

Trove user moderation process

Trove service Terms of use

<https://trove.nla.gov.au/about/policies/terms-use>

Background

Trove is a space where any member of the public can contribute and enrich the service through the creation of comments, tags or lists, or by correcting text. These postings are generally unmoderated by Trove Support Team, as they are provided in good faith by patrons keen to share their own insights gained from extensive research.

From time to time a user will abuse this privilege, using Trove's annotation features to make other users unwelcome in Trove. In the past this has taken the form of harassment of specific users, defamation of public figures or agitation around politically sensitive issues. The changes to Trove in 2020 reduced the ability for users harass other users

Reporting by other patrons brings these incidents to the Trove's attention. On investigation, some postings will directly contravene the Trove Terms of Use. As Trove's content, user base and brand recognition grows, so too do incidents of malicious postings. Trove has a staged warnings approach to dealing with these postings.

Other types of behaviour have also been reported by patrons including trivial edits with the purpose of gaining notoriety, using other Trove functionality to add spam content/promote their own cause and the use of profanity in usernames.

Moderation Procedure

Where the offending comments, tags, lists or text corrections are made by authenticated users, they will be moderated as a 'malicious' offence or an 'undetermined' offence.

Characteristics of 'malicious' offences:

- The offence is explicitly unlawful.
- The offence may contain advertising, promotion or solicitation where the user has not contributed to Trove in any other way.
- To exploit the hall of fame rankings
- Use of profanity in Trove username

Characteristics of 'undetermined' offences:

- It is likely that the user does not know they are contravening the Trove terms of use.

Undetermined offences

For 'undetermined' offences the following staged warnings approach will be taken. For example, deleting whole lines and then adding new lines with the corrected text?

Warning 1	Offending content is removed. Warning is provided to user.	Warning advises user of Trove's policy with the eventual outcome of de-activating the user account. Advises that this is the user's first warning, and include details about the offence,
-----------	---	---

		and a link to the Terms of Use and which section they have contravened.
Warning 2	Offending content is removed. Warning is provided to user.	Warning reminds user of Trove's policy with the eventual outcome of de-activating the user account. Advises that this is the user's second warning. Includes details about the first offence and the current offence, and a link to the Terms of Use and which section they have contravened.
Notification 3	Offending content is removed. Final advice is provided to user. User account is de-activated.	Advise that this is the user's third warning and that their account will now be revoked. Includes details about the first two offences and the current offence, and a link to the Terms of Use and which section they have contravened.

Malicious Offences

'Malicious' offences will not be given the first two warnings, but will immediately receive the final notification, and the user account is to be deactivated.

Where the user has not logged in and the offending comments, tags or text corrections are from an anonymous user, there is no method to get in contact with the creator. Without contact details, it is not possible to implement the three-strikes procedure. Instead, the offending comments or tags will be directly removed by Trove Support Team. Text corrections will be rolled back by Trove Support Team, where possible, or by Digital. All moderation actions will be recorded in the user moderation tracking document kept in HPRM.

Roles

- RefTracker monitor – Identify and investigate postings that contravene the Terms of Use and report them to the Program Manager, Partner Liaison. Once authorisation has been obtained, send standard communications to the user. Update the user moderation document with the number of warnings the user has received and relevant details of each offence. Record all correspondence in HPRM.
- Program Manager, Partner Liaison – investigates and authorise the sending of first warning
- Assistant-Director, Partnerships – authorise the sending of second warning messages. When notification of a third action is received, obtain consent from the Director, Partnerships to ban the user account.
- **Director, Partnerships – approve the de-activation of a user account on the third warning.**

Communication procedure

Communication will be through an email to the account's associated email address if available. Use of this generic contact method will ensure that Trove Support Team are not directly identified, and that they are not made targets of abuse or harassment by disgruntled users.

For email messages

- Use RefTracker to send the message.
- Record the RefTracker number in the user moderation document.

First warning

Dear Trove user [Trove username],

Trove Management has received a report that actions carried out on the Trove website attributed to the username [Trove username] have contravened the Trove [Terms of use](#). On investigation, Trove Support Team have determined that these actions are in contravention of the Terms.

The actions in question are [\[summary of actions\]](#)

The Terms of use for the Trove service require that [\[summary of terms of Use section that has been contravened\]](#). The Terms of use are available in full here:

<http://trove.nla.gov.au/general/termsfuse>

[\[Summary of actions taken – e.g. “The offending content has been removed...”\]](#)

This is your first warning, if Trove receives another report of actions attributed to the username [Trove username] that contravene the Trove Terms of Use, you will be issued a second warning. At the third warning you will receive advice that the user account [Trove username] is being revoked, and all content that you contributed that contravenes the Trove Terms of Use will be removed from Trove; and the associated email address de-activated in Trove.

To avoid a second warning you should review and moderate past actions attributed to the username [Trove username] to ensure they comply with the Trove Terms of use.

This incident is covered by Trove’s Terms of Use, which give us the right to withdraw or suspend the use of Trove by any person who uses Trove in breach of these terms and conditions.

To contact Trove Management about this matter fill out the [Contact form](#) or use the [Contact us](#) button at the top of all web pages on the Trove website. Please note this TroveSupport account is not monitored.

Thank you for your compliance and cooperation with this matter.

Trove Management

Second warning

Dear Trove user [Trove username],

Trove Management has received a report that actions carried out on the Trove website attributed to the username [Trove username] have contravened the Trove [Terms of use](#). On investigation, Trove Support Team have determined that these actions are in contravention of the Terms.

The actions in question are [\[summary of actions\]](#)

The Terms of use for the Trove service require that [\[summary of terms of Use section that has been contravened\]](#). The Terms of use are available in full here:

<http://trove.nla.gov.au/general/termsfuse>

[\[Summary of actions taken – e.g. “The offending content has been removed...”\]](#)

This is your second warning, you were contacted about your first warning on [date]. If Trove receives another report of actions attributed to the username [Trove username] that contravene the Trove Terms of use, you will be issued a third and final notification. At the third notification the user

account [Trove username] will be revoked, all content you contributed that contravenes the Trove Terms of Use will be removed from Trove; and the associated email address de-activated in Trove.

To avoid a third notification you should review and moderate past actions attributed to the username [Trove username] to ensure they comply with the Trove Terms of Use.

This incident is covered by Trove's Terms of Use, which give us the right to withdraw or suspend the use of Trove by any person who uses Trove in breach of these terms and conditions.

To contact Trove Management about this matter fill out the [Contact form](#) or use the [Contact us](#) button at the top of all web pages on the Trove website. Please note this TroveSupport account is not monitored.

Thank you for your compliance and cooperation with this matter.

Trove Management

Third de-activation advice message

Dear Trove user [Trove username],

Trove Management has received a report that actions carried out on the Trove website attributed to the username [Trove username] have contravened the Trove [Terms of use](#). On investigation, Trove Support Team have determined that these actions are in contravention of the Terms.

The actions in question are [\[summary of actions\]](#)

The Terms of Use for the Trove service require that [\[summary of terms of Use section that has been contravened\]](#). The Terms of use are available in full here:

<http://trove.nla.gov.au/general/termsfuse>

[\[Summary of actions taken – e.g. “The offending content has been removed...”\]](#)

This incident is covered by Trove's Terms of Use, which give us the right to withdraw or suspend the use of Trove by any person who uses Trove in breach of these terms and conditions. This is your final notification. You were contacted about your first warning on [\[date\]](#) and your second warning on [\[date\]](#). As this is the third notification the user account [Trove username] will be revoked in 7 days. All content you contributed that contravenes the Trove Terms of Use will be removed from Trove; and the associated email address de-activated in Trove.

To contact Trove Management about this matter fill out the [Contact form](#) or use the [Contact us](#) button at the top of all web pages on the Trove website. Please note this TroveSupport account is not monitored.

Trove Management

Other types of activities requiring moderation

Gaming the system (hall of fame ranking)

Multiple corrections

Users have been able to exploit the hall of fame rankings by creating a text correction with many lines in one window, then making a completely different set of corrections with many lines in another window, then, in a sequence clicking save on one, and then save on the other. Both register as valid text corrections when saved in sequence.

To slow this down a rate limit on text corrections has been added. Users get the ability to perform 1 instantaneous text correction, each correction to the same page should then take 4 seconds to do. If the user has 5 text corrections waiting to be submitted in the queue they will be disregarded.

Email reports about text correction activity

Digital have implemented a report function in Trove where each Monday, Wednesday and Friday Trove provides 2 reports identifying possible instances where user may have exploited saving corrects on multiple screens or there may be some dodgy text corrections eg only added text. This report is sent to trove@nla.gov.au.

“White space” corrections

An old version of Trove had different white space rules. It added non-breaking spaces at the end of each line (). Users are exploiting the change of white space rules to get credit for more lines than correct by entering () as the search term in the newspapers & gazettes advanced search and selecting 1000+Words as the word count. Then locate an article where the last text correction was older than 5 years. Make a single character correction in the article and view the result in the corrections history.

Possible resolutions for consideration by the Business area are:

1. Change newspapers to disregard the nbsp lines when an article is edited
2. Bulk update all legacy text corrections to remove the unnecessary white spaces
3. Prevent users searching for non-breaking spaces, so they can't find articles to correct. If it's done by Ops, in the way new Trove works it will show up as broken searches in the discovery user interface.

Use of profanity in Usernames

When setting up new Trove accounts users are able to select their own usernames. Some users have been using profanity. When identified the process is to request Digital to :

- Deactivate the user account
- Rollback any text corrections
- Issue a warning

Digital have implemented a report function in Trove where each Monday, Wednesday and Friday a list of all new users in the previous 48 hours is sent to trove@nla.gov.au for Trove Support Team to review and identify any usernames that cause concern.

Appendix 1

This list is non-exhaustive and many of the actions will be committed by the same user.

- We need a moderation tool that captures moderation deletions and writes out to a database table for reporting.
- Capability to see all user generated content for one account and flag content for action

User Behaviour	Features	Examples	Current Resolution	Suggested Resolution
Interpersonal abuse	Corrections, tags, comments, lists, flickr		Remove the text. Issue a warning.	Remove the text. Issue a warning. User timeout for a specific period
Threats of violence	Corrections, tags, comments, lists, flickr		NIL	Refer to legal for further advice and legal obligations.
Defamatory posts	Corrections, tags, comments, lists, flickr		Remove the text. Issue a warning.	Remove the text. Issue a warning. Suspend user access to specific functions (tags, comments etc). User timeout for a specific period
Deleting content	Text correction	https://trove.nla.gov.au/newspaper/article/142416719	Rollback Issue a warning	User timeout for a specific period Ignore mass text deletions (deleting characters without replacements) Remove anonymous corrections (50/50) Limiting functions for anonymous users (change characters, cannot delete or remove lines)

User Behaviour	Features	Examples	Current Resolution	Suggested Resolution
Stalking other users	Corrections, tags, comments, lists	https://trove.nla.gov.au/newspaper/article/142416719	Issue a warning. Remove text.	Remove the text. Issue a warning. Recommend make profile private
Vandalising text	Corrections	https://trove.nla.gov.au/newspaper/article/184292530	Rollback Issue a warning (known user)	Ignore mass text deletions (deleting characters without replacements) See suspect corrections button (based on list of words/ certain time)
Historical Revisionism	Comments/Tags	https://trove.nla.gov.au/newspaper/article/64634025	Contextual analysis	When this falls into other categories take action (for example defamatory)
Breach of copyright	Comment/tags		Refer to takedown page	Refer to takedown page
Spamming	Corrections, tags, comments, lists, flickr		Remove text. Issue a warning.	Ban user.
Unsolicited contact	Work emails		Issue warning when evidence of complaint	Refer users to Trove contact us page Recommend users think about safety online (Outreach)
Irrelevant comments	Comments, tags		Inoffensive leave it	Inoffensive leave it
Text correcting outside of rules	Corrections		Contact user to refer them to guidelines	Contact user to refer them to guidelines

User Behaviour	Features	Examples	Current Resolution	Suggested Resolution
Profanity	Corrections, tags, comments, lists, flickr, usernames		Rollback Delete text Issue a warning (known user)	Rollback Delete text Issue a warning (known user)
Trolling user generated content	Lists, comments	https://trove.nla.gov.au/list?id=62231	Delete text on request	Give people the ability to remove comments on lists Get rid of all star ratings
Gaming the system (whitespaces, punctuation, capitalization, single letter edits)	Corrections		NIL	Further refinement of text correction counts Masking usernames You have corrected this article in search results