



# Swiss Open Cultural Data Hackathon 2018

## Final Report

Beat Estermann, 7 January 2019

### Table of contents

1. Introduction.....	2
2. Main Objectives of the Hackathon.....	3
3. Achievement of Project Goals.....	3
3.1. Opening up cultural data and content for reuse and making them available at a central location.....	4
3.2. Improving the visibility of Swiss heritage data and content at an international level.....	7
3.3. Promoting the re-use of cultural data / content (with a special focus on sustainability).....	9
3.4. Fostering the exchange and cooperation among stakeholders from various backgrounds.....	9
3.5. Propagating the OpenGLAM principles within the Swiss heritage sector.....	11
3.6. Promoting the public visibility of OpenGLAM.....	12
4. Evaluation of the New Hackathon Format.....	12
5. Increased Involvement of Heritage Institutions.....	12
6. Potential for Improvement.....	13
7. Outlook.....	13
8. Project Resources.....	14
8.1. Financial resources.....	14
8.2. In-kind contributions.....	14
9. Final Remarks.....	16

### Appendix A: Project Goals

## 1. Introduction

On 26-28 October 2018, the OpenGLAM Working Group of the Opendata.ch association and its partners organized the fourth edition of the Swiss Open Cultural Data Hackathon (aka GLAMhack). The hackathon was kindly hosted by the Swiss National Museum in Zurich and was thus the first of its kind to take place in a museum. Further partners included infoclio.ch, the ETH Library, Zentralbibliothek Zürich, the Swiss Social Archives, the State Archives of the canton of Zurich, the Basel Historical Museum, Wikimedia CH, docuteam, and Bern University of Applied Sciences.

The hackathon was preceded by a pre-event held at the ETH Zurich on 11 September 2018 that was specifically aimed at students and gave information about the type of data provided at the GLAMhack as well as examples of projects conducted during past editions.

The present report provides a summary of event results as well as some insights with regard to future hackathons. It is based on an assessment of project goals, the results of an internal evaluation meeting, and the past years' participants' survey. An overview of the financial result is also provided. This year's participants' survey will again be carried out roughly half a year after the event, similarly to the survey of previous years.



Opening session. Photo: Thomas Bochet, Landesmuseum Zürich, CC BY-SA 4.0.

## 2. Main Objectives of the Hackathon

The main objective of the hackathon was to bring different stakeholder groups together, for them to interact around specific topics in order to share experiences, and to develop concepts and software prototypes. The event was to be used as a jumping board for the creation of software and other forms of data/content re-use that get some public visibility or have some other outside impact (e.g. by their use for research, in the context of Wikipedia/Wikimedia, or by facilitating the crowdsourcing of certain tasks). Last but not least, the hackathon was to be used as an opportunity to encourage Swiss heritage institutions to open up their data and content and to spread the word about the principles of OpenGLAM<sup>1</sup>.

The 2018 edition of the hackathon was characterized by a particular focus on museums as data providers, which had clearly been underrepresented in previous hackathons, and by a focus on hackathon projects that engage an audience – either online or in form of exhibits/installations.

However, like in the previous editions, all types of heritage institutions were encouraged to provide data and to participate in the hackathon, as it is the goal of the organizers to improve the networking among the institutions, especially also across sectors (museums, archives, libraries).

The change of orientation came with the following adjustments to the hackathon concept compared to previous years:

- The implementation of hackathon projects involving non-standard hardware were encouraged;
- The format of the hackathon was modified:
  - The 2-day hackathon on Friday and Saturday was followed by project demonstrations/presentations on Sunday afternoon, targeting both a broad public and GLAM professionals from institutions that had not been involved in OpenGLAM so far.
  - The best projects were awarded with a symbolic prize – the award was mainly introduced for presentation and communication purposes.

## 3. Achievement of Project Goals

The table in [appendix A](#) gives an overview of the goals that were set for this year's hackathon, the level of their achievement, the achievements in the previous years for comparison, as well as suggested targets for next year. Please note that not all targets could be assessed yet, as dissemination activities take more time and this year's participants' survey has not been carried out yet.

---

<sup>1</sup> <http://openglam.org/principles/>

The documentation of individual hackathon projects and an overview of the media coverage can be found on the event website<sup>2</sup>.

Specific goals were set in six areas. In the following, we will shortly discuss the level of achievement of these goals in the context of the hackathon.

### 3.1 Opening up cultural data and content for reuse and making them available at a central location

The specified targets were partly achieved: 148 open datasets / collections from 66 Swiss institutions<sup>3</sup> have so far been made available through the [make.opendata.ch](http://make.opendata.ch) website (compared to 116 open datasets / collections from 60 institutions in the previous year). Participation in the hackathon by museums fell again short of expectations. Contrary to the original plan, no pre-event was held that was specifically targeted at museums. In general, fewer outreach efforts to acquire new data providers were made than in previous years (only about 30 data owners were contacted individually, compared to approx. 135 in the previous year).



Photo: Thomas Bochet, Landesmuseum Zürich, CC BY-SA 4.0.

<sup>2</sup> <http://make.opendata.ch/wiki/event:2018-10>

<sup>3</sup> These numbers do not include all the collections from the e-rara and e-manuscripta platforms, which were officially marked as Public Domain material in the course of 2017; these platforms have been counted just as one dataset.

At the same time, the efforts that were made to win providers of cultural data platforms over to the cause of OpenGLAM did not have any direct effects in 2018, although related developments clearly go in the right direction:

- Several Swiss heritage data portals respect the OpenGLAM principles and/or actively promote the GLAM cause. Notable examples are: e-rara, e-manuscripta, FotoCH, kirchen.ch/archive, as well as Swissbib.
- The Memoriav association has indicated that they want to adopt an OpenGLAM-compatible strategy for the further development of its Memobase platform. First implementation steps are expected in 2019. Similar discussions have taken place with the KIM.bl platform.
- Several Swiss archives are moving in the direction of linked open data publication; the aLOD pilot project is about to be extended.
- In the case of platforms using federated search (Archives Online; Museums Online), it is unclear whether systematic licensing information is to be expected anytime soon.
- There is also no licensing information included at the moment on the Vallesiana platform. Discussions with providers will be further pursued.
- As of today, e-codices is the only recalcitrant cultural data platform in Switzerland that continues to claim copyright and to apply non-free licenses on content that is clearly in the public domain.

With regard to getting more museums on board, closer cooperation with platforms that cater to many museums, such as Memobase or KIM.bl, seems the most promising. Furthermore, (other) leading institutions in their respective fields could be approached in order to convince them to take an active part in the hackathon. Note that museums make up for about two thirds of the approx. 1600 heritage institutions in Switzerland that are of interest in the context of OpenGLAM.

#### Data catalogue platform strategy

The cooperation with the opendata.swiss platform has been continued. By default, data providers are expected to manage the description of their data from their own user account. Alternatively, institutions are offered the possibility to (temporarily) include their descriptions on the OpenGLAM CH account. The collaboration with the Swiss Federal Archives, who have been hosting the portal, has taken place in an agreeable atmosphere. The Swiss Federal Archives have however been slow to implement the improvements suggested. Concretely, the hackathon team would still like to see improvements to the platform in four areas:

- The ability to add references to standard licenses (e.g. Creative Commons licenses) in the metadata.
- The ability to add a picture to the metadata record (which is useful especially in the case of collections of content, like photographs, prints, scans of paintings, etc.).



- Adaptations to the functionality allowing the integration of data from opendata.swiss on third party websites (such as widgets showing up-to-date metadata from the portal).
- The ability to tag datasets at the level of the catalogue. At present, tags for datasets are managed exclusively by the organizations, which makes it awkward to tag a specific subset of datasets for a specific event (e.g. a hackathon). Furthermore, the automatic harvesting of some organizations is set up in a way that prevents the attribution of tags in language versions not covered by the harvesting routine.

In addition, the hackathon team was a bit puzzled this year at the outright rejection by the Swiss Federal Archives to provide a user account on opendata.swiss to the historical archives of a Swiss public-law institution on the grounds that this institution was purportedly not a “public sector” organization. The Opendata.ch association has requested that all Switzerland-based heritage institutions be offered the possibility to reference their data on opendata.swiss, inasmuch as the large majority of these institutions are either under public-law or predominantly publicly funded. We deem that it is in the public interest that the data from these institutions be provided as open data and referenced on opendata.swiss. This point of view is reflected in the Confederation’s new Open Government Data Strategy for the period 2019-2023<sup>4</sup>, which explicitly calls upon private sector entities detaining data of public interest to release them as open data and to reference them on opendata.swiss. We therefore call on the Federal Statistical Office, who will take over the management of the opendata.swiss portal from January 2019, to take corrective action and to provide a user account on opendata.swiss to every Switzerland-based heritage institution who requests such an account in order to reference their open data and/or content.



Photo: Thomas Bochet, Landesmuseum Zürich, CC BY-SA 4.0.

<sup>4</sup> <https://www.news.admin.ch/news/message/attachments/54908.pdf>

## Harmonizing data publication practices

Despite of respective plans, little activity has taken place to start a conversation with data providers on how to handle the heterogeneity of datasets and to develop shared best practices when it comes to publishing datasets.

As has already been pointed out in last year's report, datasets in the heritage sector are rather diverse; they typically comprise one of the following types of data or a combination thereof:

- structured data describing or documenting some cultural or historical phenomena;
- collections of digital artefacts (photographs, scans of prints, paintings, etc.);
- metadata pertaining to collections of digital or physical artefacts;
- geodata services (e.g. historical maps available through a web map service);
- ontologies, thesauri, or vocabularies;
- registers of ontologies, thesauri, or vocabularies;
- Inventories or registers of heritage institutions or collections.

Another source of heterogeneity among datasets lies in their varying granularity:

- While some datasets comprise entire catalogues, covering the collections of many institutions (e.g. catalogues of platforms such as Swissbib or e-codices), others pertain to the collection of a single institution (sometimes also available through one of the platforms).
- While some digital collections comprise many thousands of artefacts, others only comprise a few dozens of them. Whether or not to break one collection up into many smaller ones is usually at the discretion of the data owner.

A further source of heterogeneity of datasets in the context of OpenGLAM lies in the fact that some collections may comprise both artefacts that are in the public domain or freely licensed and artefacts whose copyright status is unknown or the use of which is restricted. Here again, it is up to the data owner to decide whether or not to split up existing collections into smaller ones based on the licensing criterion. From a thematic point of view, splitting collections up based on this criterion may often not make much sense.

### **3.2 Improving the visibility of Swiss heritage data and content at an international level**

The catalogue entries on the opendata.swiss platform are automatically harvested and made available on the European Data Portal<sup>5</sup>. Since late fall 2018, they are also searchable via Google Dataset Search<sup>6</sup>.

---

<sup>5</sup> <https://www.europeandataportal.eu>

<sup>6</sup> <https://toolbox.google.com/datasetsearch>

## The role of Wikimedia Commons and Wikidata

Wikipedia is a very effective channel to increase the visibility of heritage data and content at an international level. To allow its use within Wikipedia, content needs to be uploaded to Wikimedia Commons, and data needs to be made available through Wikidata. Apart from the use of the data within Wikipedia and its sister projects, Wikidata is more and more being used by heritage institutions worldwide as a hub for authority files and as a collaborative lab for data integration.

With a certain regularity, the annual hackathon brings to the fore new collections and datasets that are potentially valuable to the Wikimedia community. Many of them would be worthy of uploading to Wikimedia Commons or of ingesting in Wikidata. Since 2017, there has been a backlog of open collections and datasets to be ingested. In some cases, the institutions are taking care of the upload themselves, in others they need to be provided with some guidance on how to go about it. There have been repeated contacts with Wikimedia CH and the Wikipedia community to instigate work on this backlog, but to our knowledge, not much activity has taken place to this effect. Maybe, the launch of structured data on Wikimedia Commons in 2019<sup>7</sup> would provide a good opportunity to re-initiate efforts to upload more content to Commons.

At the same time, it has to be noted that the action required is not just about uploading content to Wikimedia Commons or about ingesting data into Wikidata, but about building communities around this content and data. As has already been noted in last year's report, there has been a radical change in this area over the past five years: While the main challenge a few years ago had been to convince institutions to open up their data and content (with online communities only waiting for their release), the challenge nowadays lies in growing the online communities and in attracting their attention to newly released datasets. The reasons for this mainly lie in the large quantities of heritage data and content that have been made available at a global level, in the modest community growth, as well as in platform interfaces and functionalities that – almost by definition – are lagging behind such developments.



Photo: Thomas Bochet, Landesmuseum Zürich, CC BY-SA 4.0.

---

<sup>7</sup> [https://commons.wikimedia.org/wiki/Commons:Structured\\_data](https://commons.wikimedia.org/wiki/Commons:Structured_data)



## The role of data portals in exposing heritage data and content at an international level

Apart from the measures described above, other efforts may be undertaken directly by the heritage institutions themselves in order to enhance the visibility of their data and content at an international level. The impact of the sole hackathon may be limited in this area, but the organizers should sensitize the institutions for this issue. In this context, conversations have been held with providers of Swiss heritage data portals, as to which are the best avenues to promote the use of Swiss heritage data and content at an international level, and how they could support this effort. These efforts need to be continued.

### **3.3 Promoting the re-use of cultural data / content (with a special focus on sustainability)**

This is a goal that is generally well achieved by means of the hackathon, as the number of hackathon projects as well as the number of responses to the 2015-2017 participants' surveys indicate: For about 60% of the participants, the hackathon usually proves (very) effective in helping them find out how data/content can be used, while more than 70% usually find that it was (very) effective in getting new inspiration or ideas.

While in previous years, a decrease in the number of hackathon projects that were pursued during the event had been observed, this trend has been reversed this year. Despite the smaller number of active participants, a larger number of projects has been pursued during the 2018 edition than in the preceding year (15 projects, compared to 11 projects in 2017), reaching the same level as in 2016. Thanks to the focus on hackathon projects that engage an audience and the encouragement to use non-standard hardware, the variety of projects was increased compared to the two previous years. At the same time, some hackathon participants of previous years were put off by the specific focus of the hackathon and the fact that the best projects were awarded with a prize (competitive element).

### **3.4 Fostering the exchange and cooperation among stakeholders from various backgrounds**

This is an area where the hackathon is doing very well, as is exemplified by the fact that approx. 80% of participants of the 2015-2017 editions appreciated the hackathon's effectiveness in terms of meeting interesting people and fostering networking. There is also strong episodic evidence that this exchange and cooperation is actually happening within hackathon project teams. Table 1 below shows that we have again been able to attract a good mix of participants from various backgrounds and profiles.

**Table 1: Different categories of participants**

Participant category	2015 edition (N = 49 of 107)	2016 edition (N = 94 of 105)	2017 edition (N = 94 of 98)	2018 edition (N = 66 of 69)
Female	19%	<b>33% ↑ **</b>	37%	39%
Male	81%	<b>67% ↓ **</b>	63%	61%
Data provider	35%	–	–	–
Data provider or expert	–	28%	32%	38%
Software programmer	35%	25%	34%	33%
Ideator	27%	25%	22%	21%
Researcher	22%	31%	21%	27%
Wikipedia editor	12%	11%	5%	3%
Wikidata editor	–	–	5%	3%
Artist	8%	7%	<b>1% ↓ **</b>	5%
Designer	4%	<b>13% ↑ *</b>	<b>3% ↓ **</b>	8%
Organizer	25%	–	–	–
Hackathon organizer	–	11%	11%	17%
Other	12%	20%	18%	11%

Changes marked \* are significant at the 0.10 level; those marked \*\* are significant at the 0.05 level.

The smaller number of participants is in part due to the fact that no side programme was offered this year. Thus, participants of previous years who mainly attended the hackathon to take part in workshops or to attend presentations, did not enter the active participant count this year. At the same time, this year's hackathon saw a significantly higher number of "visitors" than in previous years – both during the hackathon and especially for the project presentations on Sunday. Thus, the project presentation event was attended by approx. 80 visitors in addition to the regular attendants, with the visitors mingling and exchanging with the hackathon participants.

Interestingly, despite the low number of hackathon participants who self-identified as Wikidata editors, six out of fifteen hackathon projects made use of Wikidata in one way or the other.

### 3.5 Propagating the OpenGLAM principles within the Swiss heritage sector

The targets in this area have not been achieved, due to the limited efforts to reach out to museums and other potential data providers. Thus, notably fewer institutions than in previous years were directly contacted in view of the hackathon, and no pre-event targeted at heritage institutions was held. Contrary to the initial plan, no pre-event specifically targeted at museums was organized.

#### The role of data portals

In the longer term, heritage data portal providers are important allies when it comes to promoting the OpenGLAM Principles<sup>8</sup>, notably by implementing the following policies and/or functionalities on their platforms:

- Do not apply any (legal or technical) usage restrictions on public domain content.
- Declare the copyright status (and possibly other usage restrictions) of the content provided or referenced on the portal (this should be part of the platform's metadata schema).
- Publish the metadata provided through the portal under a CC-0 Waiver<sup>9</sup> and make it available for download.

In addition, portal providers may actively encourage and empower their data providers to freely license content and to increase its usability and visibility at an international level. As has been pointed out above, while efforts to win providers of heritage data platforms over to the OpenGLAM cause have met some success, further efforts are needed in this area.



Members of the Jury. Photo: Thomas Bochet, Landesmuseum Zürich, CC BY-SA 4.0.

---

<sup>8</sup> <https://openglam.org/principles/>

<sup>9</sup> <https://creativecommons.org/publicdomain/zero/1.0/>

### **3.6 Promoting the public visibility of OpenGLAM**

Despite the public project presentation session and award ceremony on Sunday, that reached the expected attendance levels (approx. 80 visitors), the hackathon received again no coverage in the classical media.

While the hackathon is a good means to promote the idea of OpenGLAM among a specialized public (heritage institutions, digital humanities, interested software programmers), and provides a good opportunity for participating institutions to promote the idea of OpenGLAM among their audiences on blogs and social media, the event does not readily lend itself to reaching a wider audience.

## **4. Evaluation of the New Hackathon Format**

The new hackathon format worked out fairly well. The projects pursued this year were varied and of high quality. The use of virtual reality equipment was an enrichment. Although not many museums participated in the event as data providers, several hackathon projects used data from museums. The trend towards larger teams working on fewer projects, which had manifested itself over the last two years, could be reverted. Thanks to the public presentation event and award ceremony on Sunday in a central location, we succeeded for the first time to attract a significant number of visitors who were able to interact with the hackathon participants. Thanks to the extra day and the increased focus on the public event and award ceremony, the quality of presentations clearly improved compared to previous years.

On the negative side, the 3-day format demanded a longer commitment from the participants, which may indeed have been too long for some. It took some extra time for the teams to interact with the jury members and to prepare their presentations, but this also made the documentation and the presentations much richer. Also, the awards may alter the “spirit” of the hackathon, and for some participants it created unnecessary stress and pressure. Some participants of the previous years stated that they did not participate this year due to the competitive element. At the same time, the event attracted new participants.

In order not to overload the event, the side programme with workshops and presentations was dropped this year, although it had been very well received over the previous years, especially also among representatives of heritage institutions.

## **5. Increased Involvement of Heritage Institutions**

As had already been noted in last year’s report, heritage institutions are more and more taking ownership of the event in a positive way: Several heritage institutions have contributed substantial amounts of staff time to the organization of the hackathon, and for the first time, the event was funded without the support from foundations. Also, project coordination was ensured without an implicit deficit guarantee by BFH.



## 6. Potential for Improvement

Potential for improvement has been identified in several areas. The most notable points that were brought up during the internal evaluation meeting were the two following:

- As had already been noted in last year's report, the event communication should be professionalized. Also, the event website should be renewed. This requires more resources; this year's tight budget only allowed for a minimal level of activities in this area.
- Outreach efforts vis-à-vis museums should be further intensified.

## 7. Outlook

The 2019 edition of the hackathon will be organized in cooperation with Museomix. The new hackathon format will largely be maintained. In addition, a make-a-thon component will be added, and outreach efforts towards museums will be increased.

Given the high quality and variety of project presentations, a follow-up event could be organized where hackathon artefacts are presented to a wider audience. This may also give the participants an extra incentive to further pursue their projects. Ideally, such a follow-up event would be integrated into a larger event that already per se attracts a larger audience, such as a museums night or similar.

In the following years, the hackathon could then again take a format similar to the one of the first years of its existence: Back to the 2-day or 2.5-day format, with a side programme of workshops and presentations, but without awards and public presentation event. The focus of such a future event could be on linked data, crowdsourcing and/or machine learning.



Award Ceremony. Photo: Andreas Hösli, Landesmuseum Zürich, CC BY-SA 4.0.

## 8. Project Resources

### 8.1 Financial resources

Table 2 below gives an overview of the financial resources of the project:

	Budget	Effective Costs / Revenues
<b>Expenses</b>	- 45'400	- 30'759
Catering	14'400	14'488
Accommodation (Youth Hostel)	5'000	2'157
Project Coordination	15'500	10'705
Association Opendata.ch (10% of revenues)	4'500	3'076
Varia / Incidentals	6'000	333
<b>Revenues</b>	+ 45'400	+ 33'100
Internal Sponsors	22'000	22'000
External Sponsors	19'150	9'000
Voluntary Participation Fees	4'250	2'100
<b>Balance</b>	0	+ 2'341

**Table 2: Overview of financial resources**

The project was kindly supported by the following sponsors:

- ETH Library (10'000 CHF)
- Wikimedia CH (9'000 CHF)
- Zentralbibliothek Zürich (6'000 CHF)
- Infoclio.ch (4'000 CHF)
- Sozialarchiv (2'000 CHF)

Note that the effective costs for accommodation have been lower than expected given the fact that the effective number of participants was lower (69) than the one budgeted for (100). Catering costs per person were significantly higher than the previous year due to higher catering prices in Zurich compared to the University of Lausanne.

The resulting benefit remains on the account of the Opendata.ch association and is earmarked for future hackathons or similar events related to cultural heritage.

Thanks to the financial contributions by the members of the Friends of OpenGLAM Network and to the reserves from previous years, we are now able to ensure the project coordination without the implicit deficit guarantee by the Bern University of Applied Sciences. In 2018, a coordinator for the Swiss Open Cultural Data Hackathon was for the first time directly hired by the Opendata.ch association.

## 8.2 In-kind contributions

The organization of the event would not have been possible without the substantial in-kind contributions made by several individuals and organizations, most notably by contributing their volunteer and/or staff time:

- Dario Donati, Thomas Bochet, Dominik Sievi, Anna Durisch (Swiss National Museum)
- Michael Gasser (ETH Library)
- Lothar Schmitt (Zentralbibliothek Zürich)
- Lionel Walter (Basel University Library)
- Tobias Hodel (State Archives of the Canton of Zurich)
- Stefan Bürer (Basel Historical Museum)
- Jan Baumann (infoclio.ch)
- Oleg Lavrovsky (opendata.ch)
- Beat Estermann (Opendata.ch / Bern University of Applied Sciences)

but also by allowing us to use their infrastructure and/or equipment:

- Swiss National Museum (hosting of the hackathon)
- ETH Library (hosting of pre-event)
- Opendata.ch (online and hardware infrastructure, financial administration).



Photo: Thomas Bochet, Landesmuseum Zürich, CC BY-SA 4.0.

## 9. Final Remarks

This year's hackathon has again been a great success, and both the participants and the organizers are looking forward to the next edition of the event.

As the present report shows, there is room for improvement in several areas and there are concrete ideas of how to reach out to new target groups and to preserve the innovative spirit of the hackathon in the future.

We would like to thank all our sponsors and partners for supporting the event and are looking forward to opportunities of future cooperation.

### Our Sponsors:

**Landesmuseum Zürich. SCHWEIZERISCHES NATIONALMUSEUM. MUSÉE NATIONAL SUISSE. MUSEO NAZIONALE SVIZZERO. MUSEUM NAZIONALE SVIZZERO.**



## Appendix A: Project Goals