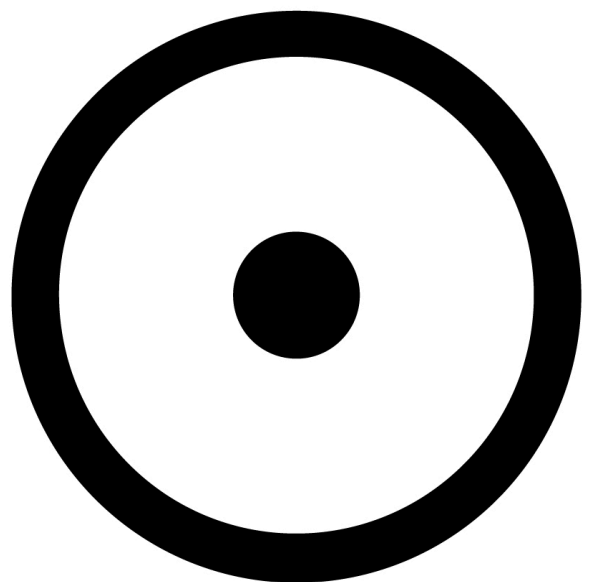


# Commons Image Moderation: Administrator Interviews

This research examines the experience moderating content on Wikimedia Commons from the perspective of extended rights users, Administrators, who perform deletions. It focuses on their workflows, challenges, perceptions of users, and ideas for improving the upload process in order to reduce the number of deletions.

By Bethany Gerdemann  
*Design Research, Wikimedia Foundation*



|                                                |    |
|------------------------------------------------|----|
| <b>Executive Summary</b>                       | 3  |
| <b>Background</b>                              | 5  |
| Methodology                                    | 6  |
| Participant Profiles                           | 6  |
| <b>Current Tools and Workflows</b>             | 6  |
| Process for Discovering Requests               | 6  |
| Case Limitations                               | 7  |
| <b>Process for Reviewing Cases</b>             | 7  |
| Speedy Deletion Requests                       | 7  |
| Regular Deletion Requests                      | 8  |
| Tools                                          | 9  |
| Beyond Moderation                              | 10 |
| <b>Language</b>                                | 10 |
| <b>Deletion Requests: Types and Difficulty</b> | 11 |
| Categories of Deletion Requests                | 11 |
| Content of Deletion Requests                   | 11 |
| Difficult Cases to Resolve                     | 12 |
| Admin Decision Making and Guidance             | 13 |
| <b>Challenges and Burdens</b>                  | 13 |
| Workflow Challenges                            | 13 |
| Organizational Challenges                      | 15 |
| <b>Upload Process</b>                          | 17 |
| Few Barriers to Uploading Images               | 18 |
| VisualEditor Upload                            | 18 |
| Learn                                          | 19 |
| Preselections                                  | 20 |
| Release Rights                                 | 20 |
| Order of Options                               | 21 |
| Interrupt the Process                          | 22 |
| Licenses                                       | 23 |
| Description                                    | 24 |
| Mass Upload & Flickr                           | 25 |
| Releasing Rights (VRT)                         | 25 |
| Content on Commons                             | 26 |
| <b>Uploaders: Perceptions and Interactions</b> | 27 |
| Interactions                                   | 27 |
| User disputes                                  | 27 |
| Copyright                                      | 28 |
| <b>Participant Suggestions</b>                 | 28 |
| Changes to cross-wiki uploads                  | 28 |
| Changes to Wikipedia upload tool               | 29 |
| Changes to UploadWizard                        | 29 |
| Changes to all upload tools                    | 29 |
| <b>Conclusion &amp; Acknowledgements</b>       | 31 |

**C**ommons is a media repository that hosts content that can be freely used by anyone for any purpose.

## Executive Summary

What is the moderator experience of reviewing deletion requests and deleting prohibited content on Commons? What are their tools, workflows, and considerations they use to handle deletion requests? What are the challenges and burdens they face? What are their ideas for reducing deletions?

### Tools and Workflows

Depending on interest, participants begin their work either on their watchlist, the speedy deletions page, or the regular deletions page. Some participants focus on speedy deletions and others on regular deletions. All gravitate to cases within their domain of expertise and actively avoid cases that cause stress.

Speedy deletions are clear and uncontroversial violations of Commons rules, most of which are either blatant copyright violations, personal (out of scope) images, advertisements, and duplicates. Administrators can quickly close these cases by viewing the reason for nomination and some other easily accessible file data. In regular deletion cases, admins rely heavily on the discussion for the case. They review the file data, although not all data are relevant for each case, review the uploader history to give context to the current case, and when applicable, review the image's associated Wikipedia article. For more complex cases, such as Freedom of Panorama or questionable copyright cases, admins will delve into more in-depth research and investigation.

Google Lens and TinEye are used universally to search the internet for earlier instances of the copyrighted images uploaded to Commons. Participants have also customized their toolboxes and streamlined their workflows with gadgets and third party tools, but each person's set up is slightly different. All participant admins remarked on the insufficiency of the current tools to perform deletion requests efficiently.

### Challenges and Burdens

Each administrator engages with the process according to their time and interest. There are a number of challenges and annoyances, but administrators take pleasure in performing this job and serving the mission of Commons.

All participants surfaced many of the same challenges, although each had a different way of framing each issue. Summarizations of the most frequently mentioned challenges are as follows:

- UI improvements, such as organizing Deletion Requests by topic or country, tool development to streamline actions, and support from the WMF to maintain essential tools and gadgets would vastly improve the work and experience of Administrators on Commons.

- There are not enough admins and functionaries on Commons to handle the immense workload, which extends beyond deleting files, and includes discussion, extensive research, workflow and tool management, interacting with users, engaging in governance, and more.
- Due to the ease of uploading ineligible and poorly described files to Commons, hundreds of thousands are untagged, unpatrolled, and undiscoverable. These images are unusable at best, and at worst, they are harmful or illegal, and degrade the integrity of Commons.
- A notably large proportion of deletion requests are cross-wiki uploads. They are most often copyright violations of various sorts, personal images, and spam. Many are uploaded to articles that also need to be deleted by local wiki administrators, duplicating the moderation work for a single contribution. The upload process on many wikis encourages uploaders to select “own work”, leading to many copyright violations.
- New users are the source of most of the problematic uploads to Commons, either directly through the UploadWizard or indirectly through the local wiki upload tool. All administrators favor some upload restrictions on new users, though specific suggestions vary widely.
- Most new users are unaware of and/or apathetic to the rules of Commons. Many are ignorant about the concept of copyright and file licensing. Some users are teachable. Many *good faith* users start their contribution journey by uploading ineligible and poor quality files and learning the rules through deletions and interactions with moderators.

## Suggestions for Reducing Deletions

Participants offered many specific suggestions for reducing deletions, which ranged from instituting earned user rights to UI and engineering upgrades in the upload tools. The following are the most widely offered suggestions:

- Use filters and machine detection to prevent ineligible images from being uploaded.
- Redesign the Wikipedia upload tool in order to increase awareness of the upload rules.
- Add complexity and hard stops to the upload process to create awareness of what belongs on Commons.
- Eliminate preselected options and other built-in nudges from the upload process to force the uploader to make conscious selections.
- Reduce and clarify license options which are numerous and confusing to average users.
- Reorder selection fields so that the most relevant option, or most important fields, comes first.
- Apply limitations on new user uploads.

## Background

The Wikimedia Foundation’s 2022-23 Annual Plan pledges to “Deepen our Commitment to Knowledge as a Service, starting with Wikimedia Commons and Wikidata”.<sup>1</sup> In April of 2022, the Foundation conducted a

---

<sup>1</sup> [https://meta.wikimedia.org/wiki/Wikimedia\\_Foundation\\_Annual\\_Plan/2022-2023](https://meta.wikimedia.org/wiki/Wikimedia_Foundation_Annual_Plan/2022-2023)

series of community conversations to elicit feedback on the plan. Conversations with Commons and Wikidata contributors surfaced the need for increased support for Wikimedia Common, specifically technological and functional support for upload and other tools.<sup>2</sup>

With this directive, the Structured Content team initiated a research request with four goals:

Goal 1: Understand the current user behavior in the image upload process for specifying release rights. This would help in identifying opportunities in the UploadWizard to prevent users from uploading content that violates the policies on Commons.

Goal 2: User testing for the new concepts and chosen solutions.

Goal 3: Understand the moderator experience of reviewing deletion requests and deleting unsuitable content or copyright violations on Commons

Goal 4: Understand current tools and workflows that moderators use in moderating content.

This research will be applied towards a series of improvements to the UploadWizard with the goals of:

- Improving the Commons upload experience for users.
- Decreasing the burden for contributors with extended rights by minimizing the possibility of uploading media that might trigger a deletion request.<sup>3</sup>

The focus areas of this project are enhancing user understanding of permitted content, clarifying the correct file data needed for upload, and simplifying licenses, thereby reducing the number of deletion requests that create a long backlog of work for volunteer administrators.

The first goal was achieved in Q4 of 2023 with a research project that focused on Spanish and English speaking users that recently opened Commons accounts. The third and fourth goals are addressed in the current report. The project addressing goal two began with the insights from the new Commons user research, and progressed concurrently with moderator research. The administrator interviews for this research took place over two weeks, while summaries were provided immediately following each interview, and were used to inform design concepts in goal two.

## Methodology

This research consisted of 90 minute in-depth interviews conducted remotely using video conferencing. In the interviews, participants were asked about work and hobbies, their involvement in the Wikimedia projects, and their involvement in Commons. Next participants were asked to share their screen as they walked through a typical speedy and regular deletion task. They demonstrated the processes along with the preferred tools, and highlighted some of the pain points they encounter in regular tasks. In some cases,

---

<sup>2</sup> [https://meta.wikimedia.org/wiki/Wikimedia\\_Foundation\\_Annual\\_Plan/2022-2023/Feedback\\_Appendix](https://meta.wikimedia.org/wiki/Wikimedia_Foundation_Annual_Plan/2022-2023/Feedback_Appendix)

<sup>3</sup> [https://commons.wikimedia.org/wiki/Commons:WMF\\_support\\_for\\_Commons/Upload\\_Wizard\\_Improvements](https://commons.wikimedia.org/wiki/Commons:WMF_support_for_Commons/Upload_Wizard_Improvements)

participants reviewed the UploadWizard to point out problems with the design and engineering that contribute to ineligible file uploads.

## Participant Profiles

- Male: 14  
Female: 1
- Canada  
France  
Germany  
India  
Italy  
Latvia  
Russia  
Netherlands  
Poland  
Turkey  
USA
- Dutch  
English  
French  
German  
Italian  
Polish  
Punjabi  
Russian  
Telugu  
Turkish
- Range of time contributing to Wikimedia projects is 3 to 20 years.
- Range of tenure as admin on Commons is less than 1 to 10 years.

## Current Tools and Workflows

### Process for Discovering Requests

Admins typically begin a moderation session in one of three places, depending on their personal interests and available time. 1. [Commons: Deletion requests](#) 2. A subcategory of the [Candidates for Speedy Deletion](#), such as [Copyright Violations](#) or [Advertisements](#). 3. Their personal watchlists, composed of images that the admin has contributed or that they have moderated and may be a target of ongoing vandalism.

Regular deletion candidates are organized chronologically. More recent dates contain the spectrum of difficulty, including easier to resolve and more difficult to resolve. As the easier cases are closed, the more difficult cases are discussed or await comments from persons with specialized knowledge. Cases that persist in the backlog are those that are most complex, controversial, or require specialized knowledge of copyright law in specific countries or specific language skills.

*Older cases in the regular deletion queue tend to be more trainwrecks.<sup>4</sup> -P10*

All admins focus their time and attention on what suits their interests or skill sets. Admins with fluency in non-English languages will look for cases that require their language skills.

*I am the first [language]-speaking admin on Commons. When I joined, there were 500 [language] files in the backlog. I have been working to close each of those cases. -P11*

---

<sup>4</sup> *Trainwreck*: A batch of files that are nominated for deletion at once, making the process for moderating each time consuming and tedious, as there is no function to process the batch with a single action. Typically each file requires debate and there are no easy answers.

Admins who wish to close cases quickly will start with recent cases, and those who wish to engage with complex research will look at older cases. This choice is made according to mood at the time of a moderating session. Admins will scan the list of deletion requests and look for low hanging fruit, cases that peak their interest, or cases that require their language skills or knowledge of copyright law in a particular country. All admins have their preferences for which types of cases to take and which they will avoid.

## Case Limitations

Every participant expressed their personal limitations for which cases to resolve. Every admin skips cases outside of their scope of knowledge, which will invariably involve copyright law in countries outside of their knowledge. Additionally, they actively avoid cases that cause too much stress or distress. Such cases mentioned are those that involve threshold of originality, copyright violations for useful but impermissible image files, or controversial decisions that may ignite outrage.

*Many images on Commons are absolutely unnecessary although completely legal and formally within the scope of Commons, but still nobody needs them. In many cases, we have images that are needed but unfortunately are not legal. So the image of this monument is in need on some local Wiki; it's in use in several local wikis, but unfortunately we cannot keep it. I prefer not to remove these images myself. -P4*

*I won't touch logos. It's difficult to determine which logos are simple enough to be considered free license, and then I also have to check the laws for the file's country to determine their rules for Threshold of Originality. -P11*

## Process for Reviewing Cases

### Speedy Deletion Requests

By the nature of their nomination to the speedy category, admins can quickly process these cases. The justification is clear and uncontroversial. The evidence is presented in the nomination. Admins quickly check the file and the evidence provided, then delete. If there is any question as to whether the justification is valid, the Admin will convert the request to a regular deletion case to elicit further discussion.

### Regular Deletion Requests

After selecting a file to review, admins will begin by looking at the reason cited for deletion and any discussion associated with the request to familiarize themselves with the case. Then they will check the file data by clicking on the file. All file data provided are used to judge cases, however not all data are relevant for each case. Below are examples of how each piece of data are used in specific case types:

- File name - Descriptive file names indicate the file was uploaded in good faith.
- Author - This field often is often the origin of copyright violations. When an uploader claims "own work" instead of the true author, it automatically creates a copyright violation. However, the file is

often under copyright and the work of another author, thus there are two copyright violations to process. Both of the following cases are common in the DR queue, and both require investigation:

- ◆ If the file has a free license, then the file can remain but the author should be changed.
  - ◆ If the author is correctly stated but the image is under copyright, then the file will be removed.
- Source - Uploaders often provide the url of the original image, which can make determining copyright violations easier.
- Description - Often the description contains the file author, while the author field contains the name of the uploader (“own work”). For Commons scope or Freedom of Panorama cases, the description may be helpful in determining the outcome of the case. Good faith contributions often contain a detailed description.
- Licensing - Admins will verify the license indicated is the true license for the file in cases of copyright violations.
- Categories - A large proportion of legitimate files are filed under categories. No categories listed is often an indication that the file is not legitimate.
- Exif data - Most legitimate photographs will contain Exif data generated by the camera. Stripped Exif data indicates the file has been taken from the internet. Additionally, images taken from Instagram or Facebook will have markers in the Exif data. In these cases, the image may be retained, but the author may need to release rights through the VRT process.
- Size of image - Small image sizes indicate the file was taken off the internet and uploaded to Commons. Legitimate files tend to be larger.
- File history
- ◆ Upload date - Used to determine how long the image has been on Commons, the urgency of the case, or to verify the file was uploaded to Commons after it appeared elsewhere on the internet in copyright violation cases.
  - ◆ Comment - This shows from where the file was uploaded; while editing a specific Wikipedia article, UploadWizard, or transferred from Flickr.
- File usage - Files used in Wikipedia articles are already deemed “within scope”. Admins use this to determine the impact of a deletion, and in some cases, to provide an alternative image to illustrate the articles that will lose the image.

A large proportion of cases require an investigation into the user that uploaded the file because neither the file data nor the file’s internet history contain enough information to make a judgment. The Commons upload history helps admins identify patterns in the user’s behavior to determine the outcome of the case in



question. They look for signs that the user is a good faith uploader, the user is currently active, the file in question matches other work authored by the user, or a history of violations. The [Global Account Information](#) shows where users have productively contributed and where they have been blocked. User talk pages show a history of deletions, other warnings from moderators, or whether a user has engaged in discussion about past violations. Looking at a user's Commons upload history often reveals more prohibited files that had not been discovered and flagged.

*I check to see if the user is blocked anywhere, if they are active, if there is more self promotion stuff in their upload history, or if they have deleted articles, etc. -P6*

*In order to get evidence that the uploader is the author I look at past contribs. Does the image seem similar to the quality of other uploaded images for which they are the author? Or have they uploaded a bunch of copyright violations in the past? -P7*

## Tools

Every participant named Google Lens and TinEye in their workflows, which are both used in copyright violation cases to determine whether an image appeared on the internet prior to being uploaded to Commons. Other tools mentioned by participants are:

- [Twinkle](#)
- [Adiutor](#)
- [DelReqHandler](#)
- [Instant Delete](#)
- 3rd party tools such as keyboard shortcut tool to populate messages with documentation, and a tool that draws frames around images in FireFox to open multiple tabs (that contain flagged images) at once, and then automatically opens the next after closing one.

*Without these external tools, I couldn't do as much as I do in the time that I have. -P9*

There is a pronounced lack of sophisticated, well maintained tools to efficiently manage deletion requests. Many admins frequently raised the need for tools to manage batches of deletion requests. Many participants remarked insufficient or nonexistent support for existing tools.

## Beyond Moderation

Participants universally demonstrated their assiduous approach to their role as administrators. They want to uphold Commons as a high quality resource and they don't want to discourage good faith contributors. Often

these two desires are at odds. They must delete files while carefully interacting with uploaders so as not to discourage future participation.

In order to prevent deletions of useful content, many try to correct the mistakes of uploaders, either by changing a license, an author, or some other piece of data. All admins are or have been Wikipedia contributors, so many admins transfer images to or from local Wikis to make images maximally accessible. Some also strive to replace deleted images used in on articles with an alternative.

## Language

English is the de-facto language on Commons, but any and all languages are represented in file names, captions, descriptions, articles, and sources. While there is fairly good coverage of languages amongst the Commons Admins group, there are still some files that require language skills not represented. Admins will handle complex cases in the languages they know. They skip the complex cases in languages they don't know because those cases require discussion, specialized regional knowledge, or both. For relatively simple cases, Admins do not need specific language skills. All Admins interviewed report moderating files in languages they do not read or speak. In fact, it is a routine part of handling speedy deletions and simple regular deletion requests.

Personal images (speedy deletion category F10) or copyrighted images taken from the internet (speedy deletion category F1) make up the bulk of the simple cases that appear in the deletion queue in non-English languages. For F10 images, admins simply look at the image to verify it is a personal image. Next, they may look at the user history to determine whether the uploader is an established contributor who has earned the right to upload a personal image to their own user page. In most cases, this is enough process to delete the image.

For copyright violations, admins check the date of upload. They check the URL of the original file if it is already provided in the deletion request or by the uploader. If not, they use TinEye and/or Google Lens to search the image for the file and determine whether the uploaded file was taken from the internet. In some cases, they may have to verify the license on the website that contains the image in question.

When some moderate investigation is needed on a foreign language article or website, admins use machine translation tools. All participants were confident in handling speedy deletion or otherwise simple cases in languages they do not read with the information already provided in the file, using Google Lens and/or TinEye, or using machine translation tools.

The most frequently reported challenge related to moderating non-English language files was lack of administrators. Some key language skills are missing from the admin group keeping some cases in the backlog over the long term.

## Deletion Requests: Types and Difficulty

### Categories of Deletion Requests

The speedy deletion nomination process requires categorization by the nominator. The categories that make up the bulk of the nominations are copyright violations, duplicates, personal images, advertisements, and courtesy deletions<sup>5</sup>. Copyright violations can be further subdivided. The most frequently mentioned subcategories by participants were: images taken from a website, incorrect author attribution, incorrect license (including [license laundering](#)), genuine confusion about public domain images, and the uploader's own images taken from social media or Flickr without releasing the image under a free license<sup>6</sup>.

### Content of Deletion Requests

Participants reported the following as frequent content flagged for deletion:

- Copyright violations - images of footballers, entertainers, singers, and politicians.

*A lot of speedy deletions come from younger folks adding pictures of celebrities, footballers, and musicians to Wikipedia articles. -P9*

- Personal images - selfies, group pictures, pictures of a single person taken by someone else.

*Selfies. Users confuse Commons and Wikipedia with social media and upload selfies, pics of their friends. Then they create a page for themself. Everyday we have 10 people who do that. -P15*

- Advertisements - imagery often associated with small businesses around the world. Often accompanied by an Wikipedia page (marked for deletion) about the business and frequently viewed as *spam*.

Regular deletion requests can include the aforementioned speedy deletion case types, but require some discussion or investigation. Other common deletion cases include:

- Public domain disputes - debate as to whether an image qualifies as public domain, including:
  - ◆ freedom of panorama - publicly displayed art or architecture
  - ◆ threshold of originality - logos

---

<sup>5</sup> Uploader requests the image be removed within 7 days of upload, often because it was uploaded by mistake.

<sup>6</sup> This may not be an exhaustive list.

- Derivative work - pictures of book covers, album covers, screenshots of copyrighted software, and more.
- De minimis - copywritten material in part of an image.
- Out of scope files - bad quality images, poorly composed images, personal images, images with no obvious encyclopedic value, and sometimes .pdfs and screenshots.

## Difficult Cases to Resolve

All admins agree that cases involving copyright law are the most difficult to resolve. Such cases involve the following copyright issues: freedom of panorama, threshold of originality, de minimis, and public domain outside of the United States. Most admins are well versed in US copyright law as it pertains to files for upload on Commons, and most have knowledge of copyright law in some other countries. Before admins can even begin to apply copyright law to their process, they have to determine which country the file is from and when it originated. This information is not always available in the file data. Further challenges arrive from countries or territories with ambiguous laws, no laws, or when the file may be out of copyright in the country of origin, but not in the United States.

*First you have to determine which country's laws apply- but sometimes you don't have information on where the image is from. -P10*

*There is a lot of disagreement about which CR laws should apply to each case. If the [WMF] legal team can have that conversation with the community, then we could know what is/isn't allowed for sure. -P2*

With respect to logos, which was the only threshold of originality (ToO) file type mentioned by participants, participants expressed a range of opinions about the difficulty of these cases. Some participants found it difficult to determine which logos are simple enough to be considered free license. Others participants appeared confident in their abilities to judge the logo complexity. Uploaders appear to be confused as evinced by the numerous logos that are flagged for deletion. Regardless of how admins perceive the difficulty of these cases, the quantity of cases is time consuming.

## Admin Decision Making and Guidance

To make decisions about deleting files, admins become familiar with the necessary laws and policies applied on Commons over time. This appears to happen primarily through reading case discussions and the documentation on Commons. The newest admin in the sample expressed the desire to learn policies in a more structured manner than currently exists.

*There are a lot of diff policies, lots of documentation, but new admins don't yet know all of the different policies. It would be nice if there were learning pathways, to learn a policy and then see how it applies. -P2*

The other resource that exists is a Village Pump where moderators can go for general discussions about copyright and to ask for opinions on cases. However, few admins reported seeking consultation of any kind outside of the individual case discussions. Admins more often skip cases that they are unsure about or require more extensive research.

## Challenges and Burdens

When asked about challenges and burdens, we saw a number of repeated responses from admins, as well as numerous other unique responses. Admins were not asked to evaluate specific ideas, and instead were encouraged to brainstorm spontaneously. We can assume there would be broad agreement on each of these issues, but we did not test for agreement in the interviews.

### Workflow Challenges

Participants surfaced an array of issues when questioned about their workflows and challenges as moderators. The focus of this section will be issues that likely necessitate engineering solutions.

### Finding & Organizing Deletion Requests

*If we have another way to organize the backlog, it would help focus admins on their interests. -P2*

Currently deletion requests are organized by date. There is no way to view the reason for the nomination ahead of clicking on the date and reading through cases. Many mentioned the desire for better categorization of deletion requests. It would help direct them to the cases they either have knowledge of or an interest in tackling. They spend time clicking on dates, scrolling, and searching around for the cases they can handle. With a categorization system, finding cases could be streamlined, and they could better use their time resolving, rather than finding, cases. Some ad hoc suggestions of categories were: reasons for nominations, cases requested by author but are greater than 7 days, images that were uploaded via Wikipedia.

**Suggestion: Provide a category structure for regular deletion requests.**

### Streamlining Deletions

*The sheer volume in the backlog is burdensome, and dealing with each is a very manual process. There are no good scripts to handle all at once. -P7*

Many participants mentioned the need for scripts or tools that enable batch processing of many deletion requests at once, and where different reasons can be selected for each image. Currently the process is manual and tedious. It requires opening each file separately and selecting individual reasons. One participation noted that such a lengthy process explains why so many requests languish for months.

**Suggestion: The ability to see Exif data on the deletion request page for each image. Currently the data is only on the file, which requires an extra click to view.**

Many participants bemoaned the lack of maintenance support for essential tools and gadgets. Every admin relies on tools and gadgets to streamline their work, however most of these essential tools are not supported by anyone other than a single developer. There is often no one to fix a tool that breaks. Lack of consistently maintained tools and gadgets cause additional friction in the moderation process.

*Some of the most basic gadgets are maintained by one person. Those people will go missing for a while and the gadget will break. -P12*

**Suggestion: Institute a system and resources to support essential gadgets and tools.**

## Autopatrol Needed

*The backlog of unpatrolled files is a problem. -P10*

Currently every file and every edit by every user is marked as “unpatrolled”. Patrolling each element creates a backlog that some moderators raise as an additional burden. On English Wiki, only new pages require patrolling. On WikiData, edits are patrollable until the user reaches 10 edits.

**Suggestion: Autopatrol activated after the user achieves X contributions.**

## Categorization

*It's easier to upload a stolen image than to find a free license image with the search/category structure we have. -P13*

The categorization system creates multiple burdens for administrators. It is poorly understood, even by good faith contributors who engage with it regularly. It is tedious to categorize an image and it is difficult to find images within the category structure. How this complexity burdens the admin workflow stems from the confusion and apathy of uploaders. Uploaders don't understand categories, and/or they do not bother to engage with it, which creates an enormous backlog of unsorted, or poorly sorted, and essentially unsearchable files. Admins worry about the files that are unpatrolled because they are hidden as a result of being uncategorized. These hidden files contain potentially harmful or ineligible content that stays undetected on Commons for years.

*There are 1 million images that need categories. It's overwhelming because there is stuff that doesn't belong there which could be hurting someone. -P5*

*My concern is not only the backlog of DR, but the huge pile of unsorted files, or grossly sorted, that are not proposed for deletion. These are not reviewed, they are not categorized. These piles might stay there forever. -P4*

Another source of friction in the workflow is finding images to use when replacing ineligible images on Wikipedia articles. It was mentioned that users will replace eligible images on an article with ineligible images. Once the ineligible image is deleted, the article is without an image. When such images are in use on an article, some admins attempt to replace the image with a preexisting image on Commons. They complain that the category structure makes it difficult to find those images.

*It would be nice if there was an automated way to show free license images for articles. The fine grained category structure that some people apply on Commons makes it hard to find the photos. Would be fantastic if commons had a way of visualizing multiple category tiers at the same time. At present, it's a tiered structure that goes from general to specific. Example: picture of that soccer from 2010, pic of that soccer player from 2011, etc. just want to see all pictures of that soccer player. -P13*

**Suggestion: Better ways to discover uncategorized files that have no or poor descriptions and titles, making them otherwise unsearchable.**

**Suggestion: Rework the category structure so that it is simplified and flatter.**

## Organizational Challenges

### Cross-Wiki Uploads

*The Commons community is not a fan of cross-wiki uploads. -P12*

There is evidence that about a third of deletion requests are images that were uploaded via Wikipedia to illustrate an article. Each wiki has a different relationship with Commons. For example, files uploaded through English Wikipedia are stored on Commons. By contrast, Italian Wikipedia hosts many files locally because they are legitimately in the public domain in Italy, but not in the U.S. In the former cases, when articles are deleted from local Wikis, the corresponding media stays on Commons. If the article is deleted, most frequently the image also has to be deleted. Currently this requires a separate moderation process. Some admins expressed wanting to couple the fate of the article with its media, thereby reducing two moderation processes to one. The reverse is also true: spammers upload on Commons and subsequently use the files on other projects, requiring two moderation processes. Thus, cross-wiki uploads necessitate a level of cross-wiki moderation that does not currently exist.

*Maybe their contributions are automatically checked when they do something on another wiki after being banned from other projects. -P12*

*Cross-wiki sock puppetry is a particular problem. -P12*

There appears to be some disagreement as to whether all files for Wikipedia should be hosted on local wikis, or all should be hosted on Commons. If hosted on local wikis, then local moderators would simultaneously deal with the article and the image. Some admins advocate for directing all image uploads to Commons, instead of the local wiki upload tool, because the UploadWizard has more barriers built in to deter ineligible material. Either way, it is clear that a large portion of copyright violations are being uploaded through Wikipedia, requiring moderation and feeding the backlog.

*Usually people don't come directly to Commons with their garbage. -P4*

*It would be desirable to remove feature of local wiki to upload directly to Commons, at least for new users. -P8*

## **Lack of admins and functionaries on Commons**

*Only 189 admins on Commons and we have a comparable workload to Admins on en.wiki. Many aren't even active. -P15*

All participants emphatically agree there are not enough admins to deal with all of the deletion requests and the other work on Commons. Despite there being 189, a total count that is already low given the amount of work to do, but the deletion by admin data shows that only a fraction do the majority of the work on Commons<sup>7</sup>.

A related issue that impacts admins is too few functionaries on Commons. One participant stated that only 3 Commons admins have CheckUser rights. Having too few users with extended rights enables harassment, abuse, and sockpuppetry, which ultimately weigh on administrators and affect the experience of other users as well.

## **Moderator Alignment**

*Currently there is this gray area about what is allowed. There are also thousands of other files that could be on Commons and we're not sure of the copyright. -P2*

Discussion is built into the deletion process and it is reasonable that some amount of disagreement will ensue given the complexity of copyright law. But in the absence of clear policy on the platform, discussions can turn into unresolvable disputes that consume a lot of time and energy from moderators. Many participants expressed frustration with a lack of clear policy or guidance with respect to applying various copyright laws, especially when they are in conflict with one another. This leads to more time spent on any given deletion request. One participant suggested the Wikimedia Foundation take more of an active role in facilitating those conversations with the Commons community.

---

<sup>7</sup> [username by deletion count, June-July 2023, Feb-July 2023](#)



*If the legal team can have that conversation with the community, then we could know what is/isn't allowed for sure. -P2*

*People who are nominating for deletion are not clear about what should be speedy vs regular. All of those things are unclear or built into the platform. -P2*

Several participants mentioned that there are varying levels of knowledge and competency among moderators, including dedicated patrollers and users who flag for deletion. When moderators don't understand the rules or practices, it can create additional confusion and work for administrators. One frequent issue mentioned is moderators are not always flagging files correctly. Files are often flagged for speedy deletion when they actually require the regular deletion process. Another participant complained that when moderators flag a large number, perhaps several dozen files at once, they should use Batch DR. When this doesn't happen then each image requires a separate process. These examples point to a lack of synchronization within the moderation function.

*There is the burden of tedium. Sometimes each image requires a different deletion justification. Processing a large number of related images (even if eligible, it's better for batch DR). -P7*

## **Upload Process**

*The upload tool allows bad files to be uploaded. Images that are too small, or they mention the original author in description, even if the uploader selected "own author", both could be caught by technology. -P2*

Among participants, there is universal agreement that the upload process, both the UploadWizard on Commons and the various upload tools on Wikipedias, are too easily allowing ineligible files onto the platform. Most participants seem sympathetic to the fact that copyright is complex, it's not within everyone's cultural context, and thus there will always be some amount of confusion on the part of uploaders. But they believe that technology should intervene to prevent uploaders from uploading bad files. There are a number of ways that the upload process exacerbates the problem. It both encourages bad uploads at times and misses the easy to catch violations. The following section describes issues related to the upload process. Some sections focus on features of the UploadWizard, some focus on features of the Wikipedia upload tool, and some address general issues related to uploading that could be addressed with design or engineering.

## **Few Barriers to Uploading Images**

*You cannot upload a video straight from your phone to Commons. You have to convert the file. So for that reason we don't have a lot of videos. We usually get them from other sources. There are fewer barriers for people uploading images or text files, that's why we have so many ineligible or out of scope files. -P2*

There is widespread agreement among participants that there are not enough barriers to uploading media, particularly images and text, on Commons. Participants feel the UI is designed to encourage participation, and move users quickly through the process, rather than forcing them to interact with the text and make conscious, accurate selections. Additionally, an uploader can populate the fields with incorrect or

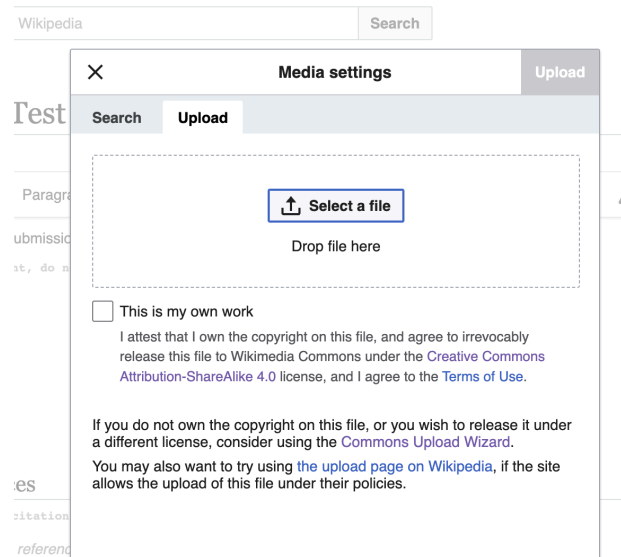
contradictory information and still successfully upload an image. For example, an uploader can upload a file after selecting “own work” even if they write in the actual author in the description.

*Some subtle hurdles for uploaders would reduce DRs. The WMF focuses on participation. For many uploaders, their first contribution to WM is uploading an improper image because there are no hurdles. Therefore, we have to delete a lot of images. -P15*

**Suggestion: Machine detection and technical solutions to prevent bad uploads.**

## VisualEditor Upload

The only prominent option on the VisualEditor upload tool on Wikipedia is “own work”. It does not give options for other scenarios. Therefore, uploaders who are unaware of copyright and licensing rules are not presented with the opportunity to learn the rules in the process. They will upload a copyright violation without ever having to consider the copyright status of their upload. Participants pointed out that this is a large source of copyright violations and other bad files in the deletion queue.



No prominent option for “not own work”.

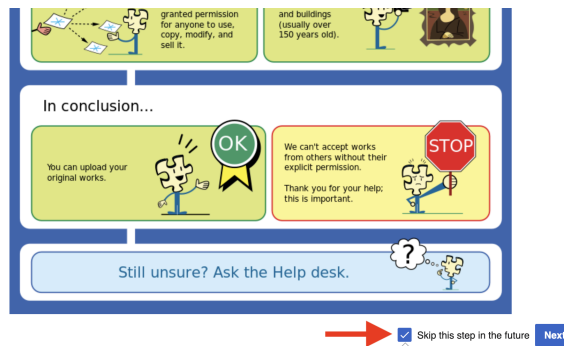
*The Wikipedia upload tool practically forces you to select “own work” and the majority of copyright vios are coming in from Wikipedia. -P8*

**Suggestion: Reconfigure Wikipedia uploads to include the same steps and required fields as the UploadWizard.**

## Learn

*Is [the infographic] really the best way to communicate the scope of Commons? I'm doubtful. -P5*

While many participants appreciate the purpose of the infographic, some expressed doubt that it was the most effective way to convey the policies and scope of Commons. Most felt like it was not adequate alone. Some suggested having a video or a tutorial in addition to the infographic. There was another suggestion to make new users demonstrate their understanding of the rules before they can make an upload. And all of the participants that critiqued the Learn screen offered the opinion that, contrary to how it currently functions, new users should not be able to skip the tutorial after only one upload.



Users can skip the Learn screen after viewing it once.

**Suggestion: Maybe one feature to consider is, say, a special user right: you can skip over the tutorial on UploadWizard when you have X edits or some threshold. Otherwise, make every user go through the tutorial. -P10**

**Suggestion: For complete beginners, make them go through a 5 min video, or some process where we can confirm they know what it is about. -P2**

## Preselections

*Most select "own work". Some choose "not my own work", but then they are selecting CC Sharealike 4.0 bc this is pre-selected as the default. In both cases the licenses they end up selecting are not correct. -P11*

Many participants expressed frustration with the nudges built into the upload process. At present, several options are preselected in the UploadWizard. Creative Commons Attribution-ShareAlike 4.0 is the default license selected as the "own work" license and the "not own work" license. If this is the most likely license to apply to valid uploads, it is not the most likely to apply to the ineligible uploads, which are numerous. Therefore, a number of participants suggested that removing the preselections would force uploaders to

think about which selection truly applies. All participants agree: more intentionality required in the upload selections should lead to fewer deletions down the line.

**Author(s)**  
 The name of the person who took the photo, or painted the picture, drew the drawing, etc.

Now tell us why you are sure you have the right to publish this work:

- The copyright holder published this work with the right Creative Commons license**

Not all Creative Commons licenses are good for this site. Make sure the copyright holder used one of these licenses.

If the work is already published online, but not under that license online, or if you aren't the copyright holder of the work, please follow the steps described at COMVRT (the copyright holder has to send an email with relevant permission) and add {{subst:PP}} to the "Source" field above.

  - Creative Commons Attribution ShareAlike 4.0 (legal code)
  - Creative Commons Attribution ShareAlike 3.0 (legal code)
  - Creative Commons Attribution ShareAlike 2.5 (legal code)
  - Creative Commons Attribution 4.0 (legal code)
  - Creative Commons Attribution 3.0 (legal code)
  - Creative Commons Attribution 2.5 (legal code)
  - Creative Commons CC0 Waiver (release all rights, like public domain: legal code)
- The copyright holder published their photo or video on Flickr with the right license**
- The copyright has definitely expired in the USA**
- This work was made by the United States government**

CC ASA 4.0 Preselected in the UploadWizard

**Suggestion: Eliminate preselections.**

## Release Rights

*Most of the time uploaders select “own work” when it is not. -P11*

*Uploaders with problematic files always select “own work”. Probably because that is the first option, or because there is too much information required for selecting “not my own work”. They just want to upload their company logo or their favorite footballer and be done. -P12*

There are two very frequent patterns in user behavior in copyright violations. First, the user selects “own work” when the file is not their own work. Often uploaders will provide the real author and source of the image in the description field. Uploaders will select “own work” in personal images when the photo is clearly taken by another individual who was not credited. Moderators spend time establishing that the uploader is in the photo and that the photo is likely taken by someone else.

This pattern suggests that uploaders are not certain or not concerned about authorship of the image. This was also identified as an outcome of the Wikipedia upload tool, which offers “own work” as the only author option. But participants perceived this as a pattern present in uploads through the Commons UploadWizard as well. More analysis is needed to know what proportion of deletion requests with “own work” erroneously selected come from which source. However, several participants believe that uploaders are selecting “own

work” because it is a quicker process than selecting “not own work”, and there are no hard stops built in to thwart the upload.

There were a number of suggestions to affirm that the work is really that of the uploader. P5 above is suggesting that the upload process confirms, more than once, that the uploader is the author. Another participant suggested that there be an additional field in the “own work” selection that asks the uploader where the work came from. Another suggestion is the upload tool accepts “own work” only if there is Exif data in the file.

*Create a conscious understanding of what is really your own work. “Is this really your own work? Like, really?” A certain prompt would help validate and maybe discourage. -P5*

**Suggestion: Build in more opportunities into the upload process to establish the uploader is the author of the file.**

In the second pattern, the user uploads a file taken from the internet, selects “not own work”, but they select the free license - public domain option on the release rights step for a copyrighted image. This pattern suggests that uploaders do not understand what constitutes free license and public domain.

*Uploaders have confusion about “free license” - they think bc it’s on the internet it is free license. -P3*

**Suggestion: Reinforce concepts of free license and public domain to increase uploader understanding when this option is selected.**

## Order of Options

*Most of the nominations are copyrighted files found on the internet and uploaded. -P11*

This site requires you to provide copyright information for this work, to make sure everyone can legally reuse it.

- This file is my own work.  
 This file is not my own work.

**Source**

Where this digital file came from — could be a URL, or a book or publication.

**Author(s)**

The name of the person who took the photo, or painted the picture, drew the drawing, etc.

Now tell us why you are sure you have the right to publish this work:

- › [The copyright holder published this work with the right Creative Commons license](#)
- › [The copyright holder published their photo or video on Flickr with the right license](#)
- › [The copyright has definitely expired in the USA](#)
- › [This work was made by the United States government](#)
- › [Another reason not mentioned above](#)
- › [I found it on the Internet — I'm not sure](#)

According to participants, the order of options as they appear in the upload process also influence uploader's selections and consequently, deletions. The most problematic step is the "not own work" screen. The last option is "I found it on the internet - I'm not sure". This is the option that applies to a large proportion of images uploaded and later deleted, yet it is not in a prominent position.

**Suggestion: *The most probable case should be first: "not my own work."* -P14**

Another field in the UploadWizard affected by order is file description. It appears third on the Describe step, below title and caption. Some participants felt it should be above the caption, because the description is used for search, not the caption. The assumption is that boosting its position will increase the likelihood that uploaders will provide a more useful description. If it appears lower than the caption, form filling fatigue may set in.

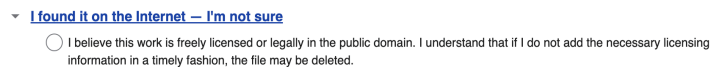
**Suggestion: *On the describe page, the "description" should be first because that is what matters for search, not the caption.* -P10**

## Interrupt the Process

*Interrupt their process more. Give information of consequences of CR violations, or consequences of some sort. More stops in the process, maybe more symbology since people don't read.* -P13

There are two predominant patterns common in deletion requests. First, uploaders upload ineligible images. Second, uploaders misrepresent images that are both eligible and ineligible. All participants expressed their irritation at there being few obstacles to uploading an image. They observe that users can just click “next” until they succeed in their goal of uploading their media, without any hard stops built into the process to deter these uploads.

The “not my own work” screen offers one option to select if the image was found on the internet. The only option presented is “I believe this work is freely licensed or legally in the public domain.” There is no option for the more likely case, that the work was found on the internet and is under copyright. That is because that image is not allowed on Commons. The UI assumes that users will understand that their image is ineligible for upload at this point and abandon the process. We cannot know how frequently that happens, but we know that some uploaders either affirm that the work is freely licensed when it is not, or turn back to select “own work” to avoid the question.



No options present in the UploadWizard that cover “not my own work” and unknown copyright status.

*People don't understand what copyright is, at least new users. We have to explain that everything is copyright by default. -P15*

**Suggestion: Give a secondary option for copyrighted images that explicitly explains the upload is not allowed and bar the upload from advancing.**

## Licenses

*There are five different options. Maybe cut it down. When you click on the CC ShareAlike 4.0 and 3.0 licenses, they have the same wording. It's unclear what the differences are. -P6*

The majority of participants felt that uploaders do not understand the license options. Typically uploaders are selecting the default license (CC BY-SA 4.0) or they are claiming the file is freely licensed/public domain.

If “own work” is selected, there are five Creative Commons license options. There is no difference in wording between the linked legal deed for Creative Commons Attribution ShareAlike 4.0 and 3.0, as with the licenses Creative Commons Attribution 4.0 and 3.0. Assuming most uploaders will not read the expanded legal code for any license, it's unlikely they will be able to discern the differences between 4.0 and 3.0. Some participants doubted the license deeds were available in other languages which may be adding to the confusion for uploaders.

## Suggestion: Reduce license options.


### You are free to:


**Share** — copy and redistribute the material in any medium or format for any purpose, even commercially.

**Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

### Under the following terms:

 **Attribution** - You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

 **ShareAlike** - If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** - You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

### Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.


### You are free to:


**Share** — copy and redistribute the material in any medium or format for any purpose, even commercially.

**Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

### Under the following terms:

 **Attribution** - You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

 **ShareAlike** - If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** - You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

### Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.

Creative Commons Attribution ShareAlike 4.0 ([legal code](#))

Creative Commons Attribution ShareAlike 3.0 ([legal code](#))

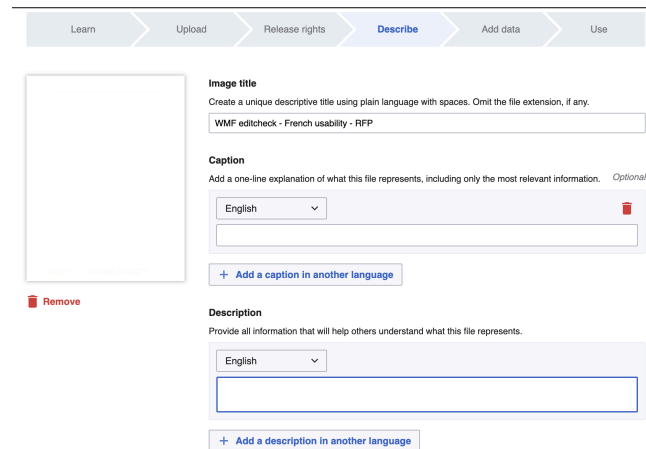
## Description

*Very often a good sign that something is actually their own work is when they write in a description, or they write in words that they took the photo. Rather than boxes are checked. -P7*

Most participants would favor machine detection to focus on the description to catch bad uploads. Often uploaders will claim “own work” but write in the original file author into the description. Files with no or poor descriptions are often out of scope or copyright violations. Several participants noted that files without good descriptions are essentially useless because they are often uncategorized and cannot be found by search.

**Suggestion: Expand the help text to succinctly explain that description is needed for images to be found using search.**





The screenshot shows the 'Describe' step of the upload wizard. At the top, a progress bar indicates the steps: Learn, Upload, Release rights, Describe (highlighted), Add data, and Use. Below the progress bar is a large empty box for the image, with a 'Remove' button below it. To the right of the image box are three sections: 'Image title' with a text input field containing 'WMF editcheck - French usability - RFP'; 'Caption' with a language dropdown set to 'English' and a text input field, plus a '+ Add a caption in another language' button; and 'Description' with a language dropdown set to 'English' and a text input field, plus a '+ Add a description in another language' button.

## Mass Upload & Flickr

*Mass upload from Flickr and other sites results in a lot of poor quality images, unsearchable in the archive, spam, and copyright vios. -P12*

*Maybe limiting the number of files someone can upload- 5 or 10 files for the first day. Several times a week, I delete 50 or 100 files someone uploaded on the first day within 20 minutes of registering. -P9*

*There is lots of garbage coming from Flickr- their uploads are within the rules but also not necessary for an encyclopedia. -P4*

Many participants lamented the mass upload function in the UploadWizard and the Flickr transfer function. Admins frequently see large batches of ineligible, out of scope, or poor quality files from new users who have uploaded multiple files, often dozens or hundreds, at once. As previously noted, deletions are a manual process, and handling mass uploads can be tedious and time consuming. All admins favor some limitations of this feature. Most made suggestions along the lines of imposing some limits on the number of files new users can upload at once.

**Suggestion: Limit mass uploads and/or Flickr transfers for new users.**

## Releasing Rights (VRT)

*We don't make it clear enough during upload when users have to go through the VRP process. -P8*

One source of copyright violations that administrators see regularly are images taken from social media. Uploaders are completely unaware that they need to release the rights of their own photos if they have previously uploaded them anywhere else on the internet, including social media. There is no obvious place in the upload process that explains that requirement. For every image in this category, a moderator has to

explain the policy and direct users to the [VRT process](#). One participant suggested creating a backend filter to throttle these images, as they often contain metadata specific to the social media platforms, and either block the upload or direct users to the VRT process.

**Suggestion: Make users aware during the upload process of the requirement to release rights.**

**Suggestion: Throttle uploads with social media meta data.**

## Content on Commons

*The vast majority of DRs are coming from new users. Another large percentage are logos or photos of people. -P12*

Photos of people and logos were specified as some of the most prevalent deletion requests by many participants. The rules concerning images of people and logos are difficult to communicate and difficult for new users to grasp. In both cases, both image types are allowed on Commons, but only in specific cases. The nuance is not easily communicated or understood. Some participants suggested that new users should not be allowed to upload these image types immediately; they must earn the right through proving their understanding of the rules or making X productive edits.

*Commons is not censored but there are a lot of unsolicited nude images. Some argue its well categorized and good quality and it's relevant. But I have concerns about whether some of the people in the images are consenting/willing/knowing. -P6*

A source of concern among all participants is the junk content that is eligible under the broad interpretation of Commons rules, but either not useful in an encyclopedic archive or unethical. One participant expressed concern about uploads of indecent photos and other material that can cause harm. Such images can be allowed under Commons rules, depending on the specific imagery depicted, but this participant doubted that the subjects in the photos were aware of or consented to their photos being taken and uploaded. Currently, [Commons policies](#) attempt to balance dual priorities of appropriate scope and remaining uncensored, but certain images with sexual overtones fall into a highly subjective gray area that can be stressful for administrators to moderate when there are strongly opposing views present in the discussions. Administrators may sometimes feel responsible for protecting people depicted in photos.

*There is a lot of junk on Commons: images that are technically acceptable and eligible but not needed. -P4*

Many admins are dismayed by the amount of “junk” that is regularly uploaded to Commons. By nature of their commitment to Commons, administrators tend to view the archive as a treasured resource, and their participation in the project comes with the personal goal of maintaining quality within the archive. With this sentiment, participants frequently expressed concern over the use of Commons to host poor quality images and images lacking any apparent educational value. Such images often require their moderation in deletion cases, including their involvement in discussions about scope. But the quantity of such images that are uploaded to Commons also burdens admins with the unique responsibility of maintaining Commons as a quality encyclopedic resource.

*I'm a child of a librarian. I want to see Commons have the quality of a traditional archive. I want it to be as good as it can be. -P12*

**Suggestion: Address primary sources of poor quality and irrelevant images by redesigning the Wikipedia upload tool and UploadWizard to make uploaders more aware of rules and scope; limiting mass uploads and transfers from Flickr; and wherever possible, enabling filters or machine detection to catch duplicates, poor descriptions, and otherwise low quality images.**

**Suggestion: Throttle images of people and logos for new users.**

## Uploaders: Perceptions and Interactions

### Interactions

*Selfie folks aren't concerned about learning. For them Commons is just a place to host their selfies. They won't respond. -P2*

*But there are good faith editors; these are all newbies. They are editing an article on ar.wiki, wanting to add an image, redirected to Commons, and they end up uploading a copyright vio. That's an opportunity in which folks can be engaged. Usually when people engage in these cases, the interaction is positive. -P2*

Several participants described common behavioral patterns of constructive and destructive users. New users who blossom into high quality contributors often start with uploading mediocre quality pictures from their phones that require moderation around categories and description. Another common productive user are those wanting to illustrate an article on Wikipedia and unknowingly upload a copyright violation. Good faith contributors are more likely to read their talkpages and engage constructively with admins. In this process they tend to learn about the scope and rules of Commons. Most apathetic or destructive contributors will not read their talkpages or respond to deletion nominations. There are a number of trolls, vandals, spammers, and otherwise dedicated but disruptive contributors on the platform who will engage in a vexatious manner. These users will often become blocked but they may then engage in sockpuppetry elsewhere in the Wikimedia projects. Often they leave behind a trail of destruction that admins feel obligated to repair.

### User disputes

*There is the burden of user disputes involving controversial and political imagery. I have to block all new maps on these issues. It's not scalable, I can't review dozens of edits to many maps. That's a problem. -P15*

Every participant expressed a low appetite for engaging in heated disputes on Commons deletion cases. Some amount of stress or discomfort comes up for participants in these exchanges. Most simply avoid these disputes after having engaged more actively in the past, but at times they are drawn into disputes by tackling objective moderation tasks. One example given was in the case of uploads that reflect certain political views,

such as new maps of Russia and Ukraine as they relate to the current war, or maps depicting abortion restrictions in the United States. Admins are also forced to answer to users on any deletion decision they make. Users who disagree with their decision may get upset, hurl insults, complain in various forums, and even go so far as to try to strip admins of their titles.

*Dealing with controversial DRs means dealing with the effects of those decisions. -P14*

## Copyright

*There is only so much you can get uploaders to understand. At the end of the day there are a lot of people, for whom the concept of copyright is not in the cultural lexicon. There is a generational gap, where younger users who have grown up on social media are not used to thinking about restrictions. -P12*

*Newbies and don't know what copyright is, or don't think applies to files on the internet, movies, and books. -P14*

Participants are sympathetic to the fact that copyright is complex and it's unrealistic to expect new users to fully understand it. New users are accustomed to social media websites that don't have such restrictions and/or they are from a country where the concept of copyright is not pervasive.

Yet the biggest complaint from all participants was about new users uploading copyright violations. This issue was brought up multiple times in every interview, which indicates there is a strong need to increase understanding of basic copyright for new users and impose technical solutions to thwart bad uploads. No admin suggested curbing the rights of experienced users, but all were in favor of placing some limitations on new user uploads. These ideas are presented in more detail in the next section.

*For the people who are well intentioned and may respond to some information, but yet unknowledgeable about copyright, we can try to educate them and throttle certain actions. -P13*

*Several times a week I delete dozens of files from someone who uploaded on the first day of registering an account. This is not possible on other platforms. Such rights are accrued stepwise according to time on the platform and contributions. -P9*

## Participant Suggestions

The following ad hoc suggestions were by admins in the interviews. We did not test for agreement on any specific ideas, but consensus naturally emerged around some general ideas. Each bullet point represents at least one participant's views and may represent multiple if the idea was repeated.

### Changes to cross-wiki uploads

- Erect a higher barrier between wikis and Commons. Local moderators would simultaneously deal with articles that are up for deletion and their associated images.

- Remove from local Wikis the ability to upload directly to Commons to stop the deluge of deletable files.
- Only allow experienced users from local wikis to upload to Commons, rather than everyone. For newcomers, it would be better to upload to local wikis only. That way the destiny of their pictures is the same destiny of their articles.
- Cross-wiki moderation is needed to deal with cross-wiki uploads.

## **Changes to Wikipedia upload tool**

- Provide the same tutorial that's on the UploadWizard on the Wikipedia upload tool. Currently there is no tutorial on the Wikipedia uploader.
- Force all uploaders to read and digest policy before proceeding.
- Change the options on the Wikipedia upload tool so that people aren't compelled to select "own work".

## **Changes to UploadWizard**

- The most probable case should be first: "not my own work."
- On the describe page, the "description" should be first because that is what matters for search, not the caption.
- The UploadWizard needs to be socially engineered to get the behavior we want. Only a few will bother to learn about the scope of Commons and licenses.
  - ◆ This problem needs an interface engineer to design things visually with some social engineer who knows how people think to create an upload process that will discourage bad actors and uninformed users.

## **Changes to all upload tools**

- If "own work" is selected, make people explain where the image came from.
  - ◆ Create a conscious understanding of what is really your own work. More prompts that make them confirm it's truly their own work.
- Develop AI to filter bad images during uploading to reduce the human effort.
  - ◆ Filters to catch metadata from social media sites. These images shouldn't be banned altogether but throttled. Ask the uploader if it's properly licensed, or inform them that they need to properly license, go through VRT, make sure it's in scope, etc. Tag those photos to

alert the Commons Community that those need to be patrolled. Don't prohibit those images wholesale, but a flag for the Commons Community that something with bad metadata is being uploaded.

- More/better notices about upload criteria and other interventions during the upload process.
  - ◆ Most users don't know about Creative Commons licenses. So there should be an info box explaining CC licenses for each of the 7 licenses.
  - ◆ A video about the process, rules, and image eligibility for new users. Some wikis have videos and so should Commons.
- Interrupt the process more so that it forces the uploader to pay attention and make decisions. More stops, stop signs, red hands throughout the process.
- Force people to think about their selections during the upload process. No pre-selections, no defaults.
- Apply restrictions on *new users* in order to slow the flow of copyright violations and out of scope images, while allowing new users to learn the rules of Commons through incremental productive contributions. The following is an aggregated list of ideas from all participants:
  - ◆ Limit file uploads and Flickr transfers to x per day.
  - ◆ Restrict new users from uploading specific image types. Suggestions were:
    - Images of people
    - Logos
    - Free license and public domain images
    - "Not my own work"
    - Files containing social media meta data
  - ◆ Eliminate the ability of new users to skip the Learn tutorial.
  - ◆ Test new users' understanding of the rules before allowing them to proceed with an upload.
  - ◆ Restrict cross-wiki uploads:
    - Allow only experienced users to transfer files from local wikis to Commons.
    - Implement a cross-wiki moderation system.
- Apply checks to *problem users* to surface and quickly address problematic files, as well as streamline cross-wiki moderation. Mark problem users such that their contributions go to a different queue for extra review.

## **Conclusion**

Due to the complexity of copyright laws and subjectivity in determining what is within scope on Commons, there will always be some number of deletion requests that require thoughtful consideration, research, and discussion. Only dedicated humans can handle these tasks. But at present, admins spend so much of their time on files that should never have been uploaded. Participants gave a wide variety of solutions that could reduce deletion requests, preserve the integrity of the archive, and lighten their workload. Machine detection, backend filters, and restrictions on new users, are widely championed as common sense solutions to reducing the flow of deletion requests.

The Wikimedia Foundation should remain attentive to our valued admins. It should strive to balance new user participation with preserving our admins over the long term. Admins care deeply about the project, and focusing on their needs would demonstrate how deeply the Foundation cares about them.

## **Acknowledgements**

I would like to acknowledge and express my sincere gratitude to all who contributed to this project. Thank you for the time and energy you put into this research. Many thanks to Alexandra Ugolnikova, Sneha Patel, the Structured Content team, Satdeep Gill, KC Velaga, Daisy Chen, Gabriel Escalante, Luca Martinelli, all of the wonderful Wikimedia Commons admins who participated in this research.