So You’re A Mastodon Moderator

CM_Noelle@mastodon.social, September 2018

Congratulations!

Either you’ve started your own instance, or someone trusts you enough to bring you on as a moderator on their instance. In either case, it’s cause for celebration!

Here’s what you’ll need to know in order to be an effective Mastodon moderator.

Your Tools

In your moderator account's Preferences page, you'll see a section called **Moderation**. If you’re a moderator and not an administrator, that gives you four sub-options: the **Mod Audit**, **Accounts**, **Reports**, and **Invites**.

Mod Audit

The Mod Audit shows you every action any moderator has taken, in reverse chronological order. It's handy to figure out who did what. **If you’re not an administrator who manages several subordinate moderators, you probably won’t need this section very much.** (If you are, this is great for working out who took action against which user, so you know who to talk to when someone complains.)

Invites

You almost certainly will not need to touch this section; it merely lists who’s used the Invite feature to let other people join your instance. If you notice that a lot of new users are being invited by a particular user, you can rescind their invitation link on this panel, but it’s pretty rare that you’ll need to do that.

Accounts

The Accounts page is where you can get information on specific accounts. You get different information depending on whether an account is **local** or **remote**. If you’re a moderator, hopefully you know that Mastodon is a distributed network, and that some users are local to **your** Mastodon instance, and others are on **other** Mastodon instances (or Mastodon-compatible instances, like GNU Social or Pleroma).

On each account's page you can also see how many reports they've filed and how many have been filed against them. On an individual account page, you can take several actions, listed in order:

- **Remove Avatar**: Use this if a person's avatar breaks your instance's rules.
- **Change Email**: **Almost never use this.** This changes the email address associated with an account. You should only do this at the explicit request of the user, and only when you can verify that you're talking to the person who actually owns the account.
- **Disable Login**: Logs the user out and prevents them from logging in again. Use this as a holding action. It's described in more detail below.
- **Silence**: Prevents a user's posts from reaching your instance's local timelines. Described in more detail below. You **almost never** want to silence a user from the account page.
- **Suspend**: **Never use this.** This destroys the account! It's described in more detail below. Only ever suspend a user as a very last resort, and **always** do it from a report, **never** from the account page.
- **Reset Password**: Sends the user a password-reset email. Use this if you believe an account has been compromised or if the user has requested it.
- **Leave Note**: This lets you leave a note on the account directly, not associated with a given report. This is handy for making sure that future moderators know your thoughts.

Reports

The Reports queue is where you'll spend most of your time. It's a list of reports made to or from your instance. By default you'll only see unresolved reports, but you can see resolved reports too. (There are a lot of those.)

You will receive reports from two sources: local and remote. **Local** reports are from users on your instance, and can be about anybody on the fediverse. (If one of your local users reports a user who's not on your instance, they can choose whether or not to send the report on to the remote user's instance. If you're not your instance's administrator, check with them to see whether or not you should assume a report has been forwarded.)

Remote reports are from users on **other** instances, and are **only** about users on your instance. Reports that come from other instances are anonymized. They do not have the username of the reporting user attached to them. This is a deliberate choice in order to protect the privacy of non-local users who want to report local users, in order to prevent bad actors who happen to be instance admins from harassing users who report people on their instances.

When you see a report from another instance, you will **only** see the name of the instance from which the report was issued. If you need more information:

- **If you're an administrator:** contact the remote instance's admin directly.
- **If you're not an administrator:** ask your admin to contact the remote instance's admin.

Their administrator will know more, but it's best to have a single point of contact.

When you view an individual report, you can see who filed the report and whom the report is about. If the report is from a remote instance, you'll just see the instance's name; if the report is about a remote user, you'll see everything but your options will be limited. You can also see the text of the report - if the reporting user filled that out; they don't always - and any toots associated with the report.

An individual report gives you seven potential actions. (There's an eighth, but you can't get to it here.)

- **Resolve report**: This will just mark the report resolved, with no note left or action taken against the reported user. Typically, instead of this, you want to use:
- **Resolve with note**: This resolves the report, but lets you leave a note saying why you resolved it. This is the most common action for many moderators.
- **Add note**: Use this if you want to leave a note but don't want to resolve the report outright. If you want to use one of the last two options, or if you want to leave the report open because you want a second opinion, here's where to leave your notes.

The behavior of the next set of tools varies depending on whether the user is local or remote.

FOR LOCAL USERS:

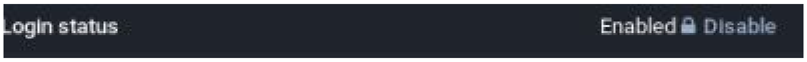
- **Add sensitive-content overlay to toot**: Select one or more toots using the checkboxes to the left, then click **Mark as sensitive** at the top of the list of toots. This will add the "Sensitive Content (18+)" overlay to any images attached to the toot. It **will not** add a content warning to the toot. (We don't currently have a way to do that.)
- **Delete individual toot:** Select one or more toots using the checkboxes to the left, then click **Delete** at the top of the list of toots. This will remove the toot from your database. This **does** request that other servers delete the toot as well, but they may not respect that request!

- **Silence user**: Prevents the user from posting to the local timeline and from appearing in the notifications of local users who aren't following the user. This is useful if someone on is harassing a local user, or if someone is, e.g., repeatedly posting untagged porn to the public timeline. Silencing someone doesn't prevent them from posting - but only the people who follow them will see it. **However**, their posts will still federate out normally.
- **Suspend user**: **Don't use this unless you mean it**. This will delete **all** the data from the user's account. All their followers, all the people they're following, all their posts, all their media. Reserve this for spammers and for people who are flagrantly breaking the rules.

FOR REMOTE USERS:

- **Add sensitive-content overlay to toot**: Select one or more toots using the checkboxes to the left, then click **Mark as sensitive** at the top of the list of toots. This will add the "Sensitive Content (18+)" overlay to any images attached to the toot. It **will not** add a content warning to the toot. (We don't currently have a way to do that.) **This only affects this particular copy of the toot.** Copies on other instances are unaffected.
- **Delete individual toot:** Select one or more toots using the checkboxes to the left, then click **Delete** at the top of the list of toots. This will remove the toot from our database, but it will not affect other instances' copies of the toot, and it can be re-added to the local instance if someone else on your instance boosts it or interacts with it.
- **Silence user**: This will prevent the user's posts from appearing on your instance's federated timeline and in the notifications of people on your instance who don't follow them. Anyone who follows the user can still see their posts and interact with them.
- **Suspend user**: This will forcibly remove the user from **all** of your timelines. It will force anyone on your instance who is following that user to unfollow them, and prevent them from ever showing up to anyone on your instance. Use this only when necessary, but you can be a little freer with it than with a local suspension, because other than unfollowing it's non-destructive.

If you're tempted to suspend a local user, **instead consider**: go to their account page. Under their email status you'll see a **Login Status** line, and at the far right you'll see "Disable".



Login status Enabled 🔒 Disable

Click "disable", and then add a note to the report and leave it open. This will prevent them from logging in and posting any further, but **won't** destroy their account like a suspension does. This lets you consider the issue, or confer with other moderators, and decide what to do about it.

A note: Silencing or suspending a user will also resolve all other reports about that user.

Quick Guidelines

If you can't figure out why someone was reported, if what was reported doesn't break your rules, or if the situation is better dealt with by one user blocking or muting the other, use Resolve With Note.

If the user's broken a rule but it seems like a one-off occurrence, use Resolve With Note and send them a DM.

If the user's broken a rule and it seems like it's going to be a pattern, silence them.

If the user's broken a major rule, disable login, add a note, and talk to your admin or co-moderators, or spend some time thinking about whether you really want to destroy their account.

If the user is flagrantly breaking your instance’s rules, suspend them. You may want to take a screenshot just so you have evidence if someone asks why they were banned.

Spam

A common position in the fediverse is that you should be a person, not a Brand. If the only thing an account exists for is to promote a business, that's spam. If the account is engaging with the user base and it's obvious there's a human being there who cares about more than just the brand, it might be best to leave them alone. And typically, an individual user self-promoting or mentioning a brand in their post is not spam. :)