

The Visual Side of Wikipedia

Fernanda B. Viégas
Visual Communication Lab, IBM Research
viegasf@us.ibm.com

Abstract

The name “Wikipedia” has been associated with terms such as collaboration, volunteers, reliability, vandalism, and edit-war. Fewer people might think of “images,” “maps,” “diagrams,” “illustrations” in this context. This paper presents the burgeoning but underexplored visual side of the online encyclopedia. A survey conducted with image contributors to Wikipedia reveals key differences in collaborating around images as opposed to text. The results suggest that, even though image editing is a more isolated activity, somewhat shielded from vandalism, the sense of community is an important motivation for image contributors. By examining how contributors are appropriating text-oriented wiki technology to support collective editing around visual materials, this paper reveals the potential and some of the limitations of wikis in the realm of visual collaboration.

1. Introduction

The Wikipedia article on “Plankton” is a lengthy page covering various aspects of planktology: definitions, functional groups, size groups, distribution, biogeochemical significance, and references, among others. In addition to these informative sections, the page also contains eight impressive photos of plankton. Close-up photographs of these microscopic organisms reveal beautiful patterns of luminescent gills and translucent bodies. The colorful display of undulating limbs and antennae resemble something out of a science fiction movie (Figure 1). The photos are so high quality that, when clicked on, their high-resolution versions reveal a multitude of delicate, minuscule hairs that adorn the extremities of these alien-looking creatures.

But wait. This is Wikipedia, an online free encyclopedia created entirely by volunteers. How is it possible that volunteers could have produced such equipment-heavy, scientifically-exact, high quality images of microscopic life forms?

This is where Uwe Kils comes in. Dr. Kils is an associate professor of planktology in the Institute of Marine and Coastal Sciences, at Rutgers University. A few years ago, he was asked to donate one of his images of plankton to Wikipedia. He liked the project

so much that he decided to donate over two hundred of his scientific images to the encyclopedia [10].

Like Dr. Kils, there are thousands of Wikipedia users around the world that contribute images to the encyclopedia. In fact, some of these users contribute to the encyclopedia mainly as image creators and as image foragers. For them, illustrating articles is the most energizing aspect of being part of Wikipedia.

With just over 9% of heavily-edited articles containing images, the encyclopedia still has a long way to go until most topics that can benefit from images are visually represented. Nevertheless, with the “Image” namespace of Wikipedia being one of the fastest growing sections of the site [8], it seems like pictures are becoming an essential part of the project.

This paper investigates the world of Wikipedia image contributors today. What kinds of images do they contribute? What are their motivations? How do they create their images? How are image contributions different from their textual counterparts? How does collaborating around visual materials change the way editors interact on Wikipedia?

To explore these questions, this paper describes the results of a survey of image contributors as well as an analysis of the documentation on image creation and usage on Wikipedia. The results present an initial snapshot of the visual side of the online encyclopedia.

2. Related Work

Various aspects of Wikipedia have attracted the attention of academics. Studies have examined the motivation of volunteers to contribute [1][3], the evolution of article editing trends [7], the reliability of information found on the site [5], and the evolution of the site itself [8]. This growing body of work has shed light on some of the strengths and weaknesses of the site.

In a study that compared Wikipedia and Everything2 articles on the same topics, Emigh and Herring [2] found that Wikipedia entries are stylistically

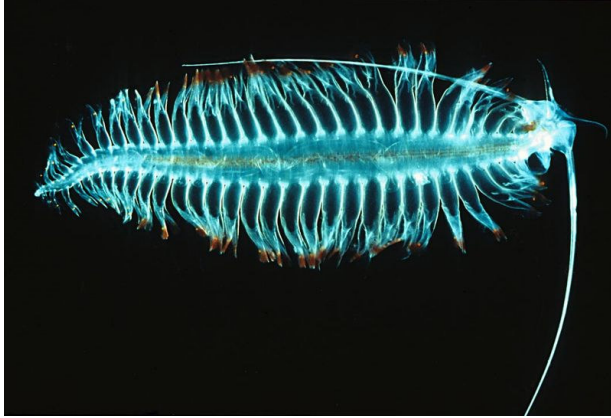


Figure 1: Image of a Tomopterus, by Prof. Uwe Kils. Photograph donated to Wikipedia by the author.

similar to traditional, printed sources such as the expert-created Columbia Encyclopedia, in terms of formality and language standardization. They attribute this phenomenon to the high degree of post-production editorial control afforded by Wikipedia—for instance, the ability to easily edit other’s entries.

Bryant and colleagues [1] have focused on the social trajectories of nine active “Wikipedians,” showing how their roles changed as they became more engrossed in the Wikipedia community. The researchers found that, as participation became more central and frequent, Wikipedians adopted new goals, new roles, and used different tools to achieve new ends. Wikipedians usually moved from a local focus on editing individual articles to a concern for the quality of Wikipedia content as a whole, taking on more “administrative” roles in the site.

Stvilia et al [6] investigated how the Wikipedia community establishes and improves information quality through discussions in “talk pages.” After analyzing the contents of a series of talk pages they found that they play a crucial role in letting users articulate what they perceive as the main issues of quality in the improvement of articles.

Viégas, Wattenberg and Dave [7] downloaded the entire archive of Wikipedia history in order to visualize the evolution of articles and analyze conflict and collaboration patterns in 2003. Through the visualizations, the researchers identified patterns such as edit wars and vandalism repair, which were then investigated further through statistical analysis.

Despite this considerable literature, little has been written about the visual side of the encyclopedia and the community of contributors that illustrates the site. The next section describes the Wikipedia pictorial framework—image formats that are allowed, image

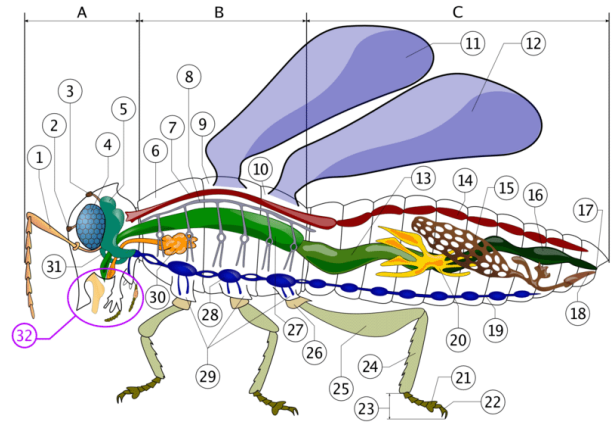


Figure 2: Diagram of insect anatomy, created by Piotr Jaworski for Wikipedia.

usage policies, sources, copyright issues, etc. After that, the survey methodology and findings are explained.

3. Images for Wikipedia

Wiki technology is designed for manipulating text, not images. As a result, collaborating around pictures turns out to be more difficult than collaborating around text.

Images are different from text in ways that fundamentally affect their usage in Wikipedia:

- Images are external, “attached” files that can be added to the site instead of being part of the basic fabric of wiki pages, like text.
- Images are language independent, meaning that they can be reused on sites that are written in different languages. This feature carries important implications for the collection of international Wikipedia sites.
- In addition to access to a computer, contributors need to have special equipment to create images.
- Because images are separate files that exist independently from the wiki site, contributors have the ability to donate or reuse their existing pictures.
- Legal issues are complex: image copyright laws vary from country to country.

These differences mean that images require two classes of consideration: technical and legal. On the legal side, the Wikipedia community needs to be knowledgeable and mindful of international image copyright laws. On the technical side, the following elements affect image contribution: (1) equipment for creating images, (2) image storage, (3) image reusability, and (4) image editing.

In addition to technical and legal considerations, the social dynamics of those collaborating around images is also of interest. Is there a sense of community? Do contributors feel like they can count on each other for help?

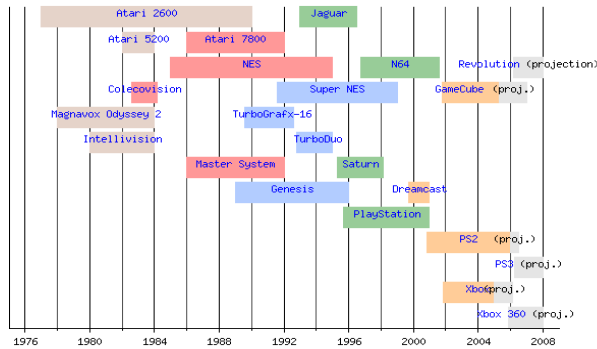


Figure 3: EasyTimeline showing different video game consoles. The name of each game console is a link to the Wikipedia article about that particular system.

3.1 The Commons

The Wikimedia Commons, or the “Commons” as it is usually referred to, is a repository of free-content images, sound and other media files. It is a project of the Wikimedia Foundation and was launched in September 2004. Files uploaded to this repository can be used like locally uploaded files on all other projects on the Wikimedia servers, including the different Wikipedia sites.

Consider the case of an editor from the Brazilian Wikipedia who finds a useful image on the English Wikipedia site. Before the inception of the Commons, the editor would have had to download the picture, save it locally on the Brazilian Wikipedia and then upload it on the article she was hoping to illustrate. The Commons allows all editors to use the same images without having to make local, redundant copies. The goal of having files that are reusable by all Wikipedia sites means that special care has to be taken for images such as maps and diagrams where text might be part of the image. For this reason, contributors are encouraged to use numbers instead of text labels whenever possible (see Figure 2).

Given its primary function as a supporting project for the collection of Wikipedia sites, the main content policy for files uploaded to the Commons is that they must be potentially useful to the encyclopedia or on any of the Wikimedia projects. This excludes material such as purely personal pictures and artwork, in contrast to image sharing repositories like Flickr and DeviantART. It should be noted, however, that large numbers of files hosted on the Commons remain unused by any of the Wikipedia projects worldwide. There are so many images that several of them may never be used by any of the encyclopedic articles.

3.2 Types of Images

The Commons supports a variety of image formats. The preferred formats are JPEG for photographic images, PNG for drawings and other iconic images, and OGG for sounds. More recently, support for Scalable Vector Graphics (SVG) files has become available. SVG support is important because it allows images such as maps, charts, graphs, diagrams and illustrations to be displayed in different sizes without the author of the image having to create different version of the same picture. For this reason, SVG means that these images can be reused more flexibly and easily than their rasterized counterparts.

Wikimedia also supports the creation and presentation of mathematical formulae [11]. In its current implementation, MediaWiki uses a subset of TeX markup, including some extensions from LaTeX and AMSLaTeX. It generates either PNG images or simple HTML markup, depending on user preferences and the complexity of the expression.

Finally, Wikimedia supports the ability for users to create visual representations of time series data in the form of “EasyTimelines” [12]. Such timelines are interactive charts that show detailed listings of events and dates, where each event is a link to the Wikipedia article that describes that occurrence. There are over 80,000 timelines in all international Wikipedia sites combined as of the writing of this paper [16].

3.3 Image Sources

Not all images found in Wikipedia are created by contributors themselves. In fact, there are thousands of images that are imported to Wikipedia without their authors ever knowing about it. The following are the two main sources for images:

1. *Self published:* Each work listed as self published was uploaded to the Commons by its creator. All such works must also have been licensed by their creators before they are uploaded.

2. *Public domain sources:* Each work in this category is available in the public domain for free. The Wikipedia community has been quite efficient in finding sites that offer public-domain images, such as government sites. For instance, several Wikipedia articles showcase images from the U.S. Department of Agriculture, the U.S. Fish & Wildlife Service, and NASA. In addition, Wikipedians have also found university sites that provide repositories of freely available scientific images, such as the Dartmouth Electron Microscope Facility, from Dartmouth College. Contributors also search the Web for repositories of

freely available stock photography. Finally, users scan images that are in the public domain but not necessarily on the Web. There are entire collections of images that have been scanned from history books, for instance.

3.4 Image Editing

Image editing works differently from text editing in Wikipedia. First, users must download the images they are hoping to edit locally to their computers. Image editing software is not provided by Wikipedia so users need to have their own copies of the editing software. Second, the different versions of an edited image are not kept in the Wikipedia server as are the edited versions of article pages. Instead, image files are overwritten when a new version of the image is uploaded. This carries serious implications for collaborative editing and for repairing acts of vandalism—previous studies show that having a permanent history of edits made to an article gives the community the ability to quickly repair vandalism [7].

Moreover, when an image file is replaced by uploading a new one with the same name (e.g. an image in an article is replaced), the change does not show up in the page history of the article, nor is it noted in the revision history of the image description page (unless the description section is also changed).

When a file is uploaded with the same name as an existing one, the old one is automatically overwritten. This suggests a weakness in how image files are managed today in Wikipedia. By allowing image files to be overwritten, the system puts the burden on the user to make sure that other images do not exist with the same name in the encyclopedia. (A natural design extension might be for the uploading interface to check for the names of images in the database before allowing a user to upload a new file.)

3.5 Image Copyright

Image copyright is a complex issue for Wikipedians because, as noted before, laws can be different in each country. In fact, a discussion of image copyright in Wikipedia and the Commons is a topic in itself, deserving much more in-depth coverage than the scope of this paper permits. Nevertheless, this section gives a brief overview of how copyright is dealt with in the Commons. Unlike the English Wikipedia site, the Commons does not permit users to upload “fair use” content. The reason for this limitation is that “fair use” is a legal concept conceived in the U.S. and, as such, it is not necessarily observed by other countries. Because the Commons is supposed to serve as a media

repository for all international Wikipedia sites, all content stored there needs to be free.

As noted above, contributors to the Commons are made to explicitly specify the license that covers each file they upload to the site. Wikimedia accepts a series of “Public Domain” licenses as well as “Free” licenses (including GNU, GFDL, and the Creative Commons licenses). For a complete list of licenses and descriptions, see [13].

Whenever the copyright standing of an image is unclear the picture may be put up as a candidate for deletion. Missing licensing information causes the file to be tagged as missing information and the person who contributed the file has seven days to correct the problem. After this period the file can be deleted by an administrator without further debate. Both the English Wikipedia site and the Commons have a page that lists all images that are candidates for deletion.

3.6 Image Categories

Because the Commons has a large number of files—over 500,000 media files when this paper was written—it is important to make it easy for users to find the images they are looking for. To that end, images can be tagged. Tagging an image automatically assigns a “category” to the picture. Like Wikipedia pages, images can have multiple categories. In fact, tagging is such an important aspect of the Wikimedia effort, that some users make “tagging images” their main duty on the site.

As an interface for finding tagged images, the “CommonSense” page was created [14], which helps users with guessing categories of images in the collection. The system also allows users to enter the name of an image file and see what categories it belongs to.

3.7 Featured Picture

Like “Featured Articles,” a “Featured Picture” (FP) is displayed every day on the front page of the English Wikipedia site. Featured articles and pictures represent what the community believes to be the best work in Wikipedia. Because they are showcased on the front page of the encyclopedia, these articles and images get high visibility. Materials are submitted to a peer review process before getting “featured” status and several candidates never make it to featured standing. As will be explained later, the FP procedure is one of the most important processes of community building around image creation in Wikipedia.

FP processes exist both on Wikipedia and in the Wikimedia Commons. However the criteria are different

on each site. On Wikipedia, FPs are supposed to be judged based on how helpful they are to illustrating concepts in the encyclopedia. In the Commons, on the other hand, the criteria are aesthetic: a “beauty contest,” in the words of one of the respondents in this survey. Another difference is that discussions about candidates to FP in Wikipedia contain a lot more posts/ commenting than their counterparts in the Commons. Decisions of FP in the Commons depends mostly on votes and no discussion. One of the main reasons for this difference is that the Commons counts on a highly international constituency, with users who do not share the same language—as opposed to the English Wikipedia, which is organized by users that speak a common language.

In Wikipedia, promotions to FP hinge on “majority consensus,” meaning that majority of votes do not automatically secure promotion. Voters need to explain their reasons for choosing to vote the way they do. The community has instituted this policy to prevent antisocial behavior on voting polls. So far, this strategy seems to be working.

FP promotion criteria are hotly debated on Wikipedia, with various editors pushing for “usefulness to article” to be the most important measure. In practice, however, users admit that FP archives may not reflect this ideal. For some, the ample dominance of photographs in these archives is a clear sign that usefulness is not the golden standard for promotion. These users argue that, if usefulness were indeed the decisive factor for promotion, diagrams and maps should be equally represented in the FP archives.

4. Survey

4.1 Method

A purposeful sample of active image contributors to Wikipedia was collected. Potential participants were selected from the list of users who had contributed images to the “Featured Pictures” page in the English Wikipedia [15], with the goal of obtaining a set of active users who had contributed a range of different image types: photographs, diagrams, and maps. Announcements of the survey were posted in 60 user talk pages¹, inviting participation. The survey was filled out by 29 respondents (48% response rate). The questionnaire was sent to participants via email. The survey consisted of 25 multiple-choice questions and 18 open-ended, essay questions. In addition, the author also interviewed one of the main FP editors.

¹ Every Wikipedia registered user gets a “user page” (their home page) and an associated “user talk page” that serves as a discussion forum. Whenever a person needs to get in touch with a given user, it is customary for this person to leave a note on the user’s “user talk page.”

4.2 Findings

4.2.1 Demographics: The races and genders of respondents in this survey were strikingly homogeneous: all respondents were male, 28 were Caucasian and one was Indian. The sample showed diversity in other respects, however. The average age was 33.6 with the youngest respondent being 16 and the oldest being 65 years old. When asked about current location, 48% of participants were located in Europe, 31% in North America (US and Canada), 7% were from Asia and the other 14% did not disclose their location. Profession wise, 34% held jobs related to computers (software engineers, IT professionals, etc), and 31% were students. Over half of respondents (62%) had been contributing images to Wikipedia for two years or more.

4.2.2 Kinds of Images Contributed: As can be seen on Table 1, the overwhelming majority of participants contributed photographs to the encyclopedia.

Over half of participants (52%) had contributed more than 200 images to Wikipedia and the Commons. The largest collection of pictures contributed by a respondent was made up of 1,200 regular pictures and 2,900 scanned images. Table 2 shows where participants were getting their images from.

In general, participants contributed images of nature—landscapes, animals, plants, flowers—people, locations, historical events and people, scientific imagery, maps, and technical diagrams—such as the workings of mechanical devices.

Table 1: Kinds of images contributed by respondents

Type of Image	% of participants	Type of Image	% of participants
Photographs	97%	Maps	48%
Diagrams	72%	Other ²	48%

Table 2: Sources of images uploaded by respondents

Image source	% of participants
I create images specifically for Wikipedia	93%
I modify my existing images for Wikipedia	78%
I donate my existing images to Wikipedia	70%
I find freely available images	40%

4.2.3 Equipment: The great majority of photography contributors said they used digital cameras to create their images for Wikipedia. Often these images would be retouched and corrected in an image-editing

² illustration, drawings, banners, flags

program. There was a vast collection of software utilized by participants. For photographic images, Adobe Photoshop was by far the most used, followed by GIMP, a freely available program for photo retouching, PTGui (a program for stitching photos into panoramic views), and Apple iPhoto.

For vector-based images such as illustrations and diagrams, participants used Adobe Illustrator, Corel Draw, Inkscape (an open source vector graphics editor), and Blender (a free software program for modeling and rendering three-dimensional graphics and animations).

In addition, four respondents said they wrote their own programs to create the diagrams and graphics they wanted to contribute. When asked whether contributing images to Wikipedia had caused them to purchase any equipment, 27% of participants said yes. Purchases consisted of: extra lenses, a tripod, new cameras, photographic gear in general, a licensed copy of Adobe Photoshop, and a new laptop with a stronger video card specifically for making maps for Wikipedia. One respondent said he had purchased high-resolution images for Wikipedia.

4.2.4 Motivation: When asked to rate how often they contributed images to Wikipedia because of a given reason, respondents said they most frequently contributed because they believe that information should be freely available (Table 3).

Tables 4 and 5 show that, on average, participants seem to balance their contribution of images and text to Wikipedia. In other words, most respondents do not “live in the world of images” but, instead, participate in Wikipedia in a larger sense.

Most participants (86%) had not heard of FPs when they started contributing images to Wikipedia. When asked what their main reason was for beginning to contribute images to Wikipedia, respondents listed as main motivations:

1. Having read an article that did not have an image but which could benefit from having one:

I guess I went to an article about a place I'd visited, and it didn't have an image. So I looked in my photo archive, and dug one out. I'm a firm believer in "a third-rate picture is better than none at all."

-- a user from Scotland

2. Having a better image than one that was being used on a Wikipedia article

The poor quality (or complete lack of) images attached to the articles on birds.

-- user from Canada

3. Wanting to make a contribution that has the potential to last for a long time

[The motivation was] that my interest in drawing nice illustrations could be used for something of long-lasting value.

-- a user from Sweden

4. Love for photography

I LOVE photography and looking at quality images. So illustrating WP (and not writing articles) was obvious for me. There is also the vanity of thinking that some of my images will outlast me and be seen in the far distant future, so that a little bit of me will outlive me.

-- a user from England

Except for “love for photography,” the other reasons reveal that image contributors to Wikipedia come to the site not necessarily as image creators but, rather, as consumers of the encyclopedic materials (i.e. articles). It is usually through reading articles that most contributors first feel compelled to add images to the site.

Table 3: Motivations for contributing images to Wikipedia

(frequency: 1 = never ↔ 5 = always)

Motivation	Frequency
I believe that information should be freely available	4
I found an article that I thought needed an image	3.7
I created an image for the intellectual challenge/stimulation	3.4
I wanted to show off my technical/artistic skills	2.8
I looked at the list of image requests or saw an image request on a page	2.2
I received a request from another person to create an image	2
Other	0.4

Table 4: Average percentage of time spent on image-related activities (+ range of responses)

Image-related activity	% of time	range
Creating/editing your own images	36%	95%-1%
Image-related discussions on talk pages	7%	35%-0%
Looking for publicly available images	3%	10%-0%
Image-related admin work	2%	20%-0%
Editing other people's images	1%	10%-0%
Verifying images' copyrights	1%	5%-0%
Helping out with the creation of image-related guidelines and policies	1%	5%-0%
other	1%	20%-0%

Table 5: Average percentage of time spent on non image-related activities (+ range of responses)

Non image-related activity	% of time	range
article editing	32%	81%-0%
discussions on talk pages	9%	40%-0%
admin work	5%	25%-0%
Helping out with the creation of guidelines and policies	2%	10%-0%
other	--	35%-0%

4.2.5 A sense of community? To find out whether image creators felt like they belonged to a community, the

last portion of the survey asked questions about the relationship between the respondent and other image contributors.

When asked whether seeing other people's pictures in Wikipedia affected their own image contributions to the site, 86% of participants said yes. Respondents described how they were affected in the following ways:

1. Source of inspiration

I'm inspired (and occasionally intimidated) by some of the very few folks who produce decent maps and diagrams.

-- user from Scotland

I get inspired by good illustrations I see at Wikipedia and try to reuse elements in them that I like (such as colors, overall style, etc.) when drawing my own illustrations.

-- user from Sweden

2. Opportunity for improvement

[An image by another contributor affects me] Only insofar as it is better or worse than an image I know I can collect. If "worse", then it's a target for replacement.

-- user from Canada

I often find images where the idea behind it is good, but the image itself is poor, e.g. low resolution, wrong choice of format, artifacts, bad colors or simply amateurish. This might make me want to improve it or if that's not possible, redraw it.

-- user from Sweden

3. Quality Standards

The bar for quality is constantly being raised, and valuable contributors like Fir0002 [another contributor] keep me interested in finding quality historic images to contribute.

-- user from the USA

4. The "less traveled" path

One can look to see what concepts are covered in existing images, and then try to fill in the gaps.

-- user from the USA

I try to contribute where it is needed, not where we have images already

-- user from the UK

5. Learning

I learned something about image-composition. Especially reading the Comments on the featured-Picture-Candidates-Page is very helpful. Taking Macroshots of insects is something I adopted from other Wikipedians.

-- user from Germany

Thanks to them I started learning Inkscape, for instance. Same goes for Blender.

-- user from Poland

6. Competition

Sometimes, I feel the need to out-photograph other users. It's sort of a friendly rivalry. Some people think this is selfish, but if it motivates us to take better pictures, I see no problem.

-- user from the USA

A little competition is good when working on the same subjects.

-- user from the USA

When asked if they felt like they could ask other image contributors for help, 96% of respondents said yes. Participants observed that often placing an image in

the "featured picture candidate" (FPC) list is a way of asking for help. Because the FPC list is a place that catalyzes discussion, respondents feel like it is one of the best forums for getting attention from a large number of fellow image contributors.

If an image is nominated for featured picture, it usually gains a lot of different feedback from people who are willing to address their concerns. [...] Besides from that I don't think there is a good forum for that kind of questions. There is actually a "picture peer review" (http://en.wikipedia.org/wiki/Wikipedia:Picture_peer_review) but it's low-traffic compared to the article peer review and the featured picture candidates.

-- user from Sweden

Some respondents also mentioned that they had sought help for identifying and categorizing the content of their images.

*I've quite often asked for help in identifying species of plants and animals. I'm often impressed by the quality of the replies, for instance [link to response to an identification question the user posted about a tree photograph he had taken]: *The tree is a Common Lime (a hybrid, *Tilia* × *europaea*; *Tilia platyphyllos* × *Tilia cordata*). This sort of twigginess is a normal character of some clones of this hybrid; it is vegetatively reproduced (propagated by cuttings, not seeds), and there appears to be a relationship between production of this twigginess and ease of rooting cuttings.**

-- user from the UK

Finally, a few respondents mentioned that they prefer to go outside of Wikipedia to ask for help on topics such as technical photography. These users felt they could get more expert advice from online forums centered specifically on photography. When asked whether they felt that their image contributions to Wikipedia were appreciated by others, 92% of respondents said yes. Participants pointed out different kinds of appreciative feedback:

1. Notes on users' talk pages

I had some nice comments on my talk page especially about my FP, like: "I would like to let you know, that I enjoyed that photo, - keep up the good work!". This gives a big motivation for continuing the work. But it isn't easy to make a FP, so not every Image gets such an appreciation. But all in all I think there is fair and motivating Feedback.

-- user from Germany

Feedback on Featured Picture Candidates is generally positive. I've occasionally received unsolicited positive comments [on my talk page], for instance: "I really liked your fish and chips photograph -- very dramatic. So I decided to look you up. Very nice job all around!"

-- user from the UK

2. Barnstars³

I got a barnstar from [User: JoJan] for illustrating the [leaf] article and then his [radula] creation. That was nice.

-- user from the USA

If you look at my user page you will see a Barnstar for photography from someone called "Neutrality," I very much appreciated that.

-- user from the UK

3. External recognition

The greatest compliment I have received was from a schoolteacher who printed a large version of my American Civil War Battles by Theater, Year diagram for use in the classroom.

-- user from the USA

I've received so much positive feedback, not to mention requests from people wishing to use my pictures in their articles, papers, or web publications.

-- user from the USA

4. Request for images

When I have done images in response to requests, I have got nothing but appreciation. But on the other hand, in those cases where I have just changed an image on a page and not responding to a request, I seldom hear anything.

-- user from Sweden

I had some positive messages, some people asked me to make pictures and the featured pictures show that people appreciate my work.

-- user from Switzerland

As can be seen from users' comments, succeeding in promoting an image to FP status generates considerable recognition from others in the community.

Belonging to a community usually means that one can recognize some other members of the group (as opposed to feeling surrounded by complete strangers). For this reason, participants were asked if they looked up the creators of images they liked. Respondents were also asked how often they recognized other image-contributors' names. 72% of participants said they looked up the authors of images they liked; with 68% of respondents saying they often recognized people's usernames. Username recognition was typically tied to one of the following circumstances:

- users have recognizable pictorial styles (this was especially true of diagram makers and illustrators)
- users shared a "niche" interest (for instance, most users working on geometry images seem to recognize one another)
- users contribute to smaller Wikipedia sites (for example, image contributors in the German Wikipedia often remarked they recognized each other's names)

³ A Wikipedia barnstar is an award given to a contributor in recognition of good work. Users can award barnstars to others by placing a barnstar image in the person's user talk page.



Figure 4: Windbeeches on the Schauinsland in Germany, photo by user Richardfabi for Wikipedia.

4.2.6 Image edits and vandalism: An open and straightforward editing architecture is one of the main reasons why wikis are such interesting collaborative platforms and, one could argue, a cause of Wikipedia's success. But, as remarked before, wikis are designed as collaboration tools around text, not images. With this in mind, the survey investigated the group dynamics of image editing in Wikipedia.

When asked how common it is for someone other than the image creator to edit a picture, 67% of respondents said it was uncommon. Users observed that the only situation when it is customary for an image to be edited by users other than its creator is when that picture is being considered for promotion to FP, that is, when the image is an FPC.

[Editing is] uncommon, although if it goes through one of the Featured Picture Candidates pages, it is more likely to get edited by another editor – sometimes for the better.

-- user from the UK

Far more common when going through the featured picture candidate process. Often, images that are submitted can be improved quite easily and there is a large number of regulars who are willing and able to do it.

-- user from Australia

Respondents remarked that most edits are usually "trivial" corrections such as fixing contrast levels or the size of an image. There was a sense that, beyond an initial level of editing, creators might be better off producing a new image instead of continuing to edit the same file.

Except for "high profile" images – 'featured picture' – it's probably uncommon. The Wikipedia interface is text-oriented; [...] editing an image requires quite a bit more effort, comparatively. For people who can create a new image, it is probably more interesting to them to do just that instead of editing an existing image.

-- user from Canada

I think it is very rare for images to be substantially edited by other users. The major exceptions in this are issues relating to cropping or image size (usually in the form of adding higher resolution versions of images), tinkering with color and contrast levels (usually for images which need the tinkering because they are poorly scanned or are otherwise washed out), or non-aesthetic manipulations (optimizing of PNG files, for example, which lowers the file size but does not change the visual content).

-- user from the USA

When asked if they had ever edited someone else's image, 89% of participants said yes. Once again, respondents explained that, for the most part, edits were minor. Typical edits include changes in white levels, contrast, brightness, sharpness, color balance, cropping, and translating an image description to other languages. In addition, 68% of participants said their images had been edited by other Wikipedians. Here too, most edits were minor optimizations such as the one mentioned above. The only major edit mentioned by a respondent, occurred to the image on Figure 4:

My picture [link to Fig 4] was enlarged. The composition was not perfect, and another user added more ceiling and ground on the right side. I must admit, I don't know how he made it, but it contributed a lot to the Image-Composition.

-- user from Germany

An important aspect of collaborative image editing is for editors to have access to pictures in "raw" formats that lend themselves more easily to being changed. For instance, vector-based files of maps and illustrations can more easily and effectively be edited than a JPG or GIF (rasterized) version of the same images. When asked whether they posted editable/raw versions of their images on Wikipedia, 35% of respondents said they did so. Most of these users referred to the creation of Scalable Vector Graphics (SVG) files for maps and diagrams.

Yes, without any exceptions, this one is important for me. When I first started with Wikipedia I was rather surprised that this wasn't more common than it was. Without providing the source data you'll make it harder for other Wikipedians to improve and reuse your work, and this is especially true for diagrams, maps, etc. Then, around August (if I recall correctly) last year, Wikipedia implemented SVG (Scalable Vector Graphics) support. This was a great improvement in many aspects. Not only could you now provide drawn illustrations that could be used at any resolution without loss in quality, but you also (automatically) included their source. This makes modification easy and allows people to use the drawings in high-quality reproduction (like printing). Nowadays I only upload drawings in SVG format. I've seen a number of contributors whose only work is to turn bitmapped illustrations into vector format. If I create the illustration in another format than SVG and then convert it to SVG, I always provide the original source along with conversions instructions.

-- user from Sweden

I usually try to convert all of my vector art to SVG format (all new vector art I upload as SVG; when I get the time I convert old vector art of mine to it as well and replace the old rasterized versions). I do this both to allow easy editing/translation should someone want to do it, and to allow re-users to scale the image to an arbitrary size (if Wikipedia can offer high-quality vector art from its encyclopedia at no cost to re-users, that is something that no other encyclopedia service can compete with in the slightest). [...] When I make graphs in Excel, I always paste a copy of my dataset to the discussion page of the image, so that others can re-create the graph if they want and/or check the data.

-- user from the USA

The flip side of an open system for collaborative editing is the possibility of vandalism. As noted in previous studies, Wikipedia often encounters vandalism in its text entries [1][7][9]. This survey asked participants whether they had ever witnessed vandalism on Wikipedia images. Over half (58%) of respondents said they had seen acts of vandalism involving images. Most of these occurrences were image switches, where the original file had been replaced with an offensive image.

Most times the vandal simply replaces the image with something shocking or funny. It's rare to see a vandal that has made the effort to actually add to or remove something from the original image. I think vandalism of articles is a bigger problem than vandalism of images. One reason might be that both more knowledge and effort is needed to vandalize a picture. Another reason is probably that anonymous and new users are not allowed to replace images on Wikimedia Commons (where a substantial amount of images are placed). An often requested feature is that watchlist notification should be implemented for image replacements -- that would decrease the time to detect and revert image vandalism.

-- user from Sweden

One very rare, but very effective, kind is to upload an animated GiF, with only two frames. The first is the innocuous and appropriate image, the second the abusive vandal image. They put a long delay (10 mins) between frames.

-- user from Scotland

The other form of image-related vandalism mentioned by respondents were changes made to a picture's caption or description.

5. Discussion

As it grows, Wikipedia is becoming a more complex site. Part of Wikipedia's evolution is the emergence of an active community of image contributors. At the same time that these contributors enrich the site with impressive pictures and informative diagrams, they help push the boundaries of wiki technology.

Collaboration around images presents a series of challenges for wiki adopters. The technical infrastructure

needed to support image editing is completely external to wiki platforms, which means that several key aspects of wiki collaboration features are not available to image creators at present. For instance, whenever images are edited, their versions are not kept on the Wikipedia site. This lack of public versioning history is a key difference from how text gets edited on wikis and it carries critical consequences to users' ability to engage in collaborative image editing. By not being able to easily revert back to earlier, public versions of pictures, image contributors do not experience the same level of flexibility that text editors encounter in a wiki site. Interestingly, respondents in this survey expressed a belief that images are more "personal" than text and, therefore, not as readily amenable to massive collaborative editing. This sentiment begs the question: do image creators feel this way because pictures are inherently more personal than text, or because they do not have a flexible platform on which to engage in collective editing? How would image editing change in Wikipedia if every version of an edited image were made public and persistent?

One of the main lessons from this survey is the importance of community building around image contributions. Even though image editing is still somewhat of an isolated activity, respondents explained that seeing each other's contributions highly affected how they contributed to Wikipedia. Moreover, as important as the FP page is for showcasing the best work on the site to visitors, its significance for respondents lies in its function as a catalyst for the community. The FPC serves both as a learning place and as a forum in which to look for help.

6. Conclusion

This paper describes the visual side of Wikipedia and presents the results from a survey of image-contributors to the encyclopedia, one of the fastest growing communities in the site.

This survey reveals that image contributors mostly create the images they contribute to the site, with the majority of these contributions being photographs. Around one third of respondents said they had purchased equipment and tools to help contribute images, showing a high degree of motivation.

The strongest reason for this motivation, according to the survey, is the idea that information (including visual materials) should be freely available. Following that, respondents said that their main incentive was the realization that a given article needed an image. This suggests that image contributors are attracted to the

encyclopedia as readers first; only afterwards feeling compelled to create images for the site.

Because it is not natively supported by wiki technology, image editing in Wikipedia is a more isolated activity than is text editing. One of the reasons for the low incidence of collective image editing may be due to the lack of persistent versioning history for images on the site. (The flip side of the image-editing limitation is the low frequency of vandalism on pictures.) Despite the potential for isolation, image contributors still share a sense of community. The FPC page serves as the "public square" of the image creation community. This is the place where users come together to discuss changes that ought to be made to pictures, to learn from each other, and to ask for help.

How well wiki technology can be tailored to support collaborative image editing remains an open question. Nonetheless, both the creation of the Commons and the recent implementation of SVG support are evidence of Wikipedians' commitment to turn images into first-class wiki citizens.

7. Acknowledgments

I am grateful to Martin Wattenberg for his helpful and insightful comments. Many thanks to Jesse Kriss for running image queries on a dump of Wikipedia data.

8. References

1. Bryant, S., Forte, A., Bruckman, A. *Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia*. Proceedings of GROUP 2005.
2. Emigh, W., Herring, S. *Collaborative authoring on the Web: A genre analysis of online encyclopedias*. Proceedings of HICSS-38, 2005.
3. Forte, A., Bruckman, A. *Why do People Write for Wikipedia? Incentives to Contribute to Open-Content Publishing*. GROUP 05 workshop position paper.
4. Giles, J. *Internet encyclopaedias go head to head*. In Nature, 14 Dec. 2005. <http://www.nature.com/news/2005/051212/full/438900a.html>
5. Lih, A. *Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media as a News Source*. Fifth International Symposium on Online Journalism, 2004.
6. Stvilia, B., Twidale, M., Gasser, L., Smith, L. *Information Quality Discussions in Wikipedia*. Technical Report ISRN UIUCLIS--2005/2+CSCW, 2005.
7. Viégas, F., Wattenberg, M., & Dave, K. *Studying Cooperation and Conflict between Authors with history flow Visualizations*. In Proceedings of SIGCHI 2004.
8. Viégas, F., Wattenberg, M., Kriss, J., & van Ham, F. *Talk Before You Type: Coordination in Wikipedia*. Proceedings of HICSS 40, 2007.
9. Voss, J. *Measuring Wikipedia*. In Proceedings 10th International Conf. of the International Society for Scientometrics and Infometrics, 2005.
10. <http://en.wikipedia.org/wiki/User:Kils>
11. <http://meta.wikimedia.org/wiki/Help:Formula>
12. <http://meta.wikimedia.org/wiki/EasyTimeline>
13. <http://en.wikipedia.org/wiki/Wikipedia:Copyrights>
14. <http://tools.wikimedia.de/~daniel/WikiSense/CommonSense.php>
15. http://en.wikipedia.org/wiki/Wikipedia:Featured_pictures
16. <http://stats.wikimedia.org/EN/TimelinesIndex.htm>