

Are Memory Institutions Ready for Open Data and Crowdsourcing? Results of a Pilot Survey from Switzerland

[Extended Abstract]

Beat Estermann
Bern University of Applied Sciences
Morgartenstrasse 2a
CH-3000 Bern 22
+41 31 848 34 38
beat.estermann@bfh.ch

ABSTRACT

Since the advent of the World Wide Web, the cultural heritage sector has undergone a series of changes. In a pilot survey among memory institutions (galleries, libraries, archives, museums) in Switzerland we have focused on two recent trends – open data and crowdsourcing – asking to what extent heritage institutions are ready to adopt open data policies and to embrace crowdsourcing strategies. The results suggest that so far, only very few institutions have adopted an open data policy. There are however signs that this may soon change: A majority of the surveyed institutions considers open data as important and believes that the opportunities prevail over the risks. Some obstacles however still need to be overcome, in particular the institutions' reservations with regard to "free" licensing and their fear of losing control. With regard to crowdsourcing the data suggest that the adoption process will be slower than for open data. Although approximately 10% of the responding institutions seem already to experiment with crowdsourcing, there is no general breakthrough in sight, as a majority of respondents remain skeptical with regard to the benefits.

Categories and Subject Descriptors

K4.3 [computers and society]: organizational impacts – computer supported collaborative work; K4.4 [computers and society]: electronic commerce – electronic data interchange (EDI); intellectual property; K6.0 [management of computing and information systems]: general – economics.

Keywords

Memory institutions, cultural heritage, open data, crowdsourcing, semantic web, digitization, participation, user engagement.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WikiSym '13 August 05 - 07 2013, Hong Kong, China
Copyright 2013 ACM 978-1-4503-1852-5/13/08 ...\$15.00.

1. INTRODUCTION

Since the advent of the World Wide Web the cultural heritage sector has undergone important changes which can be described as a series of successive and sometimes overlapping trends. Large scale digitization efforts and a tendency towards increased cooperation among memory institutions were followed by the use of web2.0 technologies and increased personalization of offers. In some cases, the users/visitors are even integrated in the "production process", thus becoming "prosumers". Thanks to projects like Wikipedia or Flickr Commons crowdsourcing and collaborative content creation have spread over the last few years. Some memory institutions cooperate with existing online communities; others have launched their own crowdsourcing projects.

Another, rather recent trend concerns the use of "free" copyright licenses and the adoption of open data policies, in order to make data available in a structured, machine readable format – "free" for anyone to be re-used, modified, integrated with other content, and re-published. Thanks to linked open data, datasets from various publishers can be integrated based on commonly shared ontologies.

While the advance of digitization efforts among memory institutions in Europe have been studied extensively (for example by the ENUMERATE project¹, the diffusion of other trends, such as open data and crowdsourcing, have hardly been investigated yet. In order to shed light on the progress in this area, we carried out a pilot survey among memory institutions of national significance in the German part of Switzerland. This extended abstract gives an overview of the study design and the key findings.

2. RESEARCH QUESTIONS

Our main research questions can be summarized as follows:

- To what extent are memory institutions ready to implement open data strategies and/or crowdsourcing approaches? How many institutions have already adopted these innovations?

¹ <http://www.enumerate.eu>

- What are the perceived risks and opportunities of open data and crowdsourcing? What are the driving forces and the hindering factors in the innovation diffusion process?
- What are the expected benefits of open data and crowdsourcing in the cultural heritage sector? Who are the beneficiaries?

3. METHODOLOGICAL APPROACH

There are around 600-700 independent memory institutions with collections of national or regional significance in Switzerland; all in all, an estimated 1000 independent memory institutions are organized in three national umbrella organizations (museums, archives, libraries).

We carried out an online survey among all the memory institutions of national significance in the German-speaking part of Switzerland. 197 organizations were contacted through 233 unique e-mail addresses. After two reminders, a total of 72 online questionnaires were completed, corresponding to 34% of the contacted organizations.

4. DESCRIPTION OF THE SAMPLE

A large majority of the responding institutions are either public institutions (58%) or private non-profits (33%). The sample consists of roughly 43% archives, 29% museums, 15% libraries, and 13% “other institutions”.

Around 70% of the overall funding of the institutions in our sample comes from public budgets (institutional funding). Individual funding situations are however quite heterogeneous: 68% of the responding institutions receive 75% or more of their overall funding from public budgets, while for 24% of the responding institutions, the share of institutional funding in overall revenues amounts to less than 25%.

With regard to the number of employees, the sample contains a good mix of institutions: Around 50% of responding institutions are small organizations with less than 5 full-time equivalents, while 10% of the sample are made up of big organizations with more than 50 full-time equivalents.

We were however able to identify several distortions in the way the institutions responded to the questionnaire (all of them are significant²):

- Archives (43%) and libraries (34%) were more likely to respond than museums (25%) and “other institutions” (20%).
- Among the institutions that had started to respond to the questionnaire, those holding “art objects” were less likely to complete the questionnaire than the others (54% compared to 79%), while those considering “collecting memory objects” as one of their core tasks were more likely to complete it than the others (80% compared to 54%).
- Interestingly, those institutions who consider “public authorities” as their main users were less likely to complete the questionnaire than the others (63% vs. 82%).

These biases and the fact that the sample size is rather small should be kept in mind when interpreting the survey results.

² Significance level: 0.05

5. KEY FINDINGS

Our pilot survey provides valuable information with regard to the diffusion of various innovative practices among Swiss memory institutions:

Digitization: The digitization of metadata and representations of memory objects as well as their deployment on the Internet are important prerequisites for Internet-based co-operations, open data, and several forms of crowdsourcing. A majority of the institutions surveyed (60%) is already active in this area and makes at least a part of their metadata and memory objects available on the Internet.

Multilateral co-operations: Over two-fifths of the institutions (43%) participate in networks in which the exchange of metadata plays an important role. A similar proportion considers the exchange of metadata as important to fulfill their core mission. There is however quite a considerable need to improve the metadata: half of the institutions surveyed indicate medium-term or even urgent need for improvement; only about a quarter say that there is no need for improvement (the other quarter wasn’t able to answer the question).

Open data: So far, only very few institutions have embraced an open data strategy. There are however signs that this may soon change: A majority of the surveyed institutions considers open data as important and believes that the opportunities prevail over the risks. Some obstacles remain however: fears of copyright violations, privacy violations and breaches of confidentiality provisions need to be overcome, and some copyright-related issues need to be resolved.³ The most important opportunities of open data from the point of view of the memory institutions in Switzerland are better visibility and accessibility of their holdings, better visibility of the institutions, and better networking among heritage institutions. The pursuit of a consistent open data strategy will however only be possible if memory institutions are able to overcome their reservations with regard to “free” licensing of works. In fact, many responding institutions would like to restrict the commercial use of works as well as their modification. The fear of losing control plays a role for 68% of the institutions surveyed and may indeed become a major stumbling block for open data in the cultural heritage sector.

Crowdsourcing: Based on the results of our survey we would expect a slower adoption process for crowdsourcing than for open data. Although approximately 10% of the responding institutions seem already to experiment with crowdsourcing, there is no general breakthrough in sight. While almost half of the institutions consider crowdsourcing as an important topic, many haven’t probably dealt with the issue yet. Furthermore, a majority of institutions is skeptical with regard to crowdsourcing: The risks are generally considered to be high, and the opportunities low. According to the surveyed institutions, crowdsourcing is most likely to be useful for the improvement of metadata as well as for correction and transcription tasks. Swiss heritage institutions however still seem to be doubtful whether crowdsourcing could result in an efficiency gain.

Linked open data: The diffusion of linked open data / semantic web technology among memory institutions in Switzerland is still

³ An issue our survey didn’t specifically cover is the orphan works problem.

in its infant stage. 29% of the institutions surveyed say it is an issue they consider; 6% are already planning their first projects.

Open data is likely to **benefit** first of all the area of education and research as well as private individuals (the general public). In addition, open data can be expected to facilitate cooperation across institutional borders and to improve the visibility of heritage institutions and their holdings. Eventually, open data might also pave the way for new data visualizations based on linked open data / semantic web technology and for various crowdsourcing approaches. The results of our study suggest however that heritage institutions in Switzerland are still far from having a clear idea how to take profit from these developments.

Also, the expected benefits need to be balanced against the **costs**. In fact, Swiss memory institutions consider the additional effort and costs related to open data and crowdsourcing as the greatest risks. In contrast, potential losses of revenue play almost no role. When it comes to promoting open data among memory institutions in Switzerland, the policies adopted by the public sector and charitable foundations, which provide the major part of the overall funding, could play a key role.

6. OUTLOOK

The questionnaire used for our pilot survey among Swiss memory institutions provides a good basis for the development of a survey instrument to be used to measure the progress with regard to the implementation of open data policies among heritage institutions in Switzerland – including those in the French and Italian speaking parts of the country that weren't included in the pilot survey. Furthermore, we are intending to use it as a basis for an international benchmark study. Please contact the author if you are interested in contributing to the international benchmark study and/or to receive a copy of the study report.

7. ACKNOWLEDGMENTS

My thanks go to Daniel Felder, David Studer, and Markus Vogler for their valuable contribution to the development of the questionnaire, its administration, and the preliminary analysis of the data. I'm also grateful to all the people who reviewed the questionnaire and made suggestions for improvement, especially to Doris Amacher (Swiss National Library), Barbara Fischer (Wikimedia Germany), André Golliez (opendata.ch), Frank von Hagel (Institute for Museum Research, Staatliche Museen zu Berlin), Alessia Neuroni (Bern University of Applied Sciences), Hartwig Thomas (Verein Digitale Allmend), and David Vuillaume (Swiss Museums Association).