



## Open Data for the Crowd: An Account of Citizen Science at ETH Library

**Stefan Wiederkehr**

ETH Library, Zurich, Switzerland

[stefan.wiederkehr@library.ethz.ch](mailto:stefan.wiederkehr@library.ethz.ch), [orcid.org/0000-0003-1339-0688](https://orcid.org/0000-0003-1339-0688)

### Abstract

This paper gives an account of ETH Library's combined strategy of implementing an open data policy and using crowdsourcing to improve metadata. Both activities go hand in hand and promote each other. ETH Library's Image Archive was the first unit to provide content for free download in high resolution. This paved the way for a very successful crowdsourcing campaign during which citizen scientists located places, dated photographs, and identified people and artefacts. This positive experience led to further crowdsourcing activities in other ETH Library units. Important conditions for success apart from the open data policy were conscious community management by social media channels and promoting competition within a gamification approach.

**Keywords:** ETH Zurich; ETH Library; crowdsourcing; open data

### 1. Introduction

Citizen science has been discussed by policy makers and library associations as one important aspect of a global Open Science strategy. (Ayrís & Ignat, 2018; Ayrís et al. 2018a; Ayrís, López de San Román, Maes, & Labastida, 2018b; European Commission, 2018). A number of authors have highlighted different aspects of citizen science and proposed a taxonomy of its forms

(Haklay, 2013; Strasser & Haklay, 2018). This paper gives an account of one specific expression of citizen science at ETH Library. It presents the combined strategy of implementing an open data policy and using crowdsourcing to improve metadata, the latter corresponding to level one of participation and engagement according to Haklay's typology (2013), or to a microtask in The Daily CrowdSource's taxonomy (Simperl, 2015, p. 21). It is argued that these two activities promote each other, making citizen science a success story.

The first part outlines the basics of open data at ETH Library. The second part traces the beginning of crowdsourcing at ETH Library's Image Archive and the expansion of citizen science activities to other ETH Library units (Map Collection, Collection of Astronomical Instruments, ETH Zurich Art Inventory, ETH University Archives). The account focuses on the conditions of success.

## **2. Open Data at ETH Library**

ETH Library has adopted an open data policy (Kyburz, 2017). This means that it renders bibliographical metadata and digital copies publicly accessible and reusable, provided that this is not opposed to by any third-party rights. Open data is consonant with the Open Science paradigm, which, according to the most common definitions, also includes Open Access, Open Source, Open Educational Resources, Open Peer Review, and Open Lab books. At ETH Library, open data pursues the following objectives (ETH Library, 2019a):

- Open licence: whenever possible, ETH Library makes its data available using the public domain mark or a CC0 licence. If the prerequisites for this are not given, an open CC license is used.
- Transparency: ETH Library indicates reliable re-use transparently for each dataset.
- Currentness: ETH Library regularly updates variable datasets.
- Freedom from discrimination: ETH Library does not restrict access to the data. The data is available to anyone at any time and without registration.
- Free download: ETH Library's data is free to obtain.
- Machine readability: ETH Library provides its data in an open and, whenever possible, machine-readable standard format.
- Availability: ETH Library provides its data via a suitable interface or platform.

Obviously, this is an adaption of the definition provided by the Open Knowledge Foundation (Open Knowledge Foundation, 2019). ETH Library promotes open cultural data. It has participated in all Swiss open cultural data hackathons so far, being a co-organiser of the 2018 edition (OpenGLAM.ch, 2019). It uploads photographs to Wikimedia Commons (Gasser, 2017) and contributes datasets to the Swiss open government data platform *opendata.swiss* (*opendata.swiss*, 2019). Finally, it supports ETH Zurich researchers and offers advice on matters of open research data and following the FAIR principles (Töwe, 2018). This, however, is beyond the scope of this paper.

### 3. Crowdsourcing at ETH Library

#### 3.1. Starting Point: The Image Archive

ETH Library has an extensive and historically valuable collection of more than three million photographs, postcards, aerial pictures and portraits in its Image Archive. More than 500,000 photographs and images have been digitised and can be searched for on the platform *e-pics Image Archive Online* (ETH Zurich, 2019).

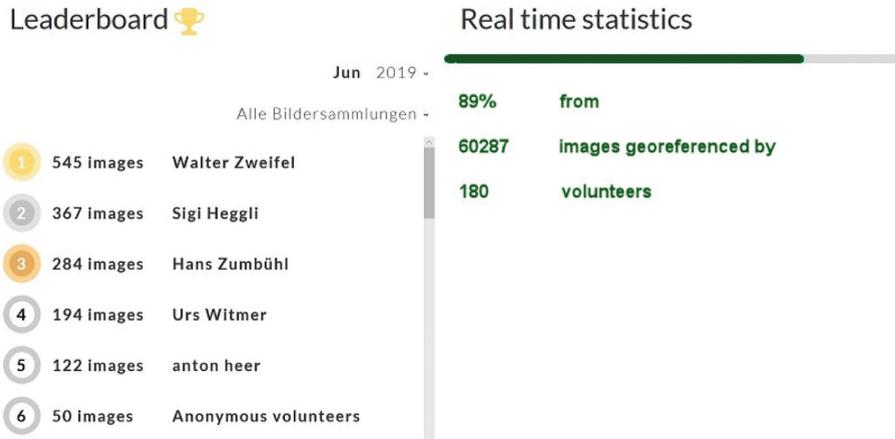
The Image Archive was the first unit to start a crowdsourcing campaign. After Swissair, the Swiss national airline, went bankrupt in 2001, ETH Library happened to acquire the Swissair Photo Archives. In 2009, former employees of Swissair were invited to improve metadata. They were given the chance to share their knowledge on around 45,000 images in an easy way, which this determined group of experts accepted enthusiastically (Graf, 2016). In this project, the volunteers took part at level one and partly also at level two of Haklay's typology, i.e. citizen scientists participated as sensors and performed basic interpretations (2013, p. 116).

The Image Archive was a pioneer in terms of implementing an open data policy, too. At the beginning of 2015, it changed its business model. Instead of trying to sell high-resolution images, which had not turned out to be profitable, it started to offer its photographs for free download, even in the highest resolution available and for commercial use whenever there were no third-party rights. As a result, resources used to fulfil orders, issue invoices and monitor accounts were freed for more forward-looking activities (Graf, 2015).

Given the outcome of the Swissair project, the Image Archive opened its database for general user comments regarding all images in December 2015. On January 18, 2016, *Neue Zürcher Zeitung*, a leading Swiss newspaper, published an article before the marketing campaign had even started (Kälin, 2016). This triggered a breakthrough, which resulted in a report being broadcast on Swiss television's main news programme the very same evening and elicited an extensive media response in the days and weeks that followed (Graf, 2017). In the wake of this unexpected promotion by the media, the Image Archive started the blog *Crowdsourcing: News and Experiences from the Community* (ETH Library, 2019b). Since May 2016, it has published quizzes and appeals inviting readers to get involved. The crowd was asked to help describe images on Mondays, and the feedback received was documented on Fridays. Rankings and community events stimulated crowd participation. All in all, 1,109 volunteers had improved the metadata on 68,488 images with 70,555 hints by December 1, 2019. The distribution of work among the volunteers confirms earlier findings from all over the world: a small minority did the majority of the work (Holley, 2010). One reason for the affirmative response is the fact that the volunteers were able to download their preferred images in high resolution and re-use them thanks to ETH Library's open data policy, and the feedback function's user-friendly design. While no systematic collaboration with the media existed, such as that in Denmark (Overgaard & Kaarsted, 2018), the Swiss media crucially supported the Image Archive's efforts by taking up the topic and encouraging citizens to participate in the project.

In 2018, the Image Archive launched a campaign on sMapshot, "the participative time machine". sMapshot is a platform for interested participants to position and geolocalise historical images on a virtual globe. The virtual globe is based on the latest satellite images and Swisstopo's 3D buildings. sMapshot is a project conducted by the Laboratoire de SIG, Haute École d'Ingénierie et de Gestion du Canton de Vaud (HEIG-VD). It allows the camera position, line of vision and height from which the photograph was taken, all the place names visible in the picture (places, rivers, fields, mountains etc.) and so-called footprints to be calculated (Graf, 2019). This tool has "become highly addictive", as the head of the Image Archive predicted at the launch (Graf, 2018). Promoting competition by publishing statistics (Figure 1) and adopting a gamification approach proved to be effective: 189 participants have georeferenced nearly 100 percent of the 68,430 aerial photographs published since January 2018 (HEIG-VD, 2019).

Fig. 1: Boosting friendly competition by publishing statistics (Screenshot from sMapshot, retrieved June 15, 2019, from <https://smapshot.heig-vd.ch/ethz>).



Why does it work? Apart from open data and gamification, community management is the key to success. Social media and the above-mentioned blog allow ETH Library's staff to interact intensively with the crowd. This is, however, not enough. The volunteers are invited to ETH Zurich at least once a year and receive awards in public for their contributions to our catalogues (ETH Library, 2019c). Six of them were thanked by publishing video interviews (Figure 2) in which they explain (in German) why they participate in our campaigns (ETH Library, 2019d). The videos are available on YouTube and accessible for the long term on ETH Zurich's video portal.

### 3.2. Applying Methodological Know-How Within Other Units

Given the positive outcome of crowdsourcing at the Image Archive the opportunity to comment on images was implemented in the Collection of Astronomical Instruments, the Art Inventory and most other E-Pics catalogues. E-Pics is ETH Zurich's platform for images, photographs and illustrations (Foulger & Wiederkehr, 2018).

ETH Library's Map Collection also adopted a crowdsourcing strategy in order to georeference old maps for further academic use and integration in

*Fig. 2: Heinz Hieke explains why he enthusiastically improves metadata of photographs (Screenshot retrieved June 15, 2019, from <https://blogs.ethz.ch/crowdsourcing/2017/06/07/video-serie-episode-4-heinz-hieke-ist-nicht-zu-bremsen/>).*



geoinformation systems (Walt, 2019). 1,135 historical maps were processed within a few months in 2017, when the Map Collection made them available on the platforms [www.oldmapsonline.com](http://www.oldmapsonline.com) and [www.georeferencer.com](http://www.georeferencer.com) and addressed the crowd via social media channels (Walt, 2017). The Map Collection's next campaign is scheduled in the autumn 2019.

e-manuscripta.ch, the co-operative digital platform for manuscript material from Swiss libraries and archives technically hosted by ETH Library, has offered a transcription tool for the crowd since 2018 (Renggli, 2018). Until recently, only image scans of archival documents were published. The crowd has not been as active in transcribing documents and thereby creating full text as in improving metadata of photographs. The reasons may be the following:

- The group of people potentially transcribing documents is not the same as the crowd interested in photographs. It is crucial to address a new community and encourage new people to co-create content.
- Transcribing documents takes more time and effort than identifying the theme of a photograph. Visual material is more attractive and evokes more emotions.

- Several institutions sponsor the content of e-manuscripta.ch. It is trying more to organise effective inter-institutional campaigns than campaigns with only one stakeholder.

As a result, ETH Library is planning a transcribathon in order to meet the potential crowd at ETH Zurich and foster friendly competition. The ETH University Archives also invite professors of history to teach courses based on materials published in e-manuscripta.ch. The idea is that the students have to transcribe sources before they analyse them. These efforts are still to come in order to transfer the positive experience with crowdsourcing of visual material to textual documents.

#### 4. Conclusion

In 2018, the University of Zurich and ETH Zurich opened the Participatory Science Academy (PWA) with the support of third-party funds. Its goal is to foster open and participatory collaboration between science and society and between scientists and citizens (Participatory Science Academy, 2019). By the time of PWA's foundation, ETH Library had become a recognised competent partner in citizen science projects. Therefore, PWA invited the head of the Image Archive to its Advisory board.

ETH Library has achieved its goals related to crowdsourcing and collaborations with citizen scientists. The volunteers have improved metadata, georeferenced maps and other materials as well as identified photographs' themes. In other words, they have completed microtasks using their cognitive ability, thus enhancing ETH Library's search tools. Taking the Image Archive as a starting point, it shared its materials in the form of open data and engaged in community building on site and in the virtual sphere. The continuous efforts and permanent interaction with the crowd prove crucial even at this level. It follows that, to date, ETH Library has limited itself to the first two levels of citizen science according to Haklay's typology. In order to reach the higher levels, the volunteers would have to be involved in problem definition and data collection ("participatory science") or even data analysis ("extreme citizen science"). This would imply that the library itself performs scientific research in the strict sense of the term. This, however, is not the case at ETH Library at the present time. Nevertheless, first

steps have been taken towards the creation of open innovation and the definition of new library services in collaboration with both scientific and non-scientific users. A further line of future action is the attempt to transfer methodological knowledge on crowdsourcing and citizen science to ETH Zurich's research groups, providing them with tools facilitating the implementation of citizen science projects.

## References

- Ayris, P., & Ignat, T. (2018). Defining the role of libraries in the Open Science landscape: A reflection on current European practice. *Open Information Science*, 2(1), 1–22. <https://doi.org/10.1515/opis-2018-0001>.
- Ayris, P., Bernal, I., Cavalli, V., Dorch, B., Frey, J., ... Horstmann, W. (2018a). *LIBER Open Science Roadmap*. <https://doi.org/10.5281/zenodo.1303002>.
- Ayris, P., López de San Román, A., Maes, K., & Labastida, I. (2018b). *Open science and its role in universities: A roadmap for cultural change*. Leuven: LERU Office. Retrieved December 13, 2019, from <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>.
- ETH Library. (2019a). *Open Data at ETH Library*. Retrieved December 8, 2019, from <https://www.library.ethz.ch/en/ms/Open-Data-at-ETH-Library>.
- ETH Library. (2019b). *Crowdsourcing. News and experiences from the community*. Retrieved December 8, 2019, from <https://blogs.ethz.ch/crowdsourcing/en/>.
- ETH Library. (2019c, May 14). 3. *Crowdsourcing-Treffen an der ETH-Bibliothek. Vorträge von Nicole Graf und Fanny Gutsche-Jones* [Video file]. Retrieved from <https://www.video.ethz.ch/events/2019/crowdsourcing/9e67a241-0926-4402-b143-093f883a5aa0.html>.
- ETH Library. (2019d). *Video-Serie. Episode 1-8* [Blog post]. Retrieved December 8, 2019, from <https://blogs.ethz.ch/crowdsourcing/tag/video/>.
- ETH Zurich. (2019). *e-pics Image Archive Online*. Retrieved December 8, 2019, from <http://ba.e-pics.ethz.ch/index.jsp?locale=en>.
- European Commission. (2018). *Open Science policy platform recommendations*. Retrieved December 13, 2019, from [https://ec.europa.eu/research/openscience/pdf/integrated\\_advice\\_opsp\\_recommendations.pdf](https://ec.europa.eu/research/openscience/pdf/integrated_advice_opsp_recommendations.pdf).
- Foulger, S., & Wiederkehr, S. (2018). E-Pics: die Plattform der ETH Zürich für Bildverwaltung und -präsentation. In A. Keller & S. Uhl (Eds.), *Bibliotheken der Schweiz: Innovation durch Kooperation: Festschrift für Susanna Bliggenstorfer anlässlich ihres Rücktrittes als Direktorin der Zentralbibliothek Zürich* (pp. 414–423). Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110553796-029>.

- Gasser, M. (2017). Research partnerships, user participation, extended outreach. Some of ETH Library's steps beyond digitisation. In A. Degkwitz & L. Romary (Eds.), *DH. Opportunities and risks. Connecting libraries and research*. IFLA Satellite meeting—Digital Humanities, Aug 2017, Berlin. Retrieved December 13, 2019, from <https://hal.inria.fr/hal-01660814>.
- Graf, N. (2015). Bilder werden frei verfügbar: Open Data im Bildarchiv. *Memoriav Bulletin*, 22, 28–30. Retrieved from [http://memoriav.ch/wp-content/uploads/2015/10/Bulletin-22\\_Web.pdf](http://memoriav.ch/wp-content/uploads/2015/10/Bulletin-22_Web.pdf).
- Graf, N. (2016). Crowdsourcing: Die Erschliessung des Fotoarchivs der Swissair im Bildarchiv der ETH-Bibliothek, Zürich. *Rundbrief Fotografie*, 23(1), 24–32.
- Graf, N. (2017). Crowdsourcing: The real people behind the crowd. In J. P. Bowen, G. Diprose, & N. Lambert (Eds.), *Proceedings of electronic visualisation and the arts (EVA) London 2017* (pp. 59–64). London: BCS Learning and Development. <https://doi.org/10.14236/ewic/EVA2017.10>.
- Graf, N. (2018, February 2). *sMapshot. We have lift-off!* [Blog Post]. Retrieved from <https://blogs.ethz.ch/crowdsourcing/en/2018/02/07/smapshot-ist-lanciert/>.
- Graf, N. (2019). *sMapshot. A historical GPS in the crowd*. In C. Busch, C. Kassung & J. Sieck (Eds.), *Kultur und Informatik. Virtual History and Augmented Present* (pp. 215–220). Glückstadt: Verlag Werner Hülsbusch.
- Haklay, M. (2013). Citizen Science and volunteered geographic information: Overview and typology of participation. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge. Volunteered geographic information (VGI) in theory and practice* (pp. 105–122). Dordrecht: Springer Science+Business Media. [https://doi.org/10.1007/978-94-007-4587-2\\_7](https://doi.org/10.1007/978-94-007-4587-2_7).
- HEIG-VD. (2019). *sMapshot. The participative time machine*. Retrieved December 8, 2019, from <https://smapshot.heig-vd.ch/ethz>.
- Holley, R. (2010). Crowdsourcing: How and why should libraries do it? *D-Lib Magazine*, 16(3–4), n.p. <https://doi.org/10.1045/march2010-holley>.
- Kälin, A. (2016, January 18). Wer kennt die Berge, Orte und Fabriken? *Neue Zürcher Zeitung*. Retrieved December 8, 2019, from <https://www.nzz.ch>.
- Kyburz, D. (2017). *Open Data: Transparency for everyone*. ETH Zürich. <https://doi.org/10.22010/ethz-exp-0004-en>.
- Open Knowledge Foundation. (2019). *What is open?* Retrieved December 8, 2019, from <https://okfn.org/opendata/>.
- Opendata.swiss. (2019). *ETH Library*. Retrieved December 8, 2019, from <https://opendata.swiss/en/organization/eth-bibliothek>.
- OpenGLAM.ch. (2019). *Hackathons*. Retrieved December 8, 2019, from <https://glam.opendata.ch/hackathons/>.

Overgaard, A.K., & Kaarsted, T. (2018). A new trend in media and library collaboration within citizen science? The case of 'A healthier Funen'. *LIBER Quarterly*, 28(1), 1–17. <https://doi.org/10.18352/lq.10248>.

Participatory Science Academy. (2019). *Aout us*. Retrieved December 8, 2019, from <https://www.pwa.uzh.ch/en/aboutus.html>.

Renggli, A. (2018). e-manuscripta.ch: Volltext – Crowdsourcing mit Qualitätssicherung. *o-bib. Das offene Bibliotheksjournal*, 5(4), 58–70. <https://doi.org/10.5282/o-bib/2018H4S58-70>.

Simperl, E. (2015). How to use crowdsourcing effectively: Guidelines and examples. *LIBER Quarterly*, 25(1), 18–39. <https://doi.org/10.18352/lq.9948>.

Strasser, B.J., & Haklay, M. (2018). *Citizen Science: Expertise, Demokratie und öffentliche Partizipation. Empfehlungen des Schweizerischen Wissenschaftsrates SWR*. Bern: Schweizerischer Wissenschaftsrat SWR. Retrieved December 13, 2019, from [https://www.wissenschaftsrat.ch/images/stories/pdf/de/Policy\\_Analysis\\_SSC\\_1\\_2018\\_Citizen\\_Science\\_WEB.pdf](https://www.wissenschaftsrat.ch/images/stories/pdf/de/Policy_Analysis_SSC_1_2018_Citizen_Science_WEB.pdf).

Töwe, M. (2018). Forschungsdatenmanagement an der ETH-Bibliothek. In A. Keller & S. Uhl (Eds.), *Bibliotheken der Schweiz: Innovation durch Kooperation: Festschrift für Susanna Bliggenstorfer anlässlich ihres Rücktrittes als Direktorin der Zentralbibliothek Zürich* (pp. 250–260). Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110553796-015>.

Walt, R. (2017, January 23). *Crowdsourcing bei Kartenmaterialien. Georeferenzieren Sie mit!* [Blog post]. Retrieved December 13, 2019, from <https://blogs.ethz.ch/crowdsourcing/2017/01/23/crowdsourcing-bei-kartenmaterialien-georeferenzieren-sie-mit/>.

Walt, R. (2019). *Use the Crowd. Georeferencing of Old Maps at ETH-library* [PowerPoint slides]. Retrieved December 13, 2019, from [http://geoportost.ios-regensburg.de/document/mapsinlibraries\\_walt.pdf](http://geoportost.ios-regensburg.de/document/mapsinlibraries_walt.pdf).