

# Crowdsourcing – The Real People behind the Crowd

Nicole Graf  
ETH Library, Image Archive  
ETH Zurich, Ramistrasse 101, 8092  
Zurich, Switzerland  
[nicole.graf@library.ethz.ch](mailto:nicole.graf@library.ethz.ch)

Following the successful crowdsourcing project involving Swissair retirees (2009–2013), ETH Library's Image Archive set up a comments function on its image database at the end of 2015. On 18 January 2016, the *Neue Zürcher Zeitung* (NZZ) reported on it and triggered a huge response in both the media and the general public, who are now helping to identify images. This article focuses on our experiences with public crowdsourcing in relation to the amendments on the image database, the volunteers' approach, the internal editing workflow and the communication with the volunteers. Community care is a key factor for successful crowdsourcing initiatives. The most important instruments tested in the image archive are presented and discussed. In order to get to know the unknown crowd better, an online survey on the motivation, work techniques and sociodemographic background among the crowd was conducted in early 2017. The results of this survey will be showcased in the presentation.

*Crowdsourcing. Online survey. Images. Image Archive.*

## 1. INTRODUCTION

ETH Library's Image Archive first experienced a crowdsourcing project between 2009 and 2013 when former Swissair employees volunteered to help index around 45,000 images from Swissair's photo archives online. The author referred to it as expert crowdsourcing as a "closed" group of experts were specifically addressed. The major advantage among the Swissair retirees was that they were well organised and had various communication channels (their own magazine, meetings, etc.), which enabled them to be contacted directly (Graf 2016, Graf 2013).

Given the success of the Swissair project, it was clear to us that we wanted to continue to draw upon the knowledge of many.

## 2. OPEN CROWDSOURCING

The most efficient way to identify untitled images is by the "swarm". Naturally, the Image Archive's six-strong team has limited knowledge. Under the banner of Do you know more? we launched the comments function on our image database ETH E-Pics BildarchivOnline (<http://ba.e-pics.ethz.ch>) on 9 December 2015. In principle, this means that

comments can be left on any image. In the age of Web 2.0 applications, a comment's function no longer really constitutes an innovation in itself. We should also add that our digital asset management system Canto Cumulus is not geared towards user participation. Instead, it focuses on integrated internal media management and production.

### 2.1 Do you know more? Amendments on the image database

However, the comments triggered a number of changes in the Image Archive. We have adjusted our image database in certain places. For instance, we make use of the motto *Do you know more?* at various points on the database to achieve a familiarity effect among the users and run it as intuitively as possible through the new function. The welcome page features the following in bold letters:

Do you know more about an image? Then please get in touch! The category Do you know more? involves images that we are unable to describe or date accurately. You will find our email function for your feedback and any questions and previous comments in the image information.

Please also indicate whether and how we might quote you in the comments.

We set up a new category under the motto *Do you know more*, which is on the same hierarchical level as the holdings or keywords. We specifically divide images that need to be identified partially or fully into general subcategories such as buildings, mountains, places, etc.

In the information view, we added a new separator right at the very bottom that is also entitled *Do you know more?* Underneath, there are the new metadata fields Feedback, Questions and Comments. The feedback field is a form that automatically generates an email for the Image Archive with the subject line Feedback on image [image code]. In the questions field, the Image Archive team can address concrete questions to the volunteers if need be. And in the comments field, the volunteers' comments are published by name with their explicit consent.

## 2.2 Unexpected media response

The launch of the comments function on 9 December 2015 was initially announced in a news message on ETH Library's homepage. To our surprise, our first feedback landed in the inbox two days later. A month on, we had received 100 emails.

At the beginning of January 2016, Adi Kälin, a journalist from the *Neue Zürcher Zeitung* (NZZ), was searching for images for an article in our image database when the appeal on our welcome page grabbed his attention. He got in touch and interviewed the author on the topic of crowdsourcing. Then, on Monday 18 January 2016, the article *Wer kennt die Berge, Orte und Fabriken?* about the Image Archive and the crowdsourcing project was published (Kälin 2016b). The online version even contained ten unidentified aerial photographs by Walter Mittelholzer. The report was met with such a huge response among readers and the media that, within one and a half days, nine of the ten images had been identified! Two days later, based on one image where an Italian palazzo was pinpointed in a suburb of Zurich, the NZZ published a follow-up article on the building identified and the crowdsourcing campaign (Hotz 2016).

The very same Monday morning, *Tagesschau*, the Swiss news programme on the television channel SRF, got in touch. By the early afternoon, a bulletin on crowdsourcing had already been filmed in the Collections and Archives reading room at ETH Library and was broadcast in the main programme that evening (SRF 1 TV 2016). The day after, Swiss radio also expressed an interest in the topic and published unidentified aerial photographs on its own homepage. Throughout the week, it kept broadcasting information on the picture puzzle campaign (SRF 1 Radio 2016b). A week later, Swiss

radio went even deeper into the topic. The author and the most prolific volunteer, Walter Aeberli, presented their own take on crowdsourcing on the live programme *Treffpunkt* (SRF 1 Radio 2016a). Evidently, this is an interesting new playing field for the media: they are trying to encourage their audiences to actively get involved while testing out whether this could help boost customer loyalty.

## 2.3 Internal Workflow

We were extremely taken aback by the major media response and the countless comments emails. We were not prepared for this flood of emails in the slightest, nor had we defined a workflow to cope with them. We answered the first 100 emails during the first month virtually on the fly. Only the link between the comments function and our general email account ([bildarchiv@library.ethz.ch](mailto:bildarchiv@library.ethz.ch)) was automated. The day the NZZ article came out, we literally watched the email subfolder "Feedback", to which the comments emails were automatically forwarded, fill up. We received 1,332 emails in the first week!

In order to get to grips with the deluge of emails and status of the project, first of all we structured the email folder by creating corresponding subfolders in the email account. After several attempts, the following structure proved viable: in the *Feedback* file, all emails with "feedback" in the subject line are filed automatically using a rule. It also contains the subfolder "Feedback – completed", to which the completed comments emails are moved. Only the initial emails are filed here as they simultaneously form the basis for the statistical evaluations generated manually in MS Excel.

In processing the first thousand emails, the biggest challenge was the fact that we had to ask virtually every individual for permission to publish their comments along with their name. Although there was a note to this effect on the welcome page ("Please indicate whether and how we may cite you in the comments."), it was the very last sentence and probably overlooked by most volunteers. Consequently, we set up a "Comment – yes" and "Comment – no" sub-folder and filed the answer emails accordingly.

When new comments emails come in, first of all we need to verify whether someone has indicated their consent or unwillingness. Every email is categorised accordingly. Finally, all further emails that follow the initial comments email are filed under "Correspondence". All general emails for the staff are filed in the "Z\_instructions" folder right at the bottom.

Initially, we were unable to respond to every email. During the first two weeks, we worked on

processing the flood of emails with one and a half fulltime positions. We then organised set days, with four of the six members of staff at the Image Archive sharing the five working days between them – not forgetting that the inbox was always fuller on Monday morning than the other days of the week.

The workflow was and is improved when necessary. For instance, an automatic reply is sent to every comments email received, thanking the volunteer for their help and, if it is the first time they have contacted us, asking for their consent to publish the comment with their name.

#### 2.4 Volunteer Searches

Around 375,000 images are currently available online on the image database. There are three typical approaches for the volunteers to select the images. The first point of entry is our category *Do you know more?* with the various subcategories, in which we deposit particular images where the title is missing or too unspecific. Many, however, embark on their own voyage of discovery on our database by studying images of the place where they live or holiday, and provide more details or inform us of erroneous image titles. Some volunteers have a specialist area, which they then (partly) work through systematically. They seek specific tips from us as to how they can pinpoint these images more easily on the database. We frequently received lists with hundreds of improved titles from these specialists. All in all, it is evident that the volunteers took to the comments function extremely well.

The Image Archive team subject the pointers from the volunteers to a plausibility check. Many volunteers also send “evidence”, whether it be in the form of a link on Google Maps, an image or a personal story. The information is then entered in the corresponding metadata fields – usually the title, description and date – and, if need be, the keywording is amended.

The original comment with the first and surname is copied into the comments field manually. It is a normal text box where all comments are entered in chronological order. It is not yet possible to display further information that is already standard in numerous Web 2.0 applications, e.g. the exact date of the comment, in our digital asset management system Cumulus by Canto and would be too much work manually. If anyone would not like to be named, we will publish the comment as “Anonymous”.

Another amendment requested by the volunteers was to assign the improved images to a specific category, too. And so we created the category *You*

*knew more! Thank you!* As with *Do you know more*, it is divided into subcategories. For their part, the volunteers verify on the image database whether and how we have included their comments. Consequently, we also integrated the comments field in the full text search – again at the volunteers’ suggestion. If anything is missing or entered incorrectly, we are immediately informed.

Since January 2016, we have been spending an average of 3.2 working hours a day on processing the emails we receive. Usually, the emails are processed within 48 hours. We invariably receive substantial and valuable information from the volunteers, which helps identify our images or improve the metadata. So far, we have not had any comments along the lines of, “Oh, what a pretty picture!” or that were offensive. The former type of comment appears surprisingly often on Flickr Commons. The high-quality added value of the contents therefore more than justifies the effort to which we go.

For some images, we were given several identical or even contradictory pieces of information that could nonetheless be resolved in the discussion (Fig. 1): for instance, our regular user Walter Aeberli had been puzzling over this aerial photograph by Walter Mittelholzer for years. Thanks to the appeal in *NZZ*, we received further suggestions, such as the Rhine Gorge near Flims, the Saanegraben or the Sensegraben in the Sense District (Canton of Fribourg). Mr Aeberli checked all the pointers and concluded that this was unlikely due to the number of river loops, the presence of a dam wall and a reservoir at the top of the picture. Moreover, Mittelholzer was flying over Lake Constance the day the photograph in question was taken. Thomas Scheidegger eventually solved the puzzle: between Flawil and Gossau (Canton of St. Gallen), reservoir in Glatt-Tobel near the ruins of Helfenberg Castle. In other words, a journey right across Switzerland.



**Figure 1:** Mittelholzer, Walter: Flawil-Gossau, reservoir in Glatt-Tobel near the ruins of Helfenberg Castle, 17.08.1932 (LBS\_MH01-006949, <http://doi.org/10.3932/ethz-a-000298816>)

Meanwhile, the volunteers have also got to know each other so well that one volunteer actually asked us to have his comments verified by another volunteer in particular before they were published!

In the end, even one of the oldest photographs in our collection could be identified (Fig. 2): the image only joined our collection last summer – merely entitled *Lucerne*. The Lucerne-based steamer historian and author Josef Gwerder (aged 85), who goes through many of the Lucerne pictures, managed to spot a lot more here using the zoom function on the image database. We were able to enlarge on the title of the photograph as follows: “Lucerne, steamer ‘St. Gotthard’ at the Schwanen jetty.” And as the description of the image now states: “The sign on the Zurgilgenhaus (with tower) that says *Pension Woerley* was only erected in 1863; it was previously *Pension Aeschmann*. The steamer on the opposite bank in front of Villa Nager-Knörr is the ‘Winkelried I’ (blurred).”

Moreover, the image could be dated to 1864, making it one of our oldest photographs of all.



**Figure 2:** Braun, Adolphe: *Lucerne, steamer “St. Gotthard” on the Schwanen jetty, around 1864 (Ans\_09918, <http://doi.org/10.3932/ethz-a-000861831>)*

## 2.5 Communication with the Crowd

A month after the *NZZ* article was published (no email had been answered personally during this period), we thanked all the volunteers with a summary of the project’s status. This triggered a new deluge of emails almost directly! In the days following the thank-you email, we received significantly more emails than in the days leading up to it. This showed us that the crowdsourcing community is grateful for information. In order to communicate with the volunteers – and not merely via a monthly circular – we decided to set up a weblog. On 9 May 2016, our blog *Crowdsourcing at ETH Library: news and experiences from the community* (<https://blogs.ethz.ch/crowdsourcing>) went online.

The blog is geared towards both the volunteers and professionals. Here, too, we use the motto *Do you know more?* to guide our users intuitively through the navigation. Every Monday morning, we present images that need to be identified or have their metadata improved in *Do you know more?* Then, on Friday morning, an article with newly identified images follows under the heading *You knew more!* One key criterion for a blog’s success is to publish new posts regularly.

The posts are also distributed via ETH Library’s Twitter account, which has revealed that specific hashtags in the tweets arguably lead to higher access numbers on the blog and bring in new volunteers. As soon as an additional publication medium or specific target group becomes aware of a topic, the figures rise. However, the access figures also indicate that the community has understood the publication rhythm: they are highest on Mondays and Fridays. Moreover, we publish key statistics at the beginning of every month: cumulative number of pointers received and identified images, number of pointers per day, number of volunteers and the top ten volunteers with the number of images processed. The top ten list is bound to spur on one or two of them even further.

A press review and information on other crowdsourcing projects round off the blog. Selected entries are also available in English. Penning two posts per month is very time-consuming as they take around half to a full day a week to write. Mostly, however, it is an opportunity to delve into a topic in more detail as the authors often have to go searching in the holdings themselves. Before posting something, they usually check the information again. On the image database, the volunteers eventually correct each other or themselves. But once something has been posted, you want to be safe in the knowledge that it is correct.

Ultimately, we wanted to get to know the crowd in person, express our thanks and appreciation, and make it easier for them to connect with each other. We invited all the volunteers to a meeting with a presentation and drinks in ETH Zurich’s main building on the Mitte campus in mid-November 2016. 105 people signed up and attended. A half-hour talk by the author focused on the appreciation of their work and the Image Archive’s internal processes. Suggestions for improvements from the crowd were implemented the very same evening, and the subsequent drinks reception was used as a welcome opportunity for discussions amongst themselves, but also to meet the Image Archive team.

NZZ subsequently reported on the event (Kälin 2016a) and Ticino's television station broadcast a feature on it in the evening news (RSI TV 2016). All this in turn led to the crowd being more motivated than ever. The monthly statistics rose significantly. A comparison of January and February 2016 and 2017 reveals that we have already reached 30 per cent of the mails received the whole of last year in 2017. Since the event, the processing time has also been significant, rising to 4.1 hours (the overall average since the beginning of the project is 3.2).

Finally, from March 2017 a short video series on crowdsourcing will be released as ETH Library starts producing brief films on various topics. *Crowdsourcing* will consist of eight episodes: about the Image Archive itself, the crowdsourcing event in November 2016, plus six testimonials on activists in the top ten.

## 2.6 Who's the Crowd?

We have been compiling statistics since we launched the comments function. As of March 2017, a total of 742 volunteers were involved in crowdsourcing, 91 per cent of whom are men. We have received 15,232 pointers, which has enabled us to enrich and improve 12,677 images. By March 2017, a very active top ten had sent 64 per cent of all information. The most industrious crowd activist alone described a whopping 1,734 images! The top 30 processed 81 per cent in all. The person in thirtieth place alone sent 50 pointers. 340 people contributed just the once, while another 312 described between two and ten images. Many volunteers have a specialist topic that they process: mountains, aerial photographs, steamers and the railway industry, local ties, power stations etc. This means that a small number of volunteers are constantly processing images. Other projects also report on such "power users", who comment regularly and on the most images. In the case of the Library of Congress's *Flickr Commons* project, there are 20 such people (Springer et al. 2008); for our Swissair project, half a dozen.

As we only know very little about the sociodemographic distribution and motivation of the volunteers, we conducted an online survey among our crowd in January/February 2016. For instance, we wanted to find out whether they were predominantly retirees. The (somewhat distorted) impression could also be based on the people who attended the November event. What is the activists' educational background? What work environment do they come from? Do they have ties to ETH Zurich? What is their motivation and how much time do they invest in our images? Are they involved in any other projects? Our survey contained 46 questions and could be completed online from 9 January to 4 February 2017. We

contacted all the volunteers via email and advertised it on our blog. More than 350 people clicked on the link and we ended up with 193 fully completed questionnaires. By the time this article went to press, the precise evaluations had not been completed; these will be presented in July.

## 3. OUTLOOK

Without the NZZ article, 2016 would probably have gone very differently at ETH Library's Image Archive. The responsibilities of the Image Archive staff were expanded accordingly. Resources that could be saved by simplifying their usage (the image database was opened for free downloads on 1 March 2015 – referred to as open data) were shifted to crowdsourcing. The Image Archive is currently contemplating expanding image crowdsourcing through georeferencing.

## 4. REFERENCES

Graf, N. (2013) Crowdsourcing – Neue Möglichkeiten und Grenzen für Bildarchive. *EVA 2013 Berlin, Berlin: Gesellschaft z. Förderung angewandter Informatik*, pp. 96–99.

Graf, N. (2016) Crowdsourcing: Die Erschließung des Fotoarchivs der Swissair im Bildarchiv der ETH-Bibliothek, Zürich. *Rundbrief Fotografie*, 23(1), [N.F. 89], pp. 24–32.

Hotz, S. (2016) Ein Florentiner Palast in Opfikon. *Neue Zürcher Zeitung*, 20.1.2016. <http://www.nzz.ch/zuerich/ein-florentiner-palazzo-in-opfikon-1.18680173> (retrieved 22 March 2017).

Kälin, A (2016a) Crowdsourcing des ETH-Bildarchivs: Freiwillige haben 5751 Bilderrätsel gelöst. *Neue Zürcher Zeitung*, 17.11.2016. <http://www.nzz.ch/zuerich/crowdsourcing-des-eth-bildarchivs-freiwillige-haben-5751-bilderraetsel-geloest-ld.128710> (retrieved 22 March 2017).

Kälin, A. (2016b) Wer kennt die Berge, Orte und Fabriken? *Neue Zürcher Zeitung*, 18.01.2016. <http://www.nzz.ch/zuerich/wer-kennt-die-berge-orte-und-fabriken-1.18678913> (retrieved 22 March 2017).

RSI TV (2016) Foto in cerca di autore, RSI, Radiotelevisione Svizzera, *Telegiornale*, 18.11.2016. <http://www.rsi.ch/play/tv/telegiornale/video/18-11-2016-foto-in-cerca-dautore?id=8336419> (retrieved 22 March 2017).

Springer, M. et al. (2008) For the Common Good: The Library of Congress Flickr Pilot Project, 30.10.2008.  
[http://www.loc.gov/rr/print/flickr\\_report\\_final.pdf](http://www.loc.gov/rr/print/flickr_report_final.pdf)  
(retrieved 22 March 2017).

SRF 1 Radio (2016a) Das Bildarchiv der ETH bietet Bilderrätsel der besonderen Art. *Treffpunkt*, SRF 1, Schweizer Radio, 26.01.2016.  
<http://www.srf.ch/sendungen/treffpunkt/das-bildarchiv-der-eth-bietet-bilderraetsel-der-besonderen-art> (retrieved 22 March 2017).

SRF 1 Radio (2016b) ETH-Bildarchiv: Kennen Sie dieses Bild? *Tagesprogramm*, SRF 1, Schweizer Radio, 19.01.2016. <http://www.srf.ch/radio-srf-1/radio-srf-1/eth-bildarchiv-kennen-sie-dieses-bild> (retrieved 22 March 2017).

SRF 1 TV (2016) ETH löst Bilderrätsel. *Tagesschau*, SRF 1, Schweizer Fernsehen, 18.01.2016.  
<http://www.srf.ch/play/tv/tagesschau/video/eth-loest-bilderraetsel?id=d00b8618-9e72-49d1-a10b-b1c158a83977> (retrieved 22 March 2017).