [@moderation@venera.social](#)

Join the IFTAS planning group: <https://groups.google.com/a/toot.wales/g/iftas/>