

## OpenGLAM CH Working Group

# Meeting Regarding Open Cultural Data Platform(s), 17 March 2014 Minutes

**Date/Time:** 17 March 2014, 15:00 – 17:00

**Place:** Bern, Hallerstrasse 8

### Participants:

- Beat Estermann (BFH) [BE]
- Carl Flisch (PD Projekt) [CF]
- Adelheid Jann (Swiss Federal Archives; also representing the newly created working group “Zugang und Vermittlung” of the Swiss Association of Archivists) [AJ]
- Ulrich Lantermann (Wikimedia/Wikipedia) [UL]
- Micha L. Rieser (Wikipedian in Residence) [MLR]
- Guillaume Rey-Bellet [GRB]

### Excused:

- Charles Andrès (Wikimedia CH)
- Florian Ducommun (Creative Commons Switzerland)
- Michael Gasser (ETH Library)
- André Golliez (itopia)
- Jürg Hagmann (RIM Consulting; ARMA) [JH]
- Enrico Natale (infoclio.ch) [EN]
- Bruno Jehle (Digitale Allmend)
- Oleg Lavrovsky (opendata.ch)
- Matthias Nepfer (Swiss National Library)
- Iolanda Pensa (SUPSI)

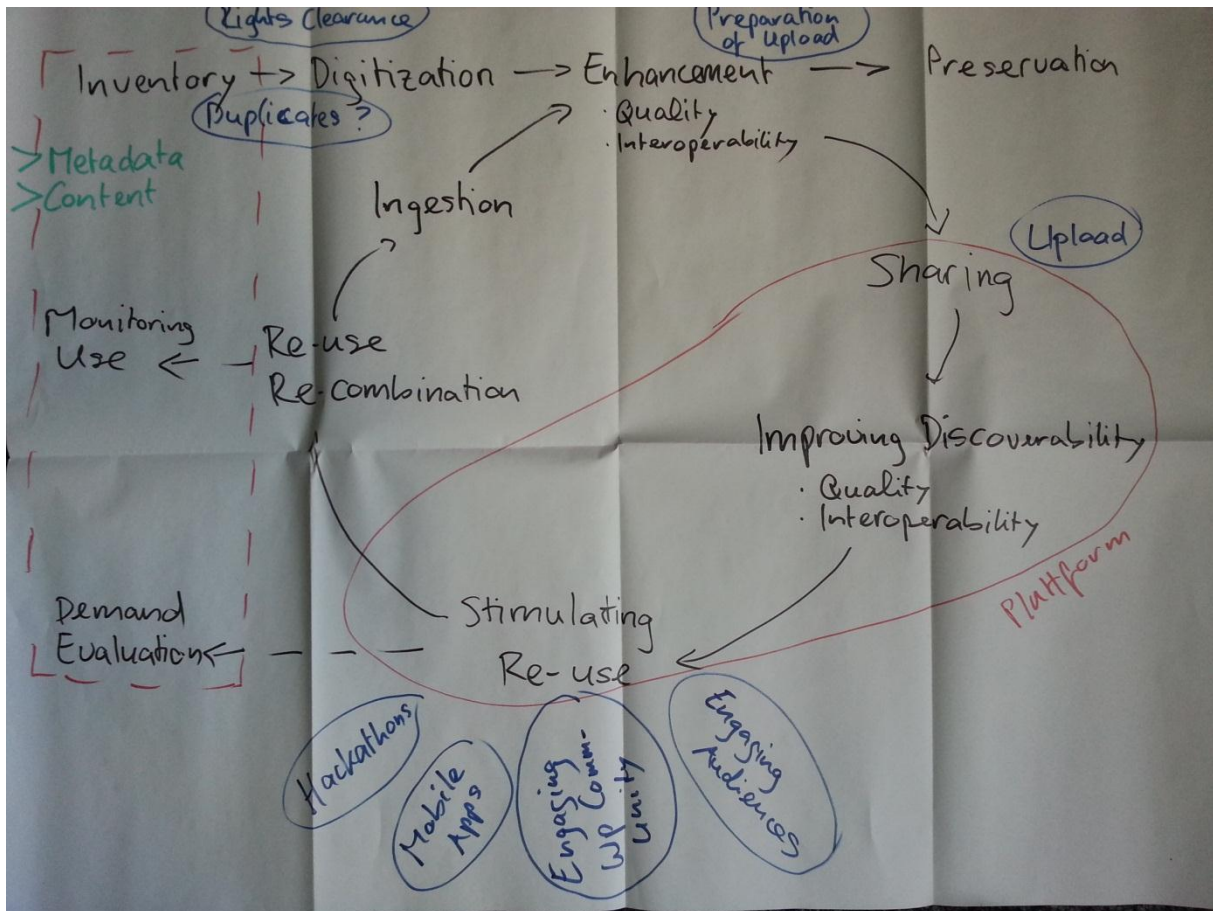
**Minutes:** Beat Estermann

### Background

The meeting was called following the plenary workshop of 31 January 2014 with the goal to constitute a task force looking into the issue of cultural data platform(s). There was a lively discussion at earlier meetings and on the mailing list whether it actually makes sense to create a new Swiss cultural data platform. The rationale behind the proposal for the creation of a Swiss cultural data platform, an overview of existing platforms, as well as inputs from the discussion are summarized in the annex. The annex is also available as a living [Google Doc](#) where further additions and corrections can be made.

## Discussion

- It is useful to discuss the platform question in the context of an OpenGLAM process model:



- There are many existing platforms; it therefore makes sense to start working with them and to see to what extent they can cover the needs of OpenGLAM.
- Running a platform can be rather resource intensive both for the platform provider and the content providers. We don't see sufficient reasons to advocate for this kind of investment right now.
- Also, we should better understand the problems we want to solve with a (new) platform. Current issues identified during the meeting comprise:
  - Insufficient use of synergies (e.g. between platforms)
  - Lack of an overview of which data and content exist, which data/content are open, etc.
  - Lack of online access to many documents (the documents wouldn't necessarily need to be open; they could be made available in a protected environment in a first step for inspection by third parties; the process of opening them up could be initiated in a second step if required).
- It appears that the Open Government Data platform of the Swiss Confederation will be open to cultural data from Swiss heritage institutions. The draft consultation version of the Open Government Data Strategy of the Swiss Confederation contains a measure regarding the creation of an inventory of all existing government data sources. It is still unclear how this inventory is going to be realized and implemented.
- The OpenGLAM working group could issue recommendations regarding the platforms to be used, based on a comparative analysis. The overview of platforms in the annex could be a good starting point for this.
- We could also document processes related to opening up collections based on experiences gained in pilot projects (UL and MLR would be ready to contribute to this), also identifying the main bottlenecks

and obstacles that prevent stakeholders from scaling up their activities. These experiences could be shared at a conference for GLAMs interested in opening up their collections. The OpenGLAM working group would thus function as an enabler.

- One of our outputs could be a white paper describing OpenGLAM related processes: what are the expectations? Quality requirements? Licensing questions, etc.
- We should partner with umbrella organizations and providers of existing platforms. Furthermore, we should get more GLAM representatives on board. In order to be a credible partner, we should further clarify our profile as an OpenGLAM working group, including:
  - Goals
  - Mission
  - Stakeholder groups (and how we want to engage with each of the stakeholder groups)
  - Requesting vs. supporting/enabling (to what extent do we see our role in requesting GLAMs to open up their collections; to what extent do we see our role in supporting/enabling them?)
  - Usefulness of our activity
  - Relationship to open data / free knowledge communities (is this the unique selling point of our Working Group compared to other groups?)

### **Decisions**

- We will focus our activities on existing platforms.
- There is presently no sufficient basis to constitute a task force related to platforms (profile of the OpenGLAM working group to be clarified; no commitment among the meeting participants regarding a hands-on approach concerning platforms).
- Feedbacks gathered during this meeting will serve as inputs for the next coordination meeting in the plenary.

# **ANNEX: Platform(s) for Open Cultural Data / Content (in Switzerland)**

## **Background / status quo**

### **We want to promote OpenGLAM in Switzerland:**

- All metadata under CC-0 or similar license
- All public domain content marked as public domain
- All other content if possible under a free copyright license (CC-by or CC-by-sa or similar)

Any platform we support should be supportive of these goals.

### **Initial motivation to propose the creation of an open cultural data / content portal Switzerland** (summarizing inputs from André Golliez)

- Good experience with Open Government Data platforms in form of metadata hubs: create an inventory of all available datasets and populate the platform by making links to them available. In some countries, contributing to the inventory is mandatory for government agencies.
- Presently, a metadata-hub for research data is in preparation. Goal: inventory of all existing datasets (even those which are not open) across disciplines, i.e. moving away from the present silo structure. During the preparation it turned out that researchers want their own platform (don't mix their data with government data) at a Switzerland level (don't just put your data on an international platform).
- Assumption 1: it would be beneficial to overcome the present silo structure also regarding cultural data/content
- Assumption 2: heritage institutions also want their own open data platform.
- Assumption 3: heritage institutions in Switzerland want/need a Swiss platform. Europeana is not sufficient as a metadata directory for cultural data in Switzerland.

## **Aspects to be taken into account when talking about an open cultural data/content platform**

- **What is the scope of the platform?**
  - Thematic/institutional/geographical scope
  - What types of data/content should be hosted on the platform?
    - catalogues, inventories, finding aids (metadata records)
    - name authority files
    - vocabularies, thesauri, ontologies
    - digital reproductions or representations of cultural heritage objects
    - digital born cultural heritage objects
    - ...
  
- **What are the functionalities of the platform?**
  - Search
  - API
  - Integration with web2.0 services
  - Mobile app
  - Contextualisation by users
  - Improvement of metadata by users
  - ...
  
- **To what extent do the main data providers and target users identify with the platform?**

# Overview of existing platforms

- *Overview of the present situation (open government data platform; various platforms in the GLAM sector; Europeana)*
- *What is presently achieved by the existing platforms? (strengths / shortcomings)*

## 01. Platforms for (open) cultural data at the international level

### a. Europeana

- i. metadata aggregator
- ii. thumbnails and links to content on other platforms (institutional platforms)
- iii. main focus on libraries and museums (?) (however, Europeana calls itself “Europe’s digital library, museum and archive”)
- iv. promotion of content through social media (blog, facebook, twitter, pinterest, google+), mobile app, and online exhibitions
- v. crowdsourcing of collection (e.g. untold stories of WWI)
- vi. metadata is licensed CC0
- vii. content licenses are up to the institutions holding the content; Europeana is lobbying the institutions in favour of labelling public domain works as public domain (a practice many heritage institutions presently don’t follow); see also Europeana’s [Public Domain Charter](#)
- viii. an API is available for registered users/uses (confidential API key required)
- ix. relationships with similar (world-)regional portals, such as the Digital Public Library of America

### b. Archives Portal Europe

- i. <http://www.archivesportaleurope.net/>
- ii. aggregator of finding aids, holding guides, and source guides
- iii. thumbnails and links to content on other platforms (institutional platforms)

### c. Wikimedia Commons

- i. <http://commons.wikimedia.org>
- ii. content portal (media files are uploaded to Wikimedia Commons, along with their metadata)
- iii. serves as a media repository for Wikipedia and Wikisource
- iv. content needs to be in the public domain or licensed under a free copyright license (such as CC-by, CC-by-sa, or CC-0)

- v. enhancement of meta data management capabilities is planned (integration with wikidata)
- vi. enhancement of mass-uploading capabilities is ongoing (GLAM-Wiki toolset)

**d. Flickr - The Commons**

- i. <http://www.flickr.com/commons>
- ii. content portal (media files are uploaded to Flickr, along with their metadata)
- iii. serves as a media repository for bloggers
- iv. heritage institutions upload media files, and users (i.e. the general public) are asked to contribute information and knowledge
- v. within "Flickr - The Commons" content needs to be published "with no known copyright restrictions"

**e. Youtube**

- i. <http://www.youtube.com>
- ii. video sharing portal (media files are uploaded to Youtube, along with their metadata)
- iii. serves as media repository for bloggers and various websites

**f. Internet Archive**

- i. <https://archive.org/>
- ii. Initial focus on born-digital materials, especially web sites; in the meanwhile accepting digitized versions of other cultural artefacts (texts, audio, moving images) as well as software.
- iii. Open Library: web database for every book ever published

**g. International Inventory of Musical Sources**

- i. <http://www.rism.info>
- ii. multinational, non-profit joint venture which aims for comprehensive documentation of extant musical sources worldwide. These primary sources are manuscripts or printed music, writings on music theory, and libretti. They are housed in libraries, archives, monasteries, schools and private collections. Founded in 1952.
- iii. metadata aggregator

**h. VIAF**

- i. <http://viaf.org/>

- ii. combines multiple name authority files into a single OCLC-hosted name authority service
- iii. ....

**i. LiAM - Linked archival metadata**

- i. <http://sites.tufts.edu/liam/>
- ii. The Linked Archival Metadata (LiAM) Project began October 1, 2012 with funding from the Institute for Museum and Library Services National Leadership Grant program (Tufts university Boston). LiAM is a planning grant project whose deliverables will facilitate the application of linked data approaches to archival description.
- iii. Extending the Encoded Archival Description (EAD) approach to the tools of linked data and semantic web (eg RDF) to define a set of relationships specific to archival collections.

**02. Platforms for (open) cultural data at the national level (Switzerland)**

**a. Swiss Open Government Data Portal**

- i. <http://opendata.admin.ch>
- ii. Portal for Open Government Data in Switzerland
- iii. metadata aggregator
- iv. references to datasets
- v. references to content (experimental, cf. WWI pictures as download in zip format); no links to Wikimedia Commons, although the pictures had been uploaded to Commons; no integration of Commons categories.
- vi. Cultural data / content is accepted on the portal; in personal exchanges, the portal provider (Swiss Federal Archives) has expressed their willingness to accept data from institutions from outside the government sphere. So far, open cultural data / content hasn't officially been part of the Open Government Data project of the Swiss Confederation (lead: ISB in cooperation with Federal Archives and Federal Chancellery), nor are heritage institutions mentioned in the corresponding draft strategy for 2014-2018 (the strategy is presently still in consultation). In contrast, the revised PSI Directive of the European Union (2013) explicitly includes heritage institutions within the scope of the directive.

**b. Swiss Electronic Library**

- i. <http://www.e-lib.ch>



- ii. search portal for the holdings of Swiss university libraries and academic institutions

**c. Swissbib**

- i. <https://www.swissbib.ch/>
- ii. metacatalogue of Swiss university libraries and the Swiss National Library
- iii. metadata aggregator

**d. archives-online.org**

- i. <http://archives-online.org>
- ii. catalogs of archival collections in Swiss state archives (12 cantons), archival collections in the Swiss National Library, municipal archives and other special collections
- iii. metadata aggregator

**e. Memobase**

- i. <http://www.memobase.ch/>
- ii. Portal providing information about audiovisual documents at 13 Swiss heritage institutions.
- iii. metadata aggregator
- iv. access to some content is provided through a network of listening/viewing stations
- v. Content is typically “all rights reserved”, even in cases where the heritage institutions are the rightholders.

**f. Kartenportal.CH**

- i. <http://www.kartenportal.ch/>
- ii. Specialist portal for maps from libraries and archives in Switzerland.
- iii. metadata aggregator

**g. e-manuscripta**

- i. <http://www.e-manuscripta.ch>
- ii. manuscript material from Swiss libraries and archives
- iii. content is hosted on the platform
- iv. content is licensed under CC-by-nc (even PD works)

**h. e-codices**

- i. <http://www.e-codices.unifr.ch>
- ii. virtual manuscript library of Switzerland
- iii. content is hosted on the platform
- iv. content (faksimiles) is licensed under CC-by-nc (even PD works)

**i. e-rara**

- i. <http://www.e-rara.ch/>
- ii. platform for digitized rare books from the 15th to 19th century from Swiss libraries
- iii. content is hosted on the platform
- iv. content (faksimiles) is licensed uncer CC-by-nc (even PD works)

**j. International Inventory of Musical Sources - Swiss branch**

- i. <http://www.rism-ch.org>
- ii. records music manuscripts and printed musical sources in Swiss libraries and archives according to international scientific standards
- iii. metadata aggregator

## What are the requirements with regard to (a) platform(s) for Open Cultural Data / Content from a Swiss perspective?

- *What do we want to achieve with a hypothetical new platform? How would it fit into the existing platform landscape?*
- *Key requirements of the new platform?*
- *Alternative scenarios? Could existing platforms be enhanced to fulfill the requirements? What would be the downsides of a new platform compared to existing platforms?*

Inputs from the discussion in the run-up of the workshop:

- To what extent would a Swiss metadata inventory be instrumental in making existing datasets visible? (the UK approach to open government data consists in making the inventory of all datasets, be they open or not, available online; potential users can ask for specific datasets to be opened)
- It is important to design a process describing how a metadata inventory is created and how the metadata will in turn be made available (according to open data principles). We should first focus on the metadata, then on the datasets, and eventually on the digital representations/reproductions of heritage items.
- The biggest challenge will be to convince institutions to open their data.
- Whether to have a separate platform for Swiss cultural data/content is a political decision. It should be supported by an alliance of Swiss cultural (heritage) institutions (cf. BAK, SRG, Memoriav, umbrella associations: BIS (libraries), VSA (archives), VMS (museums))
- Efforts shouldn't focus too much on the frontend. What is important is the metadata infrastructure.
- It makes sense to team up with partners that also aim at aggregating cultural data (e.g. Europeana).
- Create a single point of entry (the platform acting just as a façade), e.g. an intelligent search engine, allowing for category search for media across platforms
- The separation between the data world and the Wikimedia world should be overcome
- Ideally, a GLAM would need to publish their media or data on one platform only, without needing to always figure out how many platforms exist and which are the most important ones. (A counter example that doesn't make sense: the Federal Archives have published their WWI pictures on the OGD as a zip-file; the same pictures have been published earlier on Wikimedia Commons.)
- Wikimedia Commons sometimes applies more restrictive copyright rules that don't allow uploading pictures that fall in the public domain in Switzerland. This could be

an argument in favour of a separate platform.

- Is there currently a debate to add "cultural data" to [opendata.admin.ch](https://opendata.admin.ch)? Is [opendata.admin.ch](https://opendata.admin.ch) in discussion with other initiatives (e.g. e-rara) that publish material? What about private archives? Will they also publish material at [opendata.admin.ch](https://opendata.admin.ch)?
- If it substantially helps to make more cultural data accessible, the effort of creating a new platform might be worth it.
- I miss an openGLAM platform inside "The Wikimedia Stream". An information portal like an index which shows all connections of GLAMs inside Commons, Wikisource etc. Users can find a portrait of any GLAM around the globe and the links to the digital archives or the category inside commons or wikisource etc. A representation, discussion platform and a site which volunteers can find new projects.
- A platform which is connected with Europeana and all other platforms around the globe.
- GLAMs should be where people are and use existing platforms like [opendata.admin.ch](https://opendata.admin.ch) in Switzerland and Europeana in Europe for their metadata and content.
- To distribute freely their content, GLAM have access to Wikimedia Commons, the bottle neck is that Commons is not optimized for mass uploading and that the actual strategy of relying on volunteers to help for the upload is not sustainable.
- We should combine the provision of metadata for Europeana with the provision of metadata and content for Wikimedia Commons. Presently, heritage institutions need to map their metadata both to the Europeana schema and to the metadata structure on Wikimedia Commons (which can be quite a tedious process). In an ideal world, Wikimedia Commons and Europeana would work together in order to do the metadata mapping only once. Institutions that have listed their metadata within Europeana would be able to reference high-resolution mediafiles that could be harvested along with the metadata.
- What makes open content valuable is its re-use, the new ways of contextualization, and the enhancement of metadata by the crowd. There are various initiatives to promote this (e.g. the social web initiatives of Europeana, virtual exhibitions, etc.). Wikimedia Commons is a good place to promote the integration of content in Wikipedia; however, it is not such a great place to support other ways of contextualization or the enhancement of metadata (usability, missing functionality).
- On Europeana, searching for content is difficult (usability issues, multilingual setting)