

## Diffusion of Open Data and Crowdsourcing among Heritage Institutions: Results of a Pilot Survey in Switzerland

Beat Estermann<sup>1</sup>

<sup>1</sup> Bern University of Applied Sciences, E-Government Institute, Bern, Switzerland, beat.estermann@bfh.ch

Received 30 July 2013; received in revised form 13 February 2014; accepted 28 February 2014

### Abstract

In a pilot survey we examined the diffusion of open data and crowdsourcing practices among heritage institutions in Switzerland. The results suggest that so far, only very few institutions have adopted an open data / open content policy. There are however signs that many institutions may adopt this innovation in a near future: A majority of institutions considers open data as important and believes that the opportunities prevail over the risks. The main obstacles that need to be overcome are the institutions' reservations with regard to *free* licensing and their fear of losing control. With regard to crowdsourcing the data suggest that the diffusion process will be slower than for open data. Although approximately 10% of the responding institutions already seem to experiment with crowdsourcing, there is no general breakthrough in sight, as a majority of respondents remain skeptical with regard to the benefits. We argue that the observed difference in the dynamics of the diffusion of these innovations is primarily due to the fact that crowdsourcing is perceived by heritage institutions as more complex than open data, that it isn't readily expected to lead to any sizeable advantages, and that adopting crowdsourcing practices may require deeper cultural changes.

**Keywords:** Heritage institutions, Open data, Open content, Crowdsourcing, Diffusion of innovations

## 1 Introduction

In recent years, more and more heritage institutions are making their data and content available under free copyright licenses, so that they can be re-used, modified and distributed by anybody for any purpose at no cost. In fact, open data holds many promises for the heritage sector when it comes to connecting datasets of various institutions and encouraging the creation of new value-added services or new artistic creations. Heritage institutions also increasingly engage in crowdsourcing practices and online collaborative projects, such as Wikipedia, which allow them to involve their audiences in novel ways, to enhance their metadata and content, and to make cultural objects available in new contexts.

Since the advent of the World Wide Web the cultural heritage sector has undergone important changes that have taken the form of a series of successive and sometimes overlapping trends: Since the early 2000s widespread digitization of heritage objects and their metadata has been pursued as a strategic goal (as exemplified in Europe by the Lund Action Plan for Digitization [16], [17]). Digitization in turn spurred increased cooperation and coordination among heritage institutions in order to set up common catalogues with a single-point-of-access, to create virtual libraries, or to coordinate digitization efforts and long-term archiving [18], [24]. Thus, digitization not only assists preservation of cultural heritage, but has turned out to be a powerful means to expand access to collections for wider audiences [24], [26]. Half a decade later, heritage institutions started to embrace the use of web 2.0 tools, such as Facebook or Twitter, to get their messages out to their publics, and to engage them in conversations. In some cases, the users/visitors are even integrated in the *production process*, thus becoming *prosumers*. Over the last few years, crowdsourcing and collaborative content creation have spread thanks to projects like Wikipedia or Flickr Commons. Some heritage institutions cooperate with existing online communities; others have launched their own crowdsourcing projects [9], [26], [27]. 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 [22].

While the advancement of digitization efforts among heritage institutions in Europe is being monitored both at a national and international level (see [3] or [32]), the diffusion of other trends, such as open data and crowdsourcing, have hardly been investigated yet. In order to bridge that gap, a pilot survey among heritage institutions in Switzerland was carried out [15]. The purpose was to create an instrument that allows measuring the level of adoption of open data policies and crowdsourcing practices among heritage institutions in order to inform the main stakeholders about the developments in this area and to get an overview of the main challenges and driving forces. In this article we first provide an overview of previous research regarding the adoption of open data and crowdsourcing by heritage institutions. We then present key findings from the Swiss pilot survey, relating them to earlier research and discussing them in the light of innovation diffusion theory, and conclude the article with a series of suggestions in view of further research.

## 2 Definition of Core Concepts

In the following, we shall shortly introduce the core concepts referred to in this article, such as open data, linked open data and crowdsourcing, as well as the theory of innovation diffusion that serves as our primary theoretical lens.

### 2.1 Open Data

The open data movement, which had taken its origin in academic circles more than 50 years ago, experienced its worldwide breakthrough some five years ago when the Obama Administration and the UK Government adopted *Open Government Data* policies in order to promote transparency, participation, and collaboration between politicians, public authorities, private enterprises, and citizens. The term *data* includes all kinds of data: study reports, maps, satellite photographs, pictures and paintings, weather data, geographical and environmental data, survey data, the genome, medical data, or scientific formulas [7]. Open data has been hailed for its innovative capacity and transformative power [36].

According to the Sunlight Foundation's ten *Open Data Principles* [33] which serve the open data movement as a reference, data are considered as *open* if they can be re-used, modified and distributed by anybody for any purpose at no cost. In order to facilitate re-use, the data need to be made available in a machine readable format, i.e. as structured data. Typically, open data or content that is subject to copyright protection is made available under a *free* copyright license, which allows users to freely modify and to re-distribute a work.

### 2.2 Linked Open Data

While the call to open up public sector information can be seen as a logical extension of the freedom of information regulations that have been adopted by many countries since the 1990's, the open data movement is also driven by a

technical and economical vision: a *semantic web* is to be created by linking many *open* datasets from various sources. Thus, *linked open data* will serve as an infrastructure resource for third parties to build value-added services on top of it, such as new combinations of data, visualizations, or other data-driven services [5], [22].

## 2.3 Crowdsourcing

The term *crowdsourcing* was coined by Jeff Howe in 2006 in Wired Magazine, combining the two terms *crowd* and *outsourcing*: "Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers" [21]. The term has since been used with somewhat varying definitions; Estellés-Arolas and González-Ladrón-de-Guevara have compared forty original definitions of crowdsourcing in order to propose a comprehensive definition: "Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task" [14]. p. 9.

## 2.4 Innovation Diffusion

For more than half a century, scholars in various fields have studied how and under which conditions innovations spread through social systems. According to Everett M. Rogers, who has popularized the *innovation diffusion* approach, "an innovation is an idea, practice, or object that is perceived as new by an individual or other unit of adoption" [28]. p. 36. The diffusion of an innovation is a social process that unfolds as the members of a social system get acquainted with an innovation and go through the innovation decision process. Thereby, "an individual (or other decision-making unit) passes from first knowledge of an innovation, to the formation of an attitude toward the innovation, to a decision to adopt or reject, to implementation and use of the new idea, and to confirmation of this decision" [28]. p. 20.

# 3 Previous Research

This section contains an overview of previous research regarding the adoption of open data and crowdsourcing by heritage institutions, followed by a discussion of how open data and crowdsourcing relate to each other and an outline of the key elements of innovation diffusion theory that will be referred to later on in the article.

## 3.1 Open Data in the Cultural Heritage Domain

Research regarding the adoption of open data practices among heritage institutions is still relatively scarce. Baltussen et al. [4] describe the approach several organizations had been pursuing since 2011 in the Netherlands in order to create an open data ecosystem in the cultural heritage sector. Based on two expert workshops with cultural institutions they identified the main benefits and risks of opening up cultural data. They found that the number one concern among cultural heritage professionals was that opening up collections would result in material being spread and reused without proper attribution to the institution. Related to this was a perceived loss of control over the collections. Concerning financial aspects, the workshop participants did not fear a direct loss of income by making data openly available, but were afraid that they may fail to generate extra income in the future as third parties develop new business models based on their datasets. Related to the perceived loss of attribution and control was also a perceived loss of brand value. Finally, concerns regarding privacy violations were an issue for organizations that hold data containing personal information. Overall, the workshop participants agreed that open data should be part of an institution's public mission, especially if it received public funding. In their view, making collections widely accessible was at the heart of the majority of cultural heritage institutions. Furthermore, the cultural heritage professionals expected to be able to enrich data through aggregators like Europeana or other parties and to link their open data to that of other, related collections. Being able to increase the amount of channels by which end users can be reached was also seen as an important benefit of open data. As a consequence, the workshop participants also expected benefits in terms of better discoverability, which drives users to the provider's website. Further perceived benefits were increased relevance of institutions and the possibility of attracting and interacting with new customers.

These findings partly reflect earlier findings by Eschenfelder and Caswell [13] who surveyed 234 *innovative* cultural heritage institutions in the United States in order to tackle the question in which cases cultural institutions ought to control reuse of digital cultural materials. The main motives mentioned by archives, museums, and libraries for controlling the access to their collections were: (i) avoiding misuse or misrepresentation, (ii) ensuring proper object description and repository identification, (iii) avoiding legal risk, as well as (iv) donor or owner requirements. Among the top five reasons why they would limit the access to their collections, archives also mentioned generating income, libraries the impossibility to obtain the necessary rights, and museums their unwillingness to give up control over information about endangered or valuable objects, animals, or cultural events/items. The main motives against controlling the access, and thus in favor of opening up their collections, were (i) the belief that open collections have

greater impact, (ii) concerns about legal complexity when access had to be regulated for various user groups, and (iii) the institutional mission, policies or statutory requirements.

Some of the legal concerns are likely to be absent in the case of public domain works. Kelly [23] examined the policies on image rights at eleven art museums in the United States and the United Kingdom, when the underlying works are in the public domain. Investigating how and why the museums had arrived at their approach and what key changes resulted from the policy, she found that providing open access was a mission-driven decision, but that different museums looked at open access in different ways. For some it was primarily a philosophical decision, while for others it was also a business decision. For most museums, developing and adopting an open access policy was an iterative and collaborative process, with many stakeholders working together to come up with an appropriate approach. Staff at many museums cited the following critical factors that favored the adoption of an open access policy: diminishing revenues, difficulties when it came to drawing a line between scholarly and commercial uses of their images, senior management support for an open access policy, as well as technical innovations that enabled images to be made accessible with greater ease. In the process, they had to overcome a series of concerns, such as fears regarding the consequences of loss of control, challenges regarding metadata quality, technical challenges when it came to providing access to the collections through the museum's website, as well as a possible loss of revenue. Most museums reported positive outcomes of opening up their collections. Staff mentioned the goodwill and recognition that came with open access, as well as a sense of satisfaction at helping to fulfil the institution's mission. Virtually all museums experienced increased website traffic, and in some cases, curators received better and more interesting inquiries from scholars and the public. There were also positive side effects in that the policy change forced the institutions to think through their policies and their implications, as well as in form of improved technology skills among the staff members. Some museums mentioned downsides of the open access approach: For museums without automated delivery systems, increasing numbers of image requests had led to an increase in workload. Thus, an increased demand may result in a need for investments in the technical infrastructure. Unsurprisingly, most museums in the survey reported stable or lower revenue from rights and reproductions. And finally, some museum staff mentioned that it had become more difficult to track the use of images or objects in their collections.

It has to be noted that many of the cases cited by Kelly [23] relate to museums that did not comply with the Sunlight Foundation's *Open Data Principles*, but pursued open access approaches that were limited to educational or scholarly use only, even for works that were in the public domain. In the case of US institutions claiming copyright over faithful reproductions of two-dimensional works, such approaches most likely amount to copyright overreaching [10]. Copyright overreaching occurs when claims of copyright protection are made that overreach the bounds of justifiable legal rights. Examining policies from U.S. museums, Crews [10] found four varieties of copyright overreaching: assertions of false copyrights; claims to copyrights not held by the museum; assertion of control beyond rights of copyright; and claims of quasi-moral rights. He identified four motivations for copyright overreaching: protecting the integrity of art, generating additional revenue, getting credit for the museum's collections and other good work, as well as adherence to donor requirements.

### 3.2 Crowdsourcing in the Cultural Heritage Domain

There are plenty of examples of crowdsourcing approaches in the cultural heritage sector. Several authors have created inventories of crowdsourcing projects throughout the world [8], [20], [25]-[26], [30], [34]. Based on these inventories, typologies have been developed: Oomen and Aroyo [26] propose a classification scheme based on the digital content life cycle model of the National Library of New Zealand, distinguishing the following types of crowdsourcing approaches: correction, classification, contextualisation, co-curation, complementing collections, and crowdfunding. They also point to the fact that crowdsourcing initiatives in the cultural heritage domain may be executed without institutions being in the lead. They expect that more and more crossovers will take place between community- and organization-driven projects, as is the case with co-operations between heritage institutions and the Wikipedia community. This observation matches the insights gathered by Terras [34] who investigated amateur online museums, archives, and collections and concluded that the best examples of these endeavors can teach best practice to traditional heritage institutions in how to make their collections useful and to engage a broader user community. She not only recommends that heritage institutions increasingly use web 2.0 services such as Flickr, Twitter, and Facebook to build an online audience, but also encourages them to bridge the gap between pro-amateurs with their private collections of ephemera, and institutional collections.

Smith-Yoshimura and Schein [30], investigating social metadata approaches, developed a typology of crowdsourcing approaches that is slightly different from the one proposed by Oomen and Aroyo [26], and applied it to a sample of 24 websites from the cultural heritage domain which engage their communities and seek user contributions by providing social media features, such as tagging, comments, reviews, images, videos, ratings, recommendations, lists, or links to related articles. They found that within their sample of 24 websites, 16 used crowdsourcing for data enhancement in the form of improving description, 11 for collection and content building, and 10 for data enhancement in the form of improving subject access. Further areas of crowdsourcing were: ratings and reviews (i.e. for collecting subjective opinions), promoting activities outside of the site, sharing and facilitating research, as well as networking and community building.

In addition to social media functionalities built into institutions' websites, Smith-Yoshimura and Schein also investigated the use heritage institutions make of third party social media sites, such as LibraryThing, Flickr, Youtube, Facebook, Wikipedia, and blogs. Based on comparative case descriptions, they reached the following conclusion:

"LibraryThing is an excellent resource from which to harvest user-generated metadata on published works and disseminate information on one's own holdings of published materials, but impractical for unique or unpublished works. Flickr is an unparalleled vehicle for sharing still images and gathering user-generated description of the images. YouTube is the leading site for promoting and sharing moving images. Facebook provides an avenue through which LAMs [i.e. Libraries, Archives, and Museums] can communicate textually and imbed audio, video, and images. Twitter is an efficient way to push out short textual messages, such as announcements and alerts. Wikipedia offers the potential to reach a broad audience and direct web traffic to a LAM and its select resources. Blogs, especially those built in-house, are perhaps the most adaptable platform for communicating various formats of information through an interface that can be functionally and visually tailored to suit institutional needs. Establishing a presence on social networking sites, wikis and blogs enables LAMs to bring their resources to online environments where users are already active, exposing content to new audiences, encouraging user interaction, and fostering a sense of community" [30], p. 64.

Regarding the numbers of heritage institutions using third party social media sites, they report that 1600 libraries worldwide used LibraryThing to harvest user-generated content and to enhance the descriptions of published works in their online public access catalogues. For the other types of social media services, they report findings from a survey carried out among special collections and archives in academic and research libraries in the United States and Canada [11]. According to that study, 49% of the 169 responding institutions indicated that they were using institutional blogs, 39% had a social networking presence, 37% reported adding links to Wikipedia, 30% used Flickr, roughly one quarter used Twitter (25%), YouTube (24%) or Podcasting (24%), 17% had an institutional wiki, 15% collected user-contributed feedback (e.g. through social tagging), and 10% used mobile applications to reach out to their audiences. Responding institutions were also asked which of these services they were planning to implement within a year. Here, institutional blogs rated highest with 19%, followed by user-contributed feedback (16%). Regarding the publication of heritage content on Wikipedia, the first core survey of the ENUMERATE project revealed that among the 774 responding European heritage institutions, on average 3% of their digital collections is accessible through Wikipedia [31].

Holley [20] insists on the difference between social engagement (e.g. social tagging) and crowdsourcing, arguing that crowdsourcing usually entails a greater level of effort, time and intellectual input from an individual. According to her, crowdsourcing relies on sustained input from a group of people working towards a common goal, whereas social engagement may be transitory, sporadic or done just once. As a consequence, setting up a crowdsourcing project is about "using social engagement techniques to help a group of people achieve a shared, usually significant, and large goal by working collaboratively together as a group" [20]. She argues that libraries are already proficient in social engagement with individuals, as many forms of social engagement in libraries pre-date the advent of the Internet, but that they are not necessarily proficient yet at defining and working towards group goals. Oomen and Aroyo [26] point to motivating users for participation and supporting quality contributions as major challenges of crowdsourcing.

There is hardly any research into heritage institutions' motivations for crowdsourcing. In an attempt to fill that gap, Alam and Campbell [1] carried out a case study to investigate organizational motivations for crowdsourcing by the National Library of Australia. They found that the institution was motivated by a set of attributes that dynamically changed throughout the implementation of the crowdsourcing project, ranging from resource constraints to utilizing external expertise through to social engagement. The researchers noted that the project resulted in a high level of social engagement, active collaborations with and between stakeholders, and development of *bridging* social capital that in turn instigated further motivations for the organization. They concluded that this dynamic change of organizational motivation may well be crucial for the long-term establishment of crowdsourcing practices.

### 3.3 How do Open Data and Crowdsourcing Relate to each Other?

The link between heritage institutions' adoption of open data policies and their engagement in crowdsourcing approaches hasn't been studied explicitly yet; there are however several indications that these practices may converge: Flickr Commons, for example, requires that the images made available by heritage institutions be either in the public domain or *have* "no known copyright restrictions" [30]. At the same time, it invites its users to help describe the photographs they discover on Flickr Commons, either by adding tags or leaving comments [34]. Similarly, the Wikipedia/Wikimedia community requires that content provided to its projects be made available under a *free* copyright license, which allows third parties to share, modify, and re-distribute the content [35].

This convergence of open data policies and crowdsourcing approaches is fully in line with Alam's and Campbell's observation of a shift from egoistic motives towards a more public value focus as heritage institutions engage in true collaboration with their crowdsourcing communities [1]. It is also reminiscent of Wyatt's remark that Wikipedia should not be described as a product of user-generated content, *sitting alongside blogging, social-networking and video sharing websites*, but that it is far better understood as a place of community curated works, "where the individual Wikipedian is not merely a *user* of a corporation's infrastructure but also potentially the author, reader, reviewer and maintainer of every aspect of the project content, code and community" [35]. Cooperating with the

Wikimedia/Wikipedia community thus requires heritage institutions to subscribe to the community's public value orientation, which calls for a release of data and content under *free* copyright licenses.

The convergence of user participation, inter-organizational cooperation, and open data is also reflected in Oomen and Aroyo's vision of a "more open, connected and smart cultural heritage: open (the data is open, shared and accessible), connected (the use of linked data allows for interoperable infrastructures, with users and providers getting more and more connected), and smart (the use of knowledge and web technologies allows us to provide interesting data to the right users, in the right context, anytime, anywhere – both with involved users/consumers and providers)" [26].

### 3.4 Diffusion of Innovations

In our study we use the *innovation diffusion* approach as a theoretical lens to study where heritage institutions stand with regard to the adoption of open data policies and the engagement in crowdsourcing approaches. As Rogers [28] notes, the diffusion approach is particularly well suited to connect research and practice. Thanks to a wide application of the approach in various fields, many insights into the innovation diffusion process as such have been gathered that can be applied to inform stakeholders in new areas of innovation. In the following, we will shortly outline the elements of innovation diffusion theory we draw upon in this paper.

*Decision stages:* the innovation adoption process has been widely described as comprising different, successive stages, although the number of stages, their precise definition, and their naming varies according to the authors. The stage model developed by Beal and Bohlen [6] comprises five distinct stages of innovation adoption: awareness stage, interest stage, evaluation stage, trial stage, and adoption: At the *awareness stage*, agents become aware of some new idea, but lack details concerning it. At the *interest stage*, they are seeking more information about the idea, and at the *evaluation stage*, they make a mental trial of the idea by applying the information obtained in the previous stage on their own situation. At the *trial stage*, they apply the idea in a small-scale experimental setting, and if they decide afterwards in favor of a large-scale or continuous implementation of the idea, they have reached the *adoption stage*. The stage model was originally developed in order to understand the innovation adoption process of individuals. When applied to organizations, it has to be kept in mind that individual organizations may not pass through the stages in a linear fashion, but may move back and forth between stages in a process that is characterized by shocks, setbacks, and surprises [19]. In practice, a differentiation of decision stages can be useful to choose the appropriate communication channel to promote an innovative practice. As Rogers [28] notes, mass communication channels are relatively more important at the awareness stage, while interpersonal channels are relatively more important at later stages in the innovation-decision process.

*Adopter categories:* Rogers uses adopter categories to classify the members of a social system on the basis of innovativeness. Different adopter types assimilate an innovation at different moments of the innovation-diffusion process. Five adopter categories are distinguished: (i) innovators, (ii) early adopters, (iii) early majority, (iv) late majority, and (v) laggards. These categories represent *ideal types* that were created for analytical purposes. While investigations regarding the characteristics of different adopter categories and their role in the innovation process have led to many valuable insights [28], it has been criticized that the adopter categories, with their stereotypical and value-laden terms, fail to acknowledge adopters as actors who interact purposefully and creatively with complex innovations; the use of adopter categories as explanatory variables for innovation adoption should therefore be avoided [19]. In dealing with later adopters it should also be kept in mind that they have been found to be more likely to discontinue innovations than earlier adopters – either because they lack the necessary know-how to adapt the innovation to their particular circumstances, or because innovations don't fit their economic conditions [28].

*Perceived attributes of innovations:* The adoption rate of an innovation refers to the length of time required for a certain percentage of the members of a system to adopt an innovation [28]. Much of the variance in innovations' adoption rate is explained by key attributes of innovations as perceived by prospective adopters [19], [28]. Rogers identifies the following key attributes:

- "Relative advantage is the degree to which an innovation is perceived as better than the idea it supersedes" [28]. p. 15. In the assessment of an innovation, economic aspects play an important role, but also social prestige factors, convenience, and satisfaction. Thereby, the individual perception is important, and not the *objective* advantage.
- *Compatibility is the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters. An idea that is incompatible with the values and norms of a social system will not be adopted as rapidly as an innovation that is compatible. The adoption of an incompatible innovation often requires the prior adoption of a new value system, which is a relatively slow process (ibid.).*
- "Complexity is the degree to which an innovation is perceived as difficult to understand and use. [...] New ideas that are simpler to understand are adopted more rapidly than innovations that require the adopter to develop new skills and understandings" [28]. p. 16.

- *Trialability is the degree to which an innovation may be experimented with on a limited basis. New ideas that can be tried on the installment plan will generally be adopted more quickly than innovations that are not divisible. [...] An innovation that is trialable represents less uncertainty to the individual who is considering it for adoption, as it is possible to learn by doing (ibid.).*
- *Observability is the degree to which the results of an innovation are visible to others. The easier it is for individuals to see the results of an innovation, the more likely they are to adopt. Such visibility stimulates peer discussion of a new idea (ibid.).*

## 4 Research Questions and Methodology

The primary motivation for our research was to create an instrument that allowed measuring the level of adoption of open data policies and crowdsourcing practices among heritage institutions in Switzerland in order to inform the main stakeholders (heritage institutions, policy makers, as well as open data and free knowledge activists) regarding the developments in this area and to get an overview of the main challenges and driving forces. In the following we will present the research questions, describe the methodological approach and the survey instrument, and discuss the sample biases as well as the limitations of the approach.

### 4.1 Research Questions

Our main research questions can be summarized as follows:

- Where are Swiss heritage institutions situated in the innovation-decision process regarding the adoption of open data strategies and the engagement in crowdsourcing practices?
- What are the perceived risks and opportunities of open data and crowdsourcing among heritage institutions? What are the driving forces and the hindering factors regarding the diffusion of these innovations?
- What are the expected benefits of open data and crowdsourcing in the heritage domain? Which are the expected beneficiaries?

### 4.2 Methodological Approach

According to the criteria applied by the national service for the protection of cultural property, there are between 600 and 700 independent heritage institutions with collections of national or regional significance in Switzerland. An estimated total of 1000 independent heritage institutions are organized in three national umbrella organizations (museums, archives, libraries).

For the survey, a subset consisting of all the heritage institutions of national significance in the German-speaking part of Switzerland was selected. The focus on institutions with collections of national significance ensures the institution's relevance with regard to open data and crowdsourcing (excluding for example lending libraries). The limitation to the German-speaking part of Switzerland (which corresponds to a bit more than 60% of all collections of national significance) was due to time and financial constraints.

As the register of collections of national significance lists collections and not institutions, we cleaned out obvious double entries where one institution is responsible for several collections. In some cases we had contact addresses for several sub-divisions of the same legal entity where it isn't apparent from the outside to what extent they act as autonomous entities (e.g. in the case of universities where several sub-divisions have their own libraries or archives). Eventually, 197 organizations were contacted through 233 unique e-mail addresses in the first half of November 2012. After two reminders in 10 day intervals, the online questionnaires had been completed by 72 respondents from 65 different legal entities, corresponding to 34% of the contacted organizations.

Wherever possible, the e-mail invitations were sent to the official contact addresses for the collections – partly personal e-mail addresses of staff members, partly general institutional e-mail addresses. The survey was set up in a way that the link could easily be passed on to other staff members and the questionnaire could be filled in by several people and at different times, thus allowing the institutions to have the most competent staff member reply to a particular question or to gather extra information internally when needed. There is anecdotal evidence that this internal coordination took place and that closely cooperating units that received several invitations to participate in the survey filled in the questionnaire only once. As a side-effect of the flexibility of the questionnaire, several institutions that completed the questionnaire left a small number of questions without a response; these questionnaires were included in the analysis.

### 4.3 Survey Instrument

The questionnaire was elaborated in an iterative process: an initial version was produced based on an analysis of existing research literature, interviews with practitioners, and project reports. In a second step, feedback was solicited from different experts from various backgrounds (heritage professionals, open data and open knowledge activists, and researchers), and in a third step, a pre-test was carried out among 10 institutions, accompanied by a follow-up interview in order to better understand respondents' reactions to the questionnaire. After each step, an improved version of the questionnaire was produced.

In its final version, the questionnaire contained 21 questions: Seven questions related to the institutions' characteristics, such as the most characteristic type of heritage items, their main activities, their users, the number of employees, the composition of revenue sources, the institution's legal form, and the percentage of holdings pre-dating 1850. Three questions addressed the issue of metadata exchange with other institutions, metadata quality, and the availability of data and content on the Internet. One question asked whether open data and crowdsourcing (collaborative content creation) was considered important for the institution, and four batteries of questions addressed the risks and opportunities of open data and crowdsourcing respectively. The remaining questions related to the institutions' experience with *free* copyright licenses, their interest for linked data, the role different types of volunteer work (including online volunteers) play for the institution, the involvement of staff members in collaborative projects on the internet, and the institution's interest in further information about open data and crowdsourcing. Wherever possible, a 4-point Likert scale was used. The survey instrument has been published along with the study report and is available online [15].

### 4.4 Sample Biases

Compared to the entire population of heritage institutions in Switzerland, the sample has several biases that result from the selection criteria:

- All the institutions that were contacted hold collections that are rated as of *national significance* by the government office responsible for the protection of cultural heritage. We can therefore assume that virtually all larger institutions with important collections have been contacted, while smaller institutions and those with less important holdings are underrepresented.
- Institutions in the Italian and French speaking regions of Switzerland were not contacted for reasons of time and cost. Also there are no empirical observations that would suggest any notable differences between the language regions. This selection criterion introduces however a bias in favor of federal institutions and private institutions with a national scope, as many of them are located in the Bern area. On the other hand, the sample does not include the international organizations located in the Geneva region.

Several distortions in the way the institutions responded to the questionnaire were identified (all of them are significant at a confidence level of 95%):

- Archives (43% of contacted institutions) and libraries (34%) were more likely to respond than museums (25%) and *other institutions* (20%). These numbers were calculated on the basis of our own categorization based on the institutions' name and e-mail address.
- Among the institutions that had started to respond to the questionnaire (99 respondents answered at last 2 questions), those holding *art objects* were less likely to complete the questionnaire than the others (54% compared to 79%), while those considering *collecting heritage objects* as one of their core tasks were more likely to complete it than the others (80% compared to 54%).
- Interestingly, those institutions which consider *public authorities* as their main users were less likely to complete the questionnaire than the others (63% vs. 82%).

As most drop-outs took place right after the completion of the first set of questions relating to the general characteristics of the institution, it can be assumed that respondents who did not continue to fill in the questionnaire did not feel sufficiently concerned by questions relating to open data and crowdsourcing. As a consequence, the survey results may be somewhat biased in favor of institutions which think open data or crowdsourcing are relevant.

### 4.5 Limitations

Due to the small sample size we limited ourselves to analyzing the sample as a whole. We are planning to analyze the influences of various factors on the adoption of open data and crowdsourcing as well as differences between types of institutions in a future study with a larger sample, which will yield more robust results.

## 5 Description of the Sample

A large majority of the responding institutions are either public institutions (58%) or private non-profits (33%). Only 6% are or belong to private, profit-oriented institutions. 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 at least three-quarters 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 one quarter. 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 is made up of larger organizations with more than 50 full-time equivalents.

Asked about their users, the surveyed institutions most frequently mentioned private individuals (89%), education (89%), and research (73%). Cultural institutions (45%), public authorities (31%), and private enterprises (21%) were mentioned by less than half of the institutions. As to the heritage objects that are characteristic for their institutions, more than half of the respondents mentioned *images, photographs, prints* (56%). Other frequently mentioned object types were *books, periodicals* (46%), *manuscripts, autographs* (44%), and *documents, records* (44%). Roughly one quarter of responding institutions mentioned *film documents* (28%) or *audio documents* (25%), while the other object types – *objects of art* (18%), *technical objects* (14%), *craft artefacts* (10%), and *natural-history objects* (8%) – were mentioned less frequently. (Here and in the following paragraph, we are reporting the items that scored 1 – *is the case* – on a 4-point Likert scale.)

The responding institutions show a certain level of homogeneity with regard to their tasks: All the tasks mentioned in the questionnaire scored quite high, with at least 69% of responding institutions considering them at least partly as their core tasks. Thus, over 80% of the responding institutions count *collecting, archiving, and preparing, indexing, documenting* clearly among their core tasks. On the other end of the spectrum, the least often mentioned tasks were *researching, investigating* (37%), *digitizing* (39%), *lending to other institutions* (42%), *exhibiting* (45%), and *restoring, conserving, preserving* (46%).

## 6 Main Findings

In this section we relate the main findings with regard to the research questions:

### 6.1 Diffusion of Open Data and Crowdsourcing

In order to estimate the share of institutions that presently pursue a publication strategy that is in line with the *Open Data Principles*, we looked at the institutions that already make heritage objects available on the Internet and analyzed their responses to the question under which conditions they would make heritage objects available online at no charge. As shown in figure 1, between 1% and 7% of responding institutions make scans/photographs of their heritage objects *freely* available on the Internet. Over half of them make them available on the Internet, but with restrictions. 40% do not make them available at all.

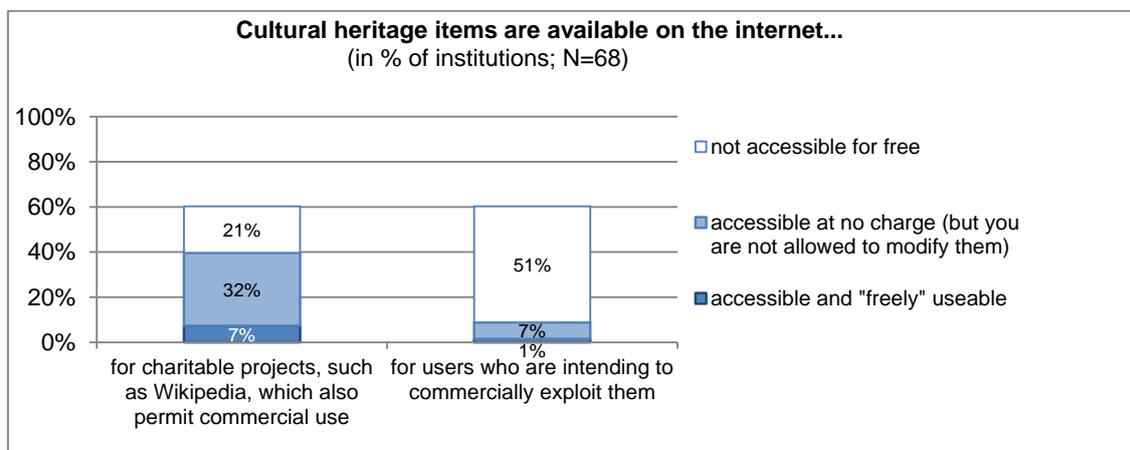


Figure 1: Availability of reproductions of heritage objects on the Internet (and limiting conditions)

When looking in more detail at the conditions under which heritage institutions would make heritage objects *freely* accessible on the Internet (provided that they would not infringe any third party rights or legal requirements), we can observe a descending order as to the type of use they would like to allow: education and research score highest

(76% clearly are in favor of free access for these groups), followed by non-profit projects (60%) and private use (59%). When asked about non-profit projects which also allow commercial use of the data, such as Wikipedia, the institutions' readiness to make their works available clearly decreased (29%), but was still much higher than for commercial use only (7%) (see figure 2); the differences between the scores obtained for charitable projects/private use, Wikipedia, and commercial users are significant at a confidence level of 95%.

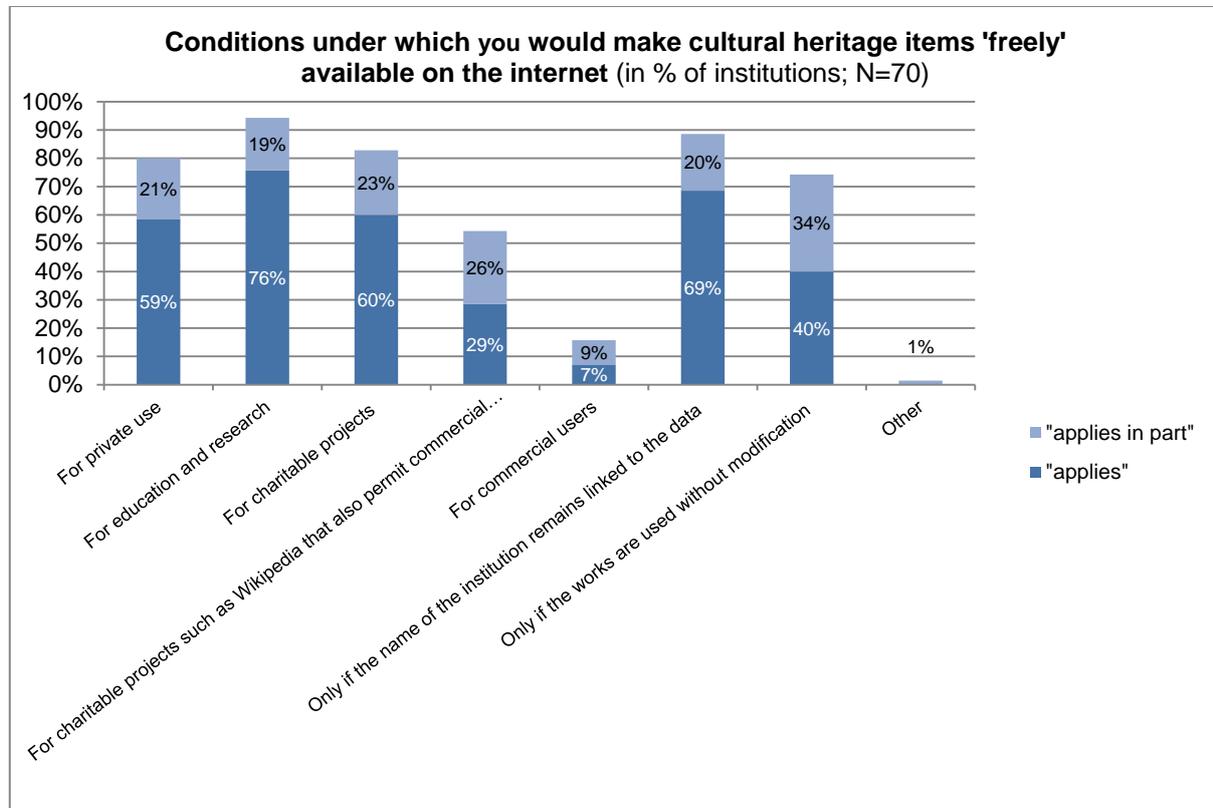


Figure 2: Conditions under which institutions would make cultural heritage items *freely* available on the internet

As the *Open Data Principles* prohibit discrimination with regard to the possible uses of the data, overcoming the reluctance among heritage institutions to admit commercial use of their data/content without requesting payment could be a major challenge. Another challenge results from the fact that 74% of the responding institutions indicated that they would at least partly want to restrict the right to modify the data/content – which is also not in line with the *Open Data Principles*. And finally, the data suggest that over 50% of the heritage institutions which make their heritage objects available on the Internet do not understand that it is impossible to make works available for the use in Wikipedia and to simultaneously prevent their modification and/or their commercial use. In fact, there seems to be a certain lack of awareness of *free* copyright licenses. This is also reflected by the fact that most institutions (83%) indicated that they did not have any experience with alternative licensing models, such as Creative Commons licenses.

Our findings are in line with the observations made in earlier studies that most museums had differential pricing for commercial, nonprofit, and scholarly clients when licensing content, and that fees are often waived for educational and scholarly use [2], [23]. Interestingly, the institutions in our sample seem to be inclined to waive fees for their main user groups, namely private individuals, education, and research. The apparent lack of understanding of *free* copyright licenses is reminiscent of Baltussen's observation that the lack of copyright knowledge and the lack of (up-to-date) information about the copyright status are seriously impeding the ability of cultural institutions to open up their collections [4].

While the above observations point to several challenges when it comes to establishing *free* licensing among heritage institutions, there are other data that suggest that the responding institutions have a rather positive attitude towards open data: When relating the perceived risks to the perceived opportunities, it appears that for 80% of the responding institutions the opportunities of open data outweigh the risks; for more than 40% this is clearly the case. Furthermore, when asked about the importance of open data, more than half of the institutions responded to the affirmative, while only about 20% said that the topic was not important to them. Among those which consider open data to be important, all but one rated the opportunities higher than the risks. This can be seen as a further indicator that we may observe a highly dynamic diffusion of open data in a near future.

Based on the data collected in our survey (and thus disregarding a possible sample bias), it can be assumed with a probability of 0.95 that 49% to 71% of cultural heritage institutions in Switzerland make at least some of their metadata and representations of their heritage objects available on the Internet. However, no more than 3% provide their content under *free* copyright licenses. At the same time, 72% to 90% perceive open data as an opportunity, and 41% to 65% consider the subject to be of importance. Thus, in terms of the innovation diffusion model, only between 0% and 3% of the heritage institutions in Switzerland had fully adopted an open data policy by the end of 2012, while roughly half of them had reached the interest or evaluation stage.

In order to estimate the share of institutions that already engage in crowdsourcing practices, two indicators were used: staff involvement in collaborative projects and the perceived importance of online volunteering. For both indicators, we got similar results: 11% of responding heritage institutions have staff members who contribute to Wikipedia as part of their professional activity, and 10% of responding institutions say that online volunteering plays partly an important role for them. Interestingly, no correlation was found between the two variables. This seems to indicate that the institutions which have some of their staff members contribute to Wikipedia during their work time do not associate this activity with the online volunteering activity of the wider Wikipedia community.

As for open data, we constructed a crowdsourcing desirability index by relating the perceived risks to the perceived opportunities. It turns out that the surveyed institutions are much less optimistic with regard to crowdsourcing than with regard to open data. For over 90% of them the risks of crowdsourcing are at least as great as the opportunities; for half of them the risks clearly prevail. Adding to this rather pessimistic outlook is the fact that even institutions which perceive crowdsourcing as important (28%) or as very important (10%), think that the risks of crowdsourcing prevail over the opportunities. This could be an indicator that heritage institutions consider it as quite a great and time consuming challenge to enter a cooperative relationship with an existing online community or to launch their own crowdsourcing project. Given the importance attributed to crowdsourcing by at least some of the institutions, it can be expected that they will be willing to confront these challenges as many others have done before them in other countries.

Based on the data collected in our survey, it can be assumed with a probability of 0.95 that between 4% and 18% of heritage institutions in Switzerland are already involved in Wikipedia projects, while 3% to 17% consider voluntary work in the online sector partly as important. There is, however, no significant correlation between the two aspects. While only 1% to 13% of cultural heritage institutions in Switzerland consider crowdsourcing as an opportunity, between 27% and 49% regard it as important. Thus, in terms of the innovation diffusion model, it appears that by the end of 2012, roughly one tenth of Swiss heritage institutions had entered the trial stage with regard to crowdsourcing. However, none of the institutions surveyed seems to have fully embraced the concept. At the same time, roughly one third of Swiss heritage institutions had reached at least the interest or evaluation stage, which is less than in the case of open data (even though this difference is not significant, possibly due to the small sample size).

The findings for open data and crowdsourcing are quite interesting in so far as they point to varying dynamics: While more institutions are already engaging in crowdsourcing practices, there seems to be more enthusiasm for open data. Based on the results of our survey, Swiss heritage institutions can therefore be expected to have higher adoption rates for open data than for crowdsourcing in a few years from now.

## 6.2 Perceived Risks and Opportunities

The perceived risks and opportunities give us further insights with regard to the driving forces and the hindering factors in view of the adoption of open data policies and crowdsourcing approaches.

### 6.2.1 Risks and opportunities of open data

When it comes to implementing an open data strategy, the responding institutions are worried most about the extra time effort and expenses to make the data available (86% consider this at least partly as a risk). 59% are also concerned about the extra time needed to respond to inquiries. Further considerable risks are loss of control (68%), potential copyright infringements (66%), violations of data protection laws (51%), and secrecy infringements (35%). Only few of them expect a loss of revenues (a mere 14% thought that this might at least partly be a risk). The opportunities mentioned most often were better visibility and accessibility of holdings (86%), better visibility of the institutions (79%), and better networking among heritage institutions (74%). 36% think that by adopting an open data strategy they would clearly improve the way they fulfill their core mission; 33% think that this is partly the case.

These findings are mostly in line with those of earlier studies [4], [13], [23], although none of them allows for a direct comparison of results. Interestingly, the extra time effort and expenses, which was perceived as the greatest challenge in our survey, was mentioned only by Kelly [23] in form of a need to improve metadata quality and investments in technical infrastructure. In the other two studies, these aspects may have been taken for granted. Similarly, the extra time needed to respond to inquiries, which was perceived as a challenge by more than half of our respondents, was mentioned only by Kelly [23]. In contrast to what might have been expected from the results of the other studies, only very few institutions in our sample were concerned about a potential loss of revenues.

Regarding possible extra investments needed when making data available online, our survey showed that more than half of the institutions felt that they needed to improve their metadata, while less than a quarter indicated that there was no need for improvement (25% of the respondents said that they couldn't answer this question). Similarly, only 23% of the responding institutions make their reproductions of heritage objects available online. An additional 37% indicated that this is partly the case. These results support Kelly's findings that the need for metadata improvement and investments in technical infrastructure are major challenges for heritage institutions that decide to make their data available under open access regimes.

### 6.2.2 Risks and Opportunities of Crowdsourcing

With regard to crowdsourcing, most of the risks respondents were asked about received very similar ratings: considerable time/effort needed for preparation and follow-up (72%), difficulties in estimating the time-effort (70%), no guarantee concerning long-term data maintenance (66%), unforeseeable results (61%), and a low level of planning reliability (60%). The only risk that was rated significantly lower was *fears among employees (job loss, changing roles and tasks)* – only 23% of the responding institutions indicated that this could partly be a problem. These findings are in line with Oomen and Aroyo's observation that motivating users for participation and supporting quality contributions are the two major challenges of crowdsourcing [26]. They also support Holley's view that many heritage institutions may be proficient in social engagement with individuals, but that they don't necessarily feel comfortable with setting up a crowdsourcing project [20].

When asked about the opportunities of crowdsourcing, the respondents were rather skeptical. The opportunity that was rated highest was *classification / completion of metadata* (31% of the respondents consider this at least partly as an opportunity), followed by *correction and transcription tasks* (30%), *enhancement and expansion of texts* (25%), *completion of collections* (25%), *crowdfunding* (24%), and *co-curatorship* (14%). Thereby it has to be noted that for the items concerning crowdsourcing opportunities the share of institutions which ticked the *not applicable* field was between 10% and 17% – which is much higher than for all the other risk and opportunity items included in the questionnaire. This could point to the fact that many institutions have not really given much thought to crowdsourcing yet. To our knowledge, this is the first quantitative assessment of the perceived importance of various types of crowdsourcing approaches in the heritage sector. Given the fact that hardly any of these institutions is actually engaging in crowdsourcing approaches, it is however primarily a hypothetical one. It also has to be noted that most of the observed differences in scores are not significant at a confidence level of 0.95, at the exception of the differences between the two highest values on the one hand, and the lowest value on the other hand.

Interestingly, the overall risk assessment by the respondents in our sample was not worse for crowdsourcing than for open data – the average scores are quite similar. What made the difference was the opportunity assessment, which was significantly better for open data than for crowdsourcing.

### 6.3 Expected Costs and Benefits

Our data suggest that extra time effort and expenses are perceived as the greatest risks or shortcomings of open data and crowdsourcing in the heritage domain. Expected losses of revenue, on the other hand, play virtually no role. This is not really surprising as the institutions in our sample reported that on average only 6% of their revenues derived from commercial activities: 3% from entrance fees, 1% from lending fees, and less than 0.5% from the sale of image rights. In fact, most institutions don't make any money by lending heritage objects or by selling image rights – the two only revenue types that one would expect to be seriously affected by a *free* licensing policy.

Concerning the expected benefits a distinction has to be made between open data and crowdsourcing: While the responding institutions expect only very limited benefits from crowdsourcing, they expect that the adoption of an open data policy would promote the networking among heritage institutions, improve the visibility of their holdings and enhance how these institutions are perceived by the general public.

The institutions were also asked about the main target groups that would benefit from an open data policy. The main target groups mentioned were research (86%), education (79%), private individuals (77%), and cultural institutions (76%). Public authorities (51%) scored significantly lower than the first three groups, and private enterprises (30%) in turn scored significantly lower than public authorities (confidence level = 0.95). These results are largely in line with the respondents' indications concerning the main users of their institutions.

## 7 Discussion

As the preceding section demonstrates, our research questions could largely be answered based on the data gathered through the pilot survey. The main limitations are the rather small sample size and the inherent inability of quantitative approaches to account for qualitative aspects and developments that have not been taken into account at the time of questionnaire development. As our review of previous research regarding open data and crowdsourcing in the heritage domain has shown, results from various studies have been published in the meanwhile, which are complementary to our approach and need to be taken into account in future quantitative studies.

While most results of our study are in line with those of earlier studies, we found rather surprisingly that very few institutions in our sample are concerned about a potential loss of revenues when adopting open data policies; in fact, they seem very much inclined to waive fees for their main user groups.

There are at least two areas where our study is breaking promising new ground: It is to our knowledge the first quantitative study examining attitudes and practices regarding open data policies and crowdsourcing among a given population of heritage institutions, and it is the first study in this area that uses the innovation diffusion model as a theoretical framework.

## 7.1 The Results in the Light of the Innovation Diffusion Model

In addition to the state of diffusion of both open data policies and crowdsourcing practices among heritage institutions in Switzerland, we were able to point out different dynamics for the diffusion of open data and crowdsourcing, which merit to be discussed in the light of earlier insights regarding innovation diffusion processes. Rogers [28] identifies a series of variables determining an innovation's rate of adoption: (i) the perceived attributes of innovations; (ii) the type of innovation-decision (optional, collective, or imposed by authority); (iii) the type of communication channels that is used to promote an innovation; (iv) the nature of the social system; as well as (v) the extent of change agents' promotion effort. In the case of Switzerland at the end of 2012, most of these variables can be assumed to be equal for crowdsourcing and open data among heritage institutions. There may have been some differences regarding the type of innovation-decision, as engaging in crowdsourcing was clearly an optional decision for each institution, whereas first official strategies had been formulated during the same year both in view of the adoption of an open government data policy in Switzerland and in view of an improved accessibility of cultural heritage on the Internet [12], [29]. As a consequence, some institutions may have anticipated an official policy in favor of open data when responding to the questionnaire. However, the main difference seems to lie in the perceived attributes of the two innovations.

As set out in section 3.4, Rogers [28] distinguishes between the five perceived attributes of an innovation. In the following, we shall shortly discuss our and earlier findings related to these five dimensions.

*Relative advantage with regard to previous solutions:* Both open data and crowdsourcing are associated with a set of risks, whereby additional effort and expense are seen as the greatest challenge. Perceived opportunities of open data are however clearly greater than those of crowdsourcing

*Compatibility with existing values, past experiences, and needs of adopters:* Regarding the adoption of an open data policy, the main cultural incompatibilities lie in the acceptance of *free* licensing of heritage objects, including for commercial use, and surmounting the fear of losing control. Possible losses of revenue are no issue for most institutions, and most of them would readily waive fees for their main users, such as research, education, and private individuals. When it comes to an engagement in crowdsourcing projects, the required cultural change may be more important. Thus, Alam and Campbell [1] describe how the motivations of the National Library of Australia changed as it engaged in a crowdsourcing project, moving from egoistic motives towards a public value orientation related to social engagement. They even conclude that the dynamic change of organizational motivation may be key to the long-term establishment of crowdsourcing practices. In a similar vein, other authors point to a shift in perceptions among cultural heritage professionals, noting that "some cultural institutions theorists argue that increased public participation should replace the façade of the infallible, omnipotent voice of the cultural institution with multiple user voices" [13]. Lori Phillips, a pioneer in the area of cooperation between heritage institutions and the Wikipedia community, has coined the term *Open Authority*. "At its most basic, Open Authority is the coming together of museum authority with the principles of the open Web, a mixing of institutional expertise with the discussions, experiences, and insights of broad audiences" [27]. She argues that museum professionals need to reconsider the definition of authority in order to remain connected to their communities, both on-site and virtual. Thus, it may well be the case that heritage institutions need to undergo a deep cultural change before being able to fully grasp and reap the benefits of crowdsourcing.

*Complexity:* The principle of open data is rather simple, especially for institutions which make reproductions of their heritage items already available online. The only thing that they would need to do in order to conform to the open data principles is to use open file formats and to apply a *free* copyright license or a public domain mark. In some cases, there may be additional challenges related to digitizing content or improving metadata quality. Also, for some heritage items there are issues related to copyright, data protection, or classified information. However, as our survey has shown, around 40% of Swiss heritage institutions have sizeable holdings that pre-date 1850, which are not concerned by these issues. It remains however to be seen to what extent this apparent simplicity of open data is confirmed in the longer term. For, as Zuiderwijk and Janssen argue, realizing the benefits of open data usually requires more from the institutions than the mere publication of the data [37]. It is also about stimulating the re-use of data by adapting the institutions' processes to the needs of the data re-users. Feeding enhanced datasets back into the institutions' own systems may further complicate things. So will the enhancement of the data in order to ensure semantic interoperability with datasets from other sources. Yet, only 29% of institutions in our sample indicated that linked data was an issue for them, 6% were planning projects in this area, but none of them had a running project. In contrast, crowdsourcing appeared to the institutions in our sample to be much more complex than open data: for them, crowdsourcing is related to many uncertainties, as they first need to learn how to set up a crowdsourcing

project and to effectively interact with a community, be it the one they build up on their own platform or an existing one, such as the Wikipedia community.

*Trialability:* Both open data and crowdsourcing practices can be set up as projects with a limited scope to gain experiences before making a definitive decision regarding their full adoption.

*Observability:* The adoption of an open data policy or the engagement in crowdsourcing practices by heritage institutions is rather easy to observe from the outside. It is however much more difficult to understand to what extent such approaches lead to benefits for the institution or third party users of the data/content. Gaining insights into where the real benefits lie usually requires direct contact with people involved in the projects.

In sum, crowdsourcing is perceived by heritage institutions as more complex than open data and isn't readily expected to lead to any sizeable advantages compared to their present situation. Furthermore, adopting crowdsourcing practices may require deeper cultural changes, although some serious reservations also need to be overcome in the case of open data. It will be interesting to see how perceptions of institutions change when they are embracing these innovations over a longer period of time, and to what extent engaging in open data and crowdsourcing practices will transform the institutions. The findings of some authors would suggest that the benefits achieved through open data may not be as low hanging fruits as perceived by the heritage institutions today, and the question remains to what extent open data and crowdsourcing practices will tend to converge in the future.

Another area where innovation diffusion theory may come into play is the selection of effective communication channels by promoters of innovations. As our survey has shown, in the case of open data and crowdsourcing we are still at a relatively early stage of the innovation-decision process. At the end of 2012, many institutions were still at the awareness stage. They therefore first needed to find out what open data and crowdsourcing are really about. As research has shown, mass communication channels are relatively more important at the awareness stage of the innovation diffusion process, while interpersonal channels are more important at the interest and evaluation stages. Also, mass communication channels are relatively more important than interpersonal channels for early adopters than for later adopters, who can more readily benefit from the information received from peers that already have firsthand experience [28]. Diffusion researchers have also come to distinguish between localite and cosmopolite communication channels. "Cosmopolite communication channels are those linking an individual with sources outside the social system under study. Interpersonal channels may be either local or cosmopolite, while mass media channels are almost entirely cosmopolite" [28]. p. 207. Earlier research has shown that cosmopolite channels are relatively more important at the awareness stage, while localite channels are relatively more important at the subsequent stage. Also, cosmopolite channels are relatively more important than localite channels for earlier adopters than for later adopters [28].

With regard to the most effective communication channels to be used to promote open data in Switzerland, we can thus conclude that given the fact that around half of the heritage institutions are still at the awareness stage, mass communication channels still play an important role. When more and more institutions start embracing open data policies, inter-personal channels to exchange experiences will gain in importance. For early adopters inter-personal channels across national boundaries or outside the heritage domain may be of particular value.

A similar situation results for crowdsourcing: Around 70% of the heritage institutions in Switzerland are still at the awareness stage. Mass communication channels are therefore even more important than for open data. Around 10% have some first experiences in the area; and another 20% have reached the interest or evaluation stage. For these, an exchange of experiences with peers would be helpful.

## 7.2 Implications for Future Research

As noted above, research into the adoption of open data and crowdsourcing is still rather scarce. Our study has shown a complementarity between qualitative and quantitative approaches. In both areas, further research is needed in order to gain a better understanding of the phenomena surrounding the adoption of these two innovations in the heritage sector.

In particular, we suggest that a similar survey be carried out on a larger scale at an international level. This survey should allow to:

- Make comparisons between museums, archives, libraries: Where do practices converge between the different types of heritage institutions? Where do they diverge?
- Investigate the factors that influence the adoption of open data policies and crowdsourcing practices; taking also into account practices in the area of web 2.0, as well as the latest insights derived from qualitative research (e.g. regarding the self-conception of heritage institutions and their role; driving and hindering factors; perceived risks; etc) and insights derived from research regarding digitization in the heritage sector.
- Further investigate the links between open data and crowdsourcing practices.

- Investigate the change of perceptions as the institutions implement open data policies or crowdsourcing approaches, e.g. by looking at institutions that are already further advanced in the adoption process.
- Make international comparisons in order to reach a better understanding of differences across countries, for example in relation to the implementation of the EU Directive on the Re-Use of Public Sector Information in the cultural heritage sector, but also with regard to financial considerations or possible differences regarding the diffusion process.
- Further corroborate findings implied by innovation diffusion theory in order to inform practice.

In parallel, we suggest that qualitative approaches be pursued that are complementary to the survey in order to reach a better understanding of the innovation adoption process, the organizational and cultural changes it may entail, and the benefits or disadvantages it may lead to. And last but not least, it might be worthwhile to compare findings related to crowdsourcing and open data for the heritage sector to those from related areas, such as open government data, open access to research data, e-participation, or the use of crowdsourcing in research, in order to get a better understanding of the similarities and differences between these fields. This would most likely encourage cross-pollination between the different strands of research.

## 8 Conclusions

The pilot survey has provided some valuable insights into the diffusion of open data and crowdsourcing among heritage institutions in Switzerland that are complementary to earlier research in the field. It could be shown where the Swiss heritage institutions stand today with regard to the innovation-decision process, and various driving forces and hindering factors could be pointed out, including a first appreciation of the expected benefits and the main beneficiaries of the innovations.

The results suggest that so far, only very few institutions have adopted an open data / open content policy. There are however signs that many institutions may adopt this practice in a near future: 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 diffusion process will be slower than for open data / open content. 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. We argued that the observed difference in the dynamics of the diffusion of these innovations is primarily due to the fact that crowdsourcing is perceived by heritage institutions as more complex than open data, that it is not readily expected to lead to any sizeable advantages, and that adopting crowdsourcing practices may require deeper cultural changes. Some caveats apply however with regard to the simplicity of open data, if the goal is to foster re-use by responding to data users' needs and preferences, to ensure semantic interoperability between datasets of different institutions, or to re-integrate enhanced datasets into the original ones.

Our data suggest that open data policies are likely to benefit first of all 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 heritage institutions consider the additional effort and costs related to open data and crowdsourcing as the greatest challenges. In contrast, potential losses of revenue play almost no role.

As a review of previous research has shown, our quantitative approach is complementary to earlier qualitative studies, and our results are mostly in line with earlier findings at the exception that only very few institutions in our sample were concerned about a potential loss of revenues when adopting open data policies. Based on the insights presented in this article we have formulated a set of recommendations with regard to further research, including the carrying out of an international benchmark survey as a natural extension of our pilot survey.

## Acknowledgements

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. Furthermore, I am much obliged 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 Neuron (Bern University of Applied Sciences), Hartwig Thomas (Verein Digitale Allmend), and David Vuillaume (Swiss Museums Association), and I am grateful to four anonymous reviewers of the OpenSym 2013 Conference and to three anonymous reviewers of the JTAER

special issue for their valuable comments. And last but not least, I would like to thank the JTAER special issue editors for their coordinating efforts.

## References

- [1] S. Alam and J. Campbell, Dynamic changes in organizational motivations to crowdsourcing for GLAMs in Proceedings Thirty Fourth International Conference on Information Systems, Milan, 2013, pp. 1-17.
- [2] N. Allen, Art Museum Images in Scholarly Publishing. Houston: Rice University Press, 2009.
- [3] S. Bakker, M. de Niet and G. J. Nauta. (2011, September) Overview of national and international initiatives. Egmus. [Online]. Available: <http://www.egmus.eu/fileadmin/ENUMERATE/deliverables/ENUMERATE-D2-01.pdf>
- [4] L. B. Baltussen, M. Brinkerink, M. Zeinstra, J. Oomen, and N. Timmermans, Open culture data: Opening GLAM data bottom-up, presented at The Annual Conference of Museums and the Web, Portland OR, USA, April 17-20, 2013.
- [5] F. Bauer and M. Kaltenböck, Linked Open Data: The Essentials: A Quick Start Guide for Decision Makers, Austria: SEMANTIC-WEB COMPANY, 2011.
- [6] G. M. Beal and J. M. Bohlen, The diffusion process, Agriculture Extension Service, Iowa State College, Iowa, Special Report No. 18, 1981.
- [7] Bundesamt für politische Bildung. (2011, October) Open Data. Bundesamt für politische Bildung. [Online]. Available: <http://www.bpb.de/gesellschaft/medien/opendata/64055/was-sind-offene-daten>
- [8] L. Carletti, G. Giannachi, D. Price, and D. McAuley, Digital humanities and crowdsourcing: An exploration, presented at The annual Conference of Museums and the Web 2013, Portland OR, USA, April 17-20, 2013.
- [9] J. R. Christensen, Four steps in the history of museum technologies and visitors' digital participation. *MedieKultur. Journal of Media and Communication Research*, vol. 27, no. 50, pp. 7-29, 2011.
- [10] K. D. Crews, Museum policies and art images: Conflicting objectives and copyright overreaching. *Fordham Intellectual Property, Media & Entertainment Law Journal*, vol. 22, no. 4, pp. 795-834, 2012.
- [11] J. M. Dooley and K. Luce, Taking our pulse: The OCLC Research survey of special collections and archives. Dublin, Ohio: OCLC, 2010.
- [12] E-Government Schweiz, Katalog priorisierter Vorhaben, Stand 24. Oktober 2012.
- [13] K. R. Eschenfelder and M. Caswell, Digital cultural collections in an age of reuse and remixes, *American Society for Information Science and Technology*, vol. 47, no. 1, pp. 1-10, 2010.
- [14] E. Estellés-Arolas and F. González-Ladrón-de-Guevara, Towards an integrated crowdsourcing definition, *Journal of Information Science*, vol. 38, no. 2, pp. 189-200, 2012.
- [15] B. Estermann, Swiss Heritage Institutions in the Internet Era. Results of a Pilot Survey on Open Data and Crowdsourcing. Bern: Bern University of Applied Sciences, E-Government Institute, 2013.
- [16] European Commission. (2001, July) European content in global networks: Coordination mechanisms for digitisation programmes. Cordis. [Online]. Available: [ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/lund\\_action\\_plan-en.pdf](ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/lund_action_plan-en.pdf)
- [17] European Commission. (2001, July) Action plan on coordination of digitisation programmes and policies. Cordis. [Online]. Available: [ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/lund\\_action\\_plan-en.pdf](ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/lund_action_plan-en.pdf)
- [18] European Commission, Information Society DG., and Salzburg Research (firm), The DigiCULT report: technological landscapes for tomorrow's cultural economy: unlocking the value of cultural heritage: executive summary, Office for official publications of the European Communities, Luxembourg, 2002.
- [19] T. Greenhalgh, G. Robert, F. Macfarlane, P. Bate, and O. Kyriakidou, Diffusion of innovations in service organizations: systematic review and recommendations, *Milbank Quarterly*, vol. 82, no. 4, pp. 581-629, 2004.
- [20] R. Holley, Crowdsourcing: how and why should libraries do it?, *D-Lib Magazine*, vol. 16, no. 3/4, 2010.
- [21] J. Howe. (2006, June) Crowdsourcing: A definition. Crowdsourcing blog. [Online]. Available: [http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing\\_a.html](http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html)
- [22] J. Jankowski, Y. Cobos, M. Hausenblas, and S. Decker, Accessing cultural heritage using the web of data, in Proceedings The 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST, Malta, 2009.
- [23] K. Kelly, Images of Works of Art in Museum Collections: The Experience of Open Access A Study of 11 Museums. Washington DC: Council on Library and Information Resources, 2013.
- [24] Z. Manžuch, Monitoring digitisation: lessons from previous experiences, *Journal of Documentation*, vol. 65, no. 5, pp. 768-796, 2009.
- [25] Nicholls, M. Pereira and M. Sani, The virtual museum, LEM The Learning Museum, Bologna, Technical Report 1, 2012.
- [26] J. Oomen and L. Aroyo, Crowdsourcing in the cultural heritage domain: Opportunities and challenges, in Proceedings 5th International Conference on Communities & Technologies – C&T 2011, Brisbane, Australia, 2011, pp. 138-149.
- [27] L. B. Phillips, The temple and the bazaar: Wikipedia as a platform for open authority in museums, *Curator: The Museum Journal*, vol. 56, no. 2, pp. 219-235, 2013.
- [28] E. M. Rogers, Diffusion of Innovations. New York: Free Press, 2003.
- [29] Schweizerische Eidgenossenschaft, Strategie des Bundesrates für eine Informationsgesellschaft in der Schweiz. Eidgenössisches Departement für Umwelt, Energie, Verkehr und Kommunikation UVEK, März 2012.
- [30] K. Smith-Yoshimura and C. Schein, Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews. Dublin, Ohio: OCLC Research, 2011.

- [31] N. Stroeker and R. Vogels. (2012, May) Survey report on digitisation in European cultural heritage institutions 2012. ENUMERATE. [Online]. Available: <http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2012.pdf>
- [32] N. Stroeker, R. Vogels, G. J. Nauta, and M. de Niet. (2013, September) Report on the enumerate thematic surveys on digital collections in European cultural heritage institutions. ENUMERATE. [Online]. Available: <http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Thematic-Survey-2013.pdf>
- [33] Sunlight Foundation. (2010, August) Ten principles for opening up government information. Sunlight Foundation. [Online]. Available: <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>
- [34] M. Terras, Digital curiosities: resource creation via amateur digitization, *Literary and Linguistic Computing*, vol. 25, no. 4, pp. 425-438, 2010.
- [35] L. Wyatt, *Wikipedia & Museums: Community Curation*. *Uncommon Culture*, no. 2, vol. 1, pp. 33-41, 2011.
- [36] A. Zuiderwijk, N. Helbig, J. R. Gil-García, and M. Janssen, Guest editors' introduction. Innovation through open data: A review of the state-of-the-art and an emerging research agenda. *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, no. 2, pp. I-XIII, 2014.
- [37] A. Zuiderwijk, and M. Janssen, A coordination theory perspective to improve the use of open data in policy-making, *Electronic Government*, no. 8074, pp. 38-49, 2013.