Tips for Running a Mastodon Instance

Alex Stamos and Sara Shah
Stanford Internet Observatory
Last modified: October 17, 2023
## Contents

1 Mastodon and the Fediverse  
2 Running your own Mastodon Instance  
3 How Mastodon Works  
   3.1 Why Bad Content is Unavoidable (For Now)  
   3.2 Moderation on Mastodon  
4 Common Issue Areas  
   4.1 Child Safety  
   4.2 Copyright Issues and Data Privacy  
   4.3 Law Enforcement Interaction  
   4.4 Porn and Sex Work  
5 Tips and Tricks Checklist
These materials have been prepared for general informational purposes only, and are not legal advice. Information in this publication may not constitute the most up-to-date legal or other information.

1 Mastodon and the Fediverse

Mastodon has become a popular alternative social networking site to platforms like X (formerly Twitter). A decentralized social network where individual servers (or “instances”) are typically grouped by topic and run by individual users, groups, or organizations, Mastodon is part of the growing “fediverse,” or network of decentralized social media sites. Users can interact across sites powered by the fediverse’s ActivityPub protocol, which manages content, connects users across decentralized networks and is developed by a volunteer working group organized by the World Wide Web Consortium. Other decentralized social networking sites like Mastodon include newcomer Bluesky1 (a rival to the fediverse, whose architecture is incompatible2 with ActivityPub) and Meta-owned Threads,3 which has announced plans to join the fediverse.

Similar to other social media platforms, Mastodon shows a “Home” feed of posts from all accounts followed by a user, as well as two additional feeds: a “Local” timeline of public posts from all users on that instance, and a “Federated” timeline of public posts from any account in the fediverse (unless blocked by the home instance). An instance federates with another when a user on an instance follows an account on another instance, causing the second instance’s content to show up on the Federated timeline of every user on the first instance. Instances don’t have to federate, however. As an instance administrator, you can create a totally isolated community, or a completely integrated one.

Every instance on any given site in the fediverse has its own rules, policies, and terms of service. Users don’t join Mastodon like they would X, because Mastodon isn’t one central website; they must choose a specific instance to open an account under, and that instance becomes part of a user’s Mastodon identity—similar4 to selecting an email provider, such as Gmail or Outlook, with which to set up an email account. Further, exiting an instance isn’t penalized like it is on centralized platforms like Facebook or X, where users lose their followers and content if they leave the social media site. Mastodon users can leave and join another instance while retaining their followers. In this way, users have the power to choose what kind of community they want to be part of, and they aren’t punished for trying again.

There are analogues to most major centralized social media sites in the fediverse, all powered by the ActivityPub protocol. For example, if Mastodon is the fediverse version of X, PeerTube is fediverse YouTube and Pixelfed is comparable to Insta-

gram. Because the fediverse is interoperable, users on one site can seamlessly interact with users on another, creating an ecosystem where users aren’t locked into one site.

2 Running your own Mastodon Instance

First, read Mastodon’s basic guide\(^5\) to starting your own instance.

Running an instance on Mastodon is different from moderating a subreddit or other online forum. No Mastodon-specific terms of service or rules exist; every instance has its own rules. Instance administrators are in complete control of the guidelines they create for their community, and often enlist volunteer moderators to help enforce them. But the freedom to set poor standards comes at a price. The online harms plaguing centralized sites—including harassment and abuse, spam, and proliferation of child sexual exploitation material—also plague the fediverse, albeit differently. These abuses, already difficult to address on other social media sites, are trickier to combat in the fediverse writ large because there isn’t a central moderation function to catch\(^6\) them. Instead, individual Mastodon instance administrators have the ability to suspend users, tag or remove sensitive content, and block problematic instances from federating with their own. Users can add content warnings to their own material, block other users, and report posts to server administrators.

These tools aren’t necessarily enough to stop bad or illegal content from reaching an instance via federation. As you get started, familiarize yourself with the issues you might face in running and moderating your instance, the tools you have at your disposal to address them, and the laws that govern them. The largest user bases of Mastodon.social, the biggest Mastodon instance, are in the U.S., UK, and Europe; important laws to be aware of in each jurisdiction are flagged below.\(^7\)

3 How Mastodon Works

3.1 Why Bad Content is Unavoidable (For Now)

Mastodon is set up in a way that only allows users to see content from instances they already know exist. Mastodon feeds show content from your instance and the instances your instance federates with, and the search bar only returns content from the same; you can’t actually search all federated instances. This means that if another instance is disseminating illegal or disturbing content, you’d never know—unless, as an administrator, a user on your instance follows an account on the aforementioned instance, causing harmful content to show up on all of your users’ federated timelines and in relevant search results.

\(^5\) https://docs.joinmastodon.org/user/run-your‑own/#so‑you‑want‑to‑run‑your‑own‑mastodon‑server

\(^6\) https://purl.stanford.edu/vb515nd6874

\(^7\) For more comprehensive legal guides aimed at instance administrators, see https://www.eff.org/deeplinks/2022/12/user‑generated‑content‑and‑fediverse‑legal‑primer or https://denise.dreamwidth.org/91757.html.
Content from Mastodon timelines is cached locally on your server. Even if you’re planning to run a very small instance with under \( \approx 100 \) users, you will likely need a lot of storage to handle cached content. There are a variety of storage options available, ranging from widely-used commercial services to boutique services that cater to Mastodon administrators specifically.

Any harmful material that feeds into your users’ federated timeline will be cached locally on your server. The amount of time that content is kept before auto deletion differs for each storage service, and often differs between text posts and media. Depending on the jurisdiction, you might be legally obligated to report illegal material that shows up in your cache and retain it for a certain period of time. This could help dictate which storage service is best for your instance.

### 3.2 Moderation on Mastodon

Mastodon administrators are responsible for outlining the Terms of Service (including a Privacy Policy) and Server Rules that govern their instance. Mastodon also has an explainer detailing the moderation actions an administrator can take against individual users, other instances, and against spam.

As an administrator, you have several options to deal with problematic or harmful users on your instance. You can freeze a user’s account, maintaining the account's content but preventing the user from logging in and using Mastodon, or limit an account, hiding the account content from everyone on the instance but the account's followers (and those who search for it). Both of these actions are reversible. For repeat offenders or those meriting more drastic action, you can suspend an account, effectively deleting it. Account suspensions can be reversed within the first 30 days after a suspension, but all content will be wiped. You can also mark an account as sensitive, which will automatically add a “sensitive” content warning to any media the account posts. Remember that these actions only apply locally, meaning only within the confines of your instance; suspended accounts can re-spawn on other instances, but they’ll have to rebuild their following.

Mastodon also gives administrators powerful tools to reject content from or block other instances entirely. You can suspend (or block) another instance, which blocks all users and content from that instance from interacting with your own. Administrators have taken to using the #fediblock hashtag to publicly communicate which instances are causing harm to users. Many instances publish their blocklists, but some don’t; this mitigates against unsavory users actively searching for harmful material. You can reject media from another instance, which ensures no files or media from that instance are processed locally on your server. You can also limit a server—similar to limiting an individual account—which hides that instance's content from anyone who isn't following an account on it.

---

8. [https://docs.joinmastodon.org/admin/moderation/](https://docs.joinmastodon.org/admin/moderation/)
To combat spam, administrators can both block by email server\(^\text{10}\) en masse (if the email domains used ultimately resolve to a single email server IP) or block visitors by IP address using a firewall.

4 Common Issue Areas

4.1 Child Safety

Individual instances have a suite of tools available to try and stop child sexual abuse material (CSAM) from ending up on their instance. Most instances don’t want this content associated with them! CSAM is illegal in some form in almost every country around the world—certainly in the United States\(^\text{11}\)—and to guard against potential liability, server administrators are incentivized to be overly cautious in this space. But because there is no central function to catch it, and the fact that certain instances not only tolerate it but disseminate it, CSAM (and other harms) manifest as unique problems on the Fediverse in ways they wouldn’t in centralized systems.

Instances have the option to defederate with other instances entirely if they don’t want any interaction with a particular server. Large servers operating in areas like the U.S., UK, or EU tend to defederate from instances based in Japan, for example, out of fear that explicit illustrations (or explicit generated content) of minors, both legal and more normalized in Japan than in other areas, might make their way onto servers in jurisdictions where they are not culturally accepted and possibly illegal. If blocking another instance entirely feels too drastic, instance admins can use the “Reject Media” function to allow text content from an instance to populate on your server but block images. Mastodon.social, the biggest Mastodon instance, rejected media\(^\text{12}\) from a large Japanese instance known for spreading “lolicon,” or sexually explicit illustrations of minors, in 2017. While explicit images or illustrations are not necessarily CSAM, they are not always treated the same in some legal jurisdictions.

In the United States, as in most of the world, CSAM involving actual minors is unequivocally illegal.\(^\text{13}\) Artificially created images or visual depictions of child sexual abuse (i.e. not involving an actual minor) are also illegal\(^\text{14}\) in the United States, if considered “obscene.”\(^\text{15}\) Obscene material—regardless of if a real minor was used or depicted—is never protected by the First Amendment, and most sexually explicit material involving a child (whether real or virtual) is likely to be found obscene. Additionally, federal law also currently prohibits\(^\text{16}\) virtual depictions of CSAM that are “indistinguishable” from an actual minor, even if not obscene. However, the courts have not ruled definitively on the constitutionality

---

\(^{10}\) [https://docs.joinmastodon.org/admin/moderation/#blocking-by-e-mail-server](https://docs.joinmastodon.org/admin/moderation/#blocking-by-e-mail-server)

\(^{11}\) [https://www.law.cornell.edu/uscode/text/18/2252](https://www.law.cornell.edu/uscode/text/18/2252)

\(^{12}\) [https://mastodon.social/@Gargron/2650049](https://mastodon.social/@Gargron/2650049)

\(^{13}\) [https://www.oyez.org/cases/1981/81-81-55](https://www.oyez.org/cases/1981/81-81-55)

\(^{14}\) [https://www.law.cornell.edu/uscode/text/18/1466A](https://www.law.cornell.edu/uscode/text/18/1466A)

\(^{15}\) [https://www.oyez.org/cases/2001/00-795](https://www.oyez.org/cases/2001/00-795)

\(^{16}\) [https://www.law.cornell.edu/uscode/text/18/2256](https://www.law.cornell.edu/uscode/text/18/2256)
of this ban. Past cases suggest\(^{17}\) that content not involving an actual minor (i.e. virtual or computer-generated sexually explicit imagery of minors), and that is also not considered obscene for a variety of reasons, is protected under the First Amendment.

Any CSAM found on an instance (or in its cached material) based in the United States must be reported to the National Center for Missing and Exploited Children (NCMEC) CyberTipline\(^{18}\) and retained for 90 days thereafter, per the requirements outlined in 18 U.S. Code § 2258A.\(^{19}\) The material should also be made publicly inaccessible. Certain bills, like the recently introduced REPORT Act,\(^{20}\) would extend the retention period from 90 days to 1 year if passed by Congress.

While there is currently no mandatory reporting requirement for CSAM in the United Kingdom, the Online Safety Bill\(^{21}\)—expected to become law in 2023—will create one. The Bill includes provisions mandating providers remove and report CSAM to the National Crime Agency, alongside a number of other online safety\(^{22}\) provisions.

There is also no central CSAM reporting mechanism for providers in EU countries, and the legality of real CSAM versus depictions of CSAM varies by country. Providers can simply delete\(^{24}\) CSAM without reporting it. The EU itself acknowledges a heavy reliance on the United States and NCMEC to intake CSAM reports. Recently proposed EU legislation,\(^{25}\) if passed, would create an EU Centre dedicated to processing reports of child sexual abuse online, amongst other new stipulations.

So what can instance administrators do to protect their users (and themselves) from any potential liability related to accidentally interacting with CSAM online and related material? Services like Microsoft’s PhotoDNA\(^{26}\) or Thorn’s Safer\(^{27}\) use perceptual hashing to proactively scan sites for known CSAM and report images to NCMEC, but there is little available tooling to integrate these with federated services, requiring significant effort on the part of administrators.\(^{28}\)

Even if stock tooling existed to allow the integration of proactive scanning services with Mastodon, only administrators running large instances—where the risk of

\(^{17}\) https://www.oyez.org/cases/2001/00-795
\(^{18}\) https://www.missingkids.org/gethelpnow/cybertipline
\(^{19}\) https://www.law.cornell.edu/uscode/text/18/2258A
\(^{20}\) https://www.govinfo.gov/app/details/BILLS-118s474rs
\(^{21}\) https://bills.parliament.uk/publications/52368/documents/3841
\(^{22}\) https://cyber.fsi.stanford.edu/news/using-safety-design-address-online-harms
\(^{23}\) This post by Heather Burns theorizes how the Online Safety Bill might affect instance administrators: https://webdevlaw.uk/2022/11/21/a-quick-hypothetical-situation-or-your-crash-introduction-to-the-real-world/
\(^{26}\) https://www.microsoft.com/en-us/PhotoDNA
\(^{27}\) https://safer.io/how-it-works/
\(^{28}\) This SIO report on child safety in the fediverse advocates for further tooling to close this critical gap: https://purl.stanford.edu/vb515nd6874.
encountering CSAM is greater—would be likely to use them. Administrators of mid-sized instances with only a few hundred users may not be worried about the low probability of CSAM making its way onto their servers, especially if prospective members are screened and vetted. For now, the reject media function is a useful tool to moderate against any instance with dubious content without restricting text speech, and hiding your blocklist will help prevent users from intentionally seeking out harmful content.

Additionally, your privacy policy should explicitly state the minimum age requirement for users on your instance. Some state-level laws in the United States like CalOPPA\(^\text{29}\) require platforms hosting users from California to publicly post a privacy policy and actually adhere to it. The Children’s Online Privacy Protection Act (COPPA)\(^\text{30}\) imposes additional requirements at the federal level for any platform catering to children under the age of 13.

### 4.2 Copyright Issues and Data Privacy

What if someone on your instance posts copyrighted material? In the United States, the Digital Millennium Copyright Act (DMCA)\(^\text{31}\) provides “safe harbor,” or immunity, from copyright infringement claims for online service providers—like instance administrators—who follow a series of steps. These include registering a designated agent with the U.S. Copyright Office to handle copyright claims; publicly posting a “repeat infringer” policy alongside the Terms of Service, privacy policy, and moderation policy; and having a clear “notice and takedown” procedure to remove offending material when notified, providing the notice submitted meets all DMCA requirements, and informing the original poster that their content has been removed.

As an administrator, you are taking on the responsibility of complying with legal frameworks related to copyright infringement. If you're running a large instance or serious about growing your user base, you could consider retaining legal counsel in case you are enmeshed in legal obligations you did not anticipate.\(^\text{32}\)

Mastodon administrators will need to comply with General Data Protection Regulation (GDPR)\(^\text{33}\) requirements if their instances are based in the EU, have EU-based users, or federate with an EU-based instance.\(^\text{34}\) Generally, it should be clear to users who can see their personal data—as an instance administrator, you can see all of it, including users’ private messages with others on and off their

---

31. [https://www.law.cornell.edu/uscode/text/17/512](https://www.law.cornell.edu/uscode/text/17/512)
32. EFF has a more comprehensive overview of copyright liability for Mastodon instance administrators in the United States: [https://www.eff.org/deeplinks/2022/12/user-generated-content-and-fediverse-legal-primer](https://www.eff.org/deeplinks/2022/12/user-generated-content-and-fediverse-legal-primer)
33. [https://gdpr.eu/](https://gdpr.eu/)
34. This privacy policy generator allows you to generate a GDPR-compliant privacy policy for your instance based on eupolicy.social’s GDPR-compliant policy: [https://blog.riemann.cc/projects/mastodon-privacy-policy-generator/](https://blog.riemann.cc/projects/mastodon-privacy-policy-generator/)
The Digital Services Act (DSA), which service providers have until January 2024 to comply with, dictates different levels of compliance and transparency for different types of intermediary services operating in the EU. Small and micro enterprises are granted exemption from certain requirements.  

4.3 Law Enforcement Interaction

There are scenarios where U.S. law enforcement might reach out to you for data about a user on your instance that they cannot otherwise find publicly (i.e., direct messages). These scenarios might include receiving a search warrant, subpoena, or National Security Letter (NSL) from law enforcement.

In the United States, the Stored Communications Act (SCA) covers the voluntary and compelled disclosure of stored wire and electronic communications and transactional records held by third party internet service providers to the U.S. government. EFF has outlined restrictions on U.S. government access under the SCA, different situations in which law enforcement can request information, and what type of documentation they need to request it.

Countries often seek information from peer nations about citizens or individuals relevant to a criminal investigation. Many countries have Mutual Legal Assistance Treaties (MLATs) that dictate how their respective law enforcements exchange information, but response times for processing and compliance are often slow. The CLOUD Act, passed in 2018, created a new paradigm of international information sharing by modernizing cross-border law enforcement access to electronic communications held by private entities. The U.S.–UK Data Access Agreement (under the CLOUD Act) specifically allows service providers in the U.S. and UK to respond to law enforcement requests for electronic data from either nation without worrying about the complexity of cross-border data transfers and disclosures.

The chances that you encounter a law enforcement request related to a user on your instance are likely low. You may consider, however, adopting a Privacy Policy for your instance that covers if and/or when you would consider releasing

---

36. This primer from decoded.legal outlines how UK law might affect fediverse instances, including from a data protection standpoint: https://decoded.legal/blog/2022/11/notes-on-operating-fediverse-services-mastodon-pleroma-etc-from-an-english-law-point-of-view.
37. This post from Martin Husovec outlines how the DSA might apply to individual Mastodon instances: https://husovec.eu/2022/12/the-dsa-newsletter-2/.
38. https://www.law.cornell.edu/uscode/text/18/part-I/chapter-121
39. https://ilt.eff.org/Privacy__Stored_Communications_Act.html
42. https://crsreports.congress.gov/product/pdf/R/R45173
43. https://www.justice.gov/opa/pr/landmark-us-uk-data-access-agreement-enters-force
44. https://mastodon.social/privacy-policy
information about users to comply with the law. Mastodon.social’s Privacy Policy has been widely adopted by other instances with minor edits.

4.4 Porn and Sex Work

What if you encounter sexually explicit material on your instance? If posted by one of your users, you might choose to tag the account’s media as “sensitive” so others have to click to reveal the content. But the presence of sexually explicit content might not be the only time you encounter sexual material or conversations surrounding sex work on your instance, especially if it’s large. The 2018 passage of FOSTA-SESTA legislation in the United States, meant to curb online sex trafficking, fundamentally changed how online platforms address such content.

FOSTA-SESTA was passed as a carve out from the safe harbor that Section 230 of the Communications Decency Act allows providers, purportedly to crack down on sex trafficking. Previously granted general immunity from liability for user-generated content, providers are now held liable for any sex work ads (including consensual sex work) posted on their platforms in the United States. This has resulted in platforms more heavily moderating sexual content in order to ensure compliance—including by removing spaces for sex workers to congregate online, out of an abundance of caution. Switter, a Mastodon instance founded as a safe space for sex workers to discuss their work, was shut down by Cloudflare days after FOSTA-SESTA was signed into law. It was the fifth largest instance on Mastodon at the time. VerifyHim, a system enabling sex workers to access past reviews of potential clients and “Bad John” lists, was also shut down after the legislation was signed into law.

Many sex workers have vocalized that FOSTA-SESTA takedowns make their work more dangerous. Taking down such material may also contribute to the chilling effect First Amendment advocates are concerned about. As an instance administrator, you may be incentivized to take down any material you worry may violate FOSTA-SESTA. Ultimately, whether to leave it up or take it down depends on your risk tolerance. If you feel strongly, it may be beneficial to talk to legal counsel for advice about how best to comply with SESTA-FOSTA without adversely affecting the sex worker community.

5 Tips and Tricks Checklist

• Consider defederation with servers based in areas that have different cultural or legal views on sexually explicit material of minors (i.e. Japan).

45. https://survivorsagainstsesta.org/documentation/
46. https://www.law.cornell.edu/uscode/text/47/230
51. https://survivorsagainstsesta.org/documentation/
• Publish a Terms of Service (ToS) and Server Rules clearly outlining the moderation and speech restrictions you set for your instance.

• Copy blocklists from large instances with ToS/Server Rules that you like or are similar to yours, and hide them publicly so users can't search for bad instances.

• Publish a strong, GDPR-compliant Privacy Policy that includes a “repeat infringer” policy, notes on user data collection, and minimum user age requirements, and actually do what it says.

• Register with the U.S. Copyright Office to take advantage of safe harbor protections.

• Make liberal use of the reject media function for any picture-based content you want to restrict without limiting text speech from another instance.

• Consider in advance how you will address sexual or sex worker-driven content.

• Follow developments in how the DSA and UK Online Safety Bill might apply to the fediverse, as well as upcoming U.S. Supreme Court cases that might affect Section 230, which may impact your instance.

Acknowledgements

This guide would not have been possible without the feedback and contributions of Dan Bateyko, Elena Cryst, Gareth Fowler, Daphne Keller, John Perrino, Nate Persily, Riana Pfefferkorn, Chris Riley, Alan Rozenshtein, and David Thiel.

Thank you.
The Stanford Internet Observatory is a cross-disciplinary program of research, teaching and policy engagement for the study of abuse in current information technologies, with a focus on social media. The Stanford Internet Observatory was founded in 2019 to research the misuse of the internet to cause harm, formulate technical and policy responses, and teach the next generation how to avoid the mistakes of the past.