Diffusion of Open Data and Crowdsourcing among Heritage Institutions

Based on data from Finland, Poland, Switzerland, and The Netherlands

Beat Estermann

Bern University of Applied Sciences, Switzerland; beat.estermann@bfh.ch

Abstract

In a survey among heritage institutions in Finland, Poland, Switzerland, and The Netherlands we examined the diffusion of various Internet-related practices within the heritage sector. The practices examined comprise the exchange of data with other institutions, the digitization of heritage objects, open data/open content, the use of social media, as well as collaborative content creation. We draw on the innovation diffusion model popularized by Everett Rogers in order to assess the diffusion stage of each of the examined practices: The results suggest that digitization and the use of social media are already widespread practices within the European heritage sector, with adoption rates above 50%. In contrast, open data and open content are just about to diffuse; with adoption rates of 26% and 17% respectively. While open data and open content are expected to diffuse relatively rapidly as these practices are considered as important by almost half of the respondents, collaborative content creation is seen as less important and appears to be diffusing at a slower pace.

By means of regression analyses, we explored the influence of various characteristics of heritage institutions on their adoption of open data and crowdsourcing. It appears that the "country" factor has an influence on the adoption of almost all the Internet-related practices under examination. Several other factors, such as institution type, types of heritage objects held by an institution, geographical reach, size of an institution, staff skills, or attitudes towards making content available for others to re-use were found to be correlated with the adoption of only one or two of the examined practices.

Our data also shows that over the next five years we will see leaps in digitization activities, which opens up many new opportunities for opening up collections. While problems related to copyright clearance may be a serious issue for individual institutions, the data suggest that copyright issues are not a general show-stopper in the short run, but may become a serious limitation regarding the opening up of content within a few years from now. The greatest challenges regarding open content are the time effort and expense related to the digitization and the documentation of content as well as the negative attitudes among many heritage institutions regarding the opening up of content, inspired by a feeling of loss of control and the wish to prevent commercial use of content by third parties without due compensation.

Keywords: heritage institutions, open data, open content, crowdsourcing, diffusion of innovations

2 Introduction

In the context of widespread adoption of Open Government Data policies, this article sheds light on a sector that is adjacent to classical public administration: the cultural heritage sector. While some heritage institutions (galleries, libraries, archives, museums) are governed by public law, many others are constituted as private non-profit organizations, a large fraction of which are mainly publicly funded and thus directly affected by public funding policies. The emerging collaborative culture on the Internet provides heritage institutions with new opportunities, but creates also new challenges for them. Several pioneer institutions have already managed to harvest the first fruits of the new collaborative culture, sharing their data and content online, embracing innovative crowdsourcing approaches, and reaching out to online communities. Many heritage institutions, however, are still in an exploratory phase, trying to find out how they could best adapt to the new circumstances.

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; European Commission 2001a and 2001b). 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 (European Commission and Salzburg Research 2002, Manžuch 2009). Thus, digitization has turned out to be a powerful means to expand access to collections for wider audiences (Manžuch 2009, Oomen & Aroyo 2011). 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 (Christensen 2010, Oomen & Aroyo 2011, Phillips 2013). 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. The increasing trend towards open data is presently being institutionalized in the European heritage sector in the wake of the 2013 extension of the European Union's Public Sector Information Directive (2003/98/EC) to include also libraries, museums, and archives. By linking open data from various publishers, data can be integrated based on commonly shared ontologies – an approach that is commonly referred to as "linked data", giving rise to the so-called "semantic web" (Jankowski et al. 2009).

While the advancement of digitization efforts among heritage institutions in Europe is being monitored both at a national and international level (see Stroeker et al. 2012 or Bakker et al. 2012), 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 in 2012 (Estermann 2013 and 2014). 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 various stakeholders about the developments in this area and to get an overview of the main challenges and driving forces regarding the diffusion of these practices. Based on an improved questionnaire, an international benchmark survey is being carried out in 2014/2015 in a number of countries across the globe.

In this article we first introduce a series of core concepts and present the methodological approach. We then present findings from the international benchmark survey based on data from Finland, Poland, Switzerland, and The Netherlands, relating them to earlier research and discussing them in the light of innovation diffusion theory as set out in an earlier publication (Estermann 2014). We conclude the article with a series of suggestions in view of further research.

3 Definition of core concepts

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

3.1 Heritage institutions

In the context of our survey we follow the definition of the ENUMERATE project, which has defined the cultural heritage domain to consist of the "memory institutions", such as museums, libraries, archives and records offices, audio-visual and film archives, organizations with curatorial care for monuments, sites and the historic environment, as well as hybrid types of organizations. The defining criterion of a "heritage institution" is that "curatorial care for, at least part of, the collections of the institution are included in its mission. Institutions that do not hold heritage collections or that have collections of heritage materials (like for example of books, films, and music) to be lent by or sold to contemporary users without the explicit task of safeguarding the collections for future generations, will not be included in the survey. This essentially leaves out both school libraries [...] and public libraries without cultural heritage collections" (Nauta et al. 2011, p. 5).

3.2 Open data / open content

The open data movement experienced its worldwide breakthrough a bit over 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. In its general form, 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 (Bundesamt für politische Bildung 2011). In the heritage sector, a distinction is however frequently made between "data" and "content": while the term "data" is generally used to designate different types of metadata, such as catalogues, inventories, finding aids, glossaries, vocabularies, or name authority files, the term "content" is used to refer to digital versions of heritage objects.

According to the Sunlight Foundation's ten *Open Data Principles* (Sunlight Foundation 2010) 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.

3.3 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 1990ies, 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 (Bauer & Kaltenböck 2011, Jankowski et al. 2009).

3.4 Crowdsourcing / collaborative content creation

The term 'crowdsourcing' was coined by Jeff Howe in 2006 in Wired Magazine, by 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"* (Howe 2006). 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 one: *"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"* (Estellés-Arolas & González-Ladrón-de-Guevara 2012, p. 9). In our survey, the term "collaborative content creation" is used alongside "crowdsourcing" to refer to activities taking place within existing online communities, such as the Wikipedia community.

3.5   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"* (Rogers 2003, 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"* (Rogers 2003 p. 20). In the present paper, we use the 'innovation diffusion' approach as a theoretical lens to study where heritage institutions stand with regard to the adoption of various innovative practices. As Rogers (2003) 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 present paper we mainly draw on the following elements of innovation diffusion theory:

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 (1957) 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 (Greenhalgh et al. 2004). In practice, a differentiation of decision stages can be useful to choose the appropriate communication channel to promote an innovative practice. As Rogers (2003) 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 (2003) 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, 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 (Greenhalgh et al. 2004). 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 (Rogers 2003).

4 Research questions and methodology

4.1 Research questions

The primary motivation for our research was to create an instrument that allows measuring the level of adoption of open data policies and crowdsourcing practices among heritage institutions in various countries 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 extension of the Swiss pilot survey, the international survey should allow for international comparisons in order to reach a better understanding of the progress and the particularities of each country. Furthermore, the larger sample size should also allow for comparisons between museums, archives, and libraries, as well as for other more sophisticated data analyses.

The research questions covered by this article can be summarized as follows:

- Where are heritage institutions in the participating countries 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? How do the results of the survey relate to findings from previous research?

- What are the expected benefits of open data and crowdsourcing in the heritage domain?

- What are the differences between different types of heritage institutions? Where do practices converge between the different types? Where do they diverge?

4.2 Survey instrument

The questionnaire was elaborated in an iterative process: an initial version was produced based on the questionnaire used for the Swiss pilot survey (Estermann 2013) and the ENUMERATE Core Survey 2 (Stroeker & Vogels 2014) and complemented by new questions based on a thorough review of the previous research regarding open data, crowdsourcing, and social media in the heritage sector (Estermann 2014). This initial version was then reviewed and discussed by a number of experts from various countries in an open feedback process that led to a revised version. The revised version was in turn pretested among a small number of institutions in The Netherlands and Denmark; the ensuing changes led to the final version. Minor adaptions were made to the final version of the questionnaire after its deployment in the first two countries, Poland and Finland.

In its final version, the questionnaire contains 34 questions: Ten questions relate to the institution's characteristics, such as the type of institution, the most characteristic type of heritage items, its main users, its geographical reach, the number of employees and volunteers, its annual revenues, the composition of revenue sources, and the institution's legal form. Two questions concern the assessment of various practices related to the Internet, regarding their importance for the institution and the institution's evaluation of risks and opportunities. Two questions address the availability of metadata in form of open data and linked data respectively, while two further questions focus on the digitization of heritage objects. Seven questions cover various aspects of open content: conditions under which the institution is ready to make its content available on the Internet without receiving payment in exchange, the copyright situation of the objects in the institution's collection, the percentage of objects published as open content, copyright licenses used to make content available as open content, as well as benefits, challenges, and risks related to open content. Two questions relate to the use of social media, while four questions cover various aspects of crowdsourcing, such as staff involvement in collaborative content creation by online communities, crowdsourcing approaches used, as well as the purpose, risks and challenges of crowdsourcing. Four questions address the skills and know-how of the staff in the areas covered by the survey as well as the institution's need for further information, training, and external consulting. And finally, the last question asks the survey participants to list the professional role(s) of the people who have responded to the questionnaire. Several questions are conditional questions, and some haven't been included in all the countries. The questionnaire has been translated into the local languages by national teams, whereby a strict quality assurance procedure has been followed. The questionnaire in its various language versions is available for download on the project portal[1].

---

[1] http://survey.openglam.ch

4.3 Sampling approach and response rates

The survey has been organized in a federative manner, relying on national teams in the different participating countries. In the four countries covered by this article, the national teams were asked to include – to the extent possible – all the known heritage institutions of the given country in the distribution list. In all four countries, a large majority of the institutions could be reached by e-mail. In Finland, individual museums are often grouped together into one organization with a central contact address. In this case, only the central contact address has been used.

Judging by the number of institutions included in the distribution lists, it appears that there are stark differences between the countries regarding the structure of the heritage domain, even if we account for methodological differences in the way the distribution lists have been assembled (see table 1): When looking at the number of institutions per million inhabitants, Switzerland shows by far the highest density of heritage institutions, outnumbering the other countries by factor 2 to 10. When it comes to the number of institutions per square kilometer, Switzerland and The Netherlands have the highest density of heritage institutions, with the other two countries showing values that are 15 to 30 times smaller. Given the varying percentages of small and very small institutions in the different samples (see section 5 below), it can be assumed that the long tail of smaller institutions has been removed in Finland by grouping smaller institutions together into one bigger organization. In the case of Poland, the long tail of smaller museums, provided that they do exist, may be missing in the distribution list.

	CH	FI	NL	PL
N institutions contacted	1532	356	1393	669
N inhabitants (in mio., as of 2013)	8.1	5.4	16.8	38.5
Surface (in 1000 sq km)	41	338	42	313
Density of heritage institutions (institutions per mio. inhabitants)	189.1	65.9	82.9	17.4
Density of heritage institutions (institutions per 1000 sq km)	37.4	1.1	33.2	2.1

Table 1: Density of heritage institutions in the participating countries

The institutions were sent an e-mail invitation and two reminders in ca. 2 weeks' intervals. In the case of Poland, additional follow-up calls were made to non-responding institutions. Table 2 shows the response rates per country and institution type: While the overall response rate for the different countries lies between 12% and 26%, there are also quite pronounced differences in the response rates for the different institution types, with museums generally less inclined to fill in the questionnaire than archives or libraries.

	CH (incl. FL)	FI	NL	PL
N institutions contacted [1]	1543 (11)	356	1393	669
N responses [2]	278 (2)	81	146	79
Response rate [3]	19%	26%	12%	12%
– Archives	27%	50%	16%	21%
– Libraries	29%	37%	15%	17%
– Museums	16%	21%	10%	10%

Notes:
[1] Number of institutions which have been contacted (each institution is counted once, even if it has been contacted via several email addresses or through different channels, e.g. follow-up phone calls in addition to e-mail reminders).
[2] Number of institutions which have completed the questionnaire (almost) in full, i.e. institutions which have responded to at least 20 out of the 24 non-conditional questions of the questionnaire.
[3] For the purpose of the calculation of the response rate, institutions which have started to fill in the questionnaire but dropped out after question A2 as they did not correspond to the survey's definition of heritage institutions, are counted as responses.

Table 2: Response rates per country and institution type

4.4 Sample biases

There is an initial sample bias due to the fact that institutions without a publicly available email address have not been contacted. For the countries in question, the percentage of institutions thus excluded from the survey ranges between less than 5% (e.g. Switzerland) to almost 15% (Poland). We did not make any extra efforts to reach these institutions, as it could be assumed that the survey would not make much sense thematically to most of them.

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

- Archives (22% of contacted institutions) and libraries (23%) were more likely to respond than museums (14%). These numbers were calculated on the basis of our own categorization based on the institutions' name and e-mail address and/or their membership in professional associations.

- In order to further estimate the self-selection bias, we analyzed the respondents who had at least partly filled in the questionnaire. Among these, the institutions with the following characteristics have been more likely to fill in the questionnaire (almost) in full than the others: institutions for which text based or time based resources are characteristic; institutions which count public authorities, research institutions, or journalists among their main users; institutions with a local / regional focus (compared to institutions with a global focus), and institutions for which institutional funding from public funds accounts for at least 80% of total revenues (compared to institutions for which institutional funding from public funds accounts for less than 20% of total revenues). Across all countries, no distortion has been observed regarding the size of institutions or the institutions' form of organization.

- More critically, a certain number of biases have been identified which do not pertain to objective characteristics of the institutions, but to some of the attitudes that are at the center of our investigation. Thus, institutions considering digitization or collaborative content creation as important have been more likely to complete the questionnaire; and the same is the case for institutions holding positive attitudes regarding the risks and opportunities of various Internet-related practices. To name just the most extreme one: Institutions indicating that the opportunities of exchanging data prevail over the risks had an 87% likelihood to complete the questionnaire, compared to 65% of their counterparts which do not think so. Interestingly, such topic-related dropout rates are unevenly distributed across the countries. While they hardly occur in Finland and in Switzerland, they can be very pronounced in Poland and The Netherlands. Thus, 44% of Polish institutions indicating that digitization is not important for them did not complete the survey, while only 10% of their counterparts indicating that digitization is important dropped out. Or in the case of the Netherlands, 70% of institutions considering that the opportunities of exchanging data do not prevail over the risks dropped out (compared to 16% of their counterparts). These strong self-selection biases in some countries pose serious methodological challenges and should be kept in mind when interpreting the results of the survey.

4.5 Limitations

The main limitations of the survey are related to the methodological challenges posed by the heterogeneity of the heritage sectors in the participating countries and the stark differences regarding the responding behavior of institutions in the various countries. In addition, there may have been differences in the way the national inventories of heritage institutions have been assembled (e.g. stronger focus on public sector institutions in some countries than in others). The findings presented in this article are based on a combined sample of 584 institutions from four European countries; while they are certainly more robust than the findings of the Swiss pilot survey with only 72 respondents (Estermann 2014), we are planning to add further countries, including countries from other continents, in order to improve the robustness of the results.

5   Description of the sample

A large majority of the responding institutions are either public institutions (54%) or private non-profits (33%). Only 2% are or belong to private, profit-oriented institutions. 11% of respondents indicated that their institution has a mixed form (e.g. premises provided by a public institution; exploitation taken care of by a private nonprofit). There are important differences between the countries: While the Polish sample contains 94% public institutions, this percentage is much smaller in the other three countries (NL: 43%; CH: 46%; FI: 63%). 64% of responding institutions are predominantly funded by public funds with a share of institutional funding from public funds amounting to at least 50% of overall revenues (NL: 44%; CH: 61%; FI: 78%; PL: 87%). Individual funding situations are, however, quite heterogeneous: in 6% of the cases the majority of revenues comes from institutional funding from private funds; in 9% of the cases from sponsorship or donations; in 11% of the cases from commercial operations; and in 2% of the cases from project funding.

With regard to the size, the sample contains a good mix of institutions: 47% of responding institutions are small organizations with a total annual budget of 100'000 € or less, while 21% report an annual budget of at least 1 million €. Similarly, 59% have no more than 5 FTE paid staff, while 15% report at least 25 FTE staff. As noted above, Switzerland and The Netherlands seem to have a higher density of heritage institutions than the other two countries. This is reflected by a large percentage of very small and small institutions (more than 25% of institutions in Switzerland and The Netherlands report annual revenues of 10'000€ or less, compared to 12% in Poland and 3% in Finland; and at least 50% of institutions in Switzerland and The Netherlands report no more than 1 FTE paid staff, compared to 6% in Finland, and 3% in Poland).

Asked about their main users, the surveyed institutions most frequently mentioned private individuals (95%), education (79%), and research (51%). Journalists (36%), cultural institutions (32%), public authorities (28%), and private enterprises (17%) were mentioned by less than half of the institutions. Interestingly, journalists are mentioned clearly more frequently in Finland (56%), while research is mentioned less frequently in the Netherlands (36%). Regarding their geographical reach, 41% of institutions reported that they had a "local/regional" focus, compared to 28% with a "regional/national", 25% with a "national/international", and 6% with a "global" focus. In line with their larger percentage of small institutions, Switzerland and The Netherlands have higher percentages of institutions with a "local/regional" focus than the other two countries (CH: 48%; NL: 45%; FI: 31%; PL: 19%). While Finland has the largest percentage of institutions with a "regional/national" focus (42%), Poland has the largest percentage of institutions with a "national/international" focus (42%). This is another indicator that the long tail of small, locally focused institutions is missing in the Polish sample.

As to the heritage objects that are characteristic for their institutions, more than half of the respondents mentioned "text based resources" (70%), "two-dimensional visual resources" (70%), "three-dimensional man-made movable objects" (59%), or "archival resources" (57%). "Time based resources" were mentioned by 43%, "digital interactive resources" by 27%, "geography based resources" by 25%, and "natural resources" by 9% of responding institutions.

6 Main findings

6.1 Diffusion of Internet-related practices among heritage institutions

The questionnaire of our survey has been designed in a way to allow for the responding institutions to be assigned to the different stages of the innovation-decision process as suggested by innovation diffusion theory. Thereby, the following criteria have been taken into account: By default, institutions are assigned to the "no interest" stage. Institutions which indicated that they require further information, training, or external consulting in a given area are assigned at least to the "interest" stage. Institutions which anticipate a minimal level of activity in a given area (e.g. at least 0.5% of content released as open content over the coming 5 years or at least one social media type used over the coming year) are assigned at least to the "evaluation" stage. Institutions which already report this minimal level of activity today are assigned at least to the "trial stage". Institutions which report already quite a high level of activity in a given area are assigned either to the "adoption" or the "advanced implementation" stage (see table 3 for the criteria that have been applied). In addition, institutions which show decreasing (in the case of social media and collaborative content creation) or stagnating (in the case of open data, linked data, digitization, and open content) levels of activity, are assigned to the "stagnation / discontinuance" stage in order to capture those which are planning to abandon or to significantly reduce a given practice. As this latter group is very small (ranging from 0% to 2.6% depending on the practice), it has been excluded from further analyses.

	"Adoption"	"Advanced implementation"
Open data	• More than 10% of metadata (average of all metadata types) are available as open data at present, and within the next 5 years, over 10 additional percent will be made available as open data.	• More than 50% of metadata (average of all metadata types) and more than 50% of the institution's catalogues, inventories, and finding aids are presently available as open data.
Linked data	*In analogy to open data*	*In analogy to open data*
Digitization	• More than 10% of content have already been digitized, and within the next 5 years, over 5 additional percent will be digitized. OR • More than 5% of content have already been digitized, and within the next 5 years, over 10 additional percent will be digitized.	• More than 50% of content have already been digitized.
Open content	*In analogy to digitization*	*In analogy to digitization*
Social media	• At least one type of social media is being used at present, and within the next year, at least one more will be used. OR • At least two types of social media are being used at present, and within the next year the number of social media types being used remains stable or increases.	• More than 3 (out of 10) different types of social media are being used at present.
Collaborative content creation	*In analogy to social media*	• More than 2 (out of 5) different types of collaborative content creation are being used at present.

Table 3: Criteria used for the "adoption" and the "advanced implementation" stage

Figure 1 gives an overview of the diffusion of the various practices among heritage institutions in the four countries, according to the innovation diffusion model: Digitization and the use of social media are the most widespread practices with adoption rates of over 50%. They are followed by open data (26% adoption) and open content (17% adoption). In both cases, the "early majority" is starting to adopt the practice. In contrast, collaborative content creation (13% adoption) and linked data (6% adoption) are for the moment just being embraced by the "early adopters", although collaborative content creation may soon reach the "early majority".

As can be seen in figure 2 and 3, adoption rates vary according to country and institution type (the following differences are significant at the 0.95 level): While digitization is widespread practice in the Dutch and Finnish heritage sector, Switzerland and Poland are lagging considerably behind. There are also important differences between institution types, with the adoption rate of museums (59%) being more than twice as high as the one of libraries (23%). The use of social media is a widespread practice in the Finnish, Polish, and Dutch heritage sectors, with adoption rates of 70% and more, while Switzerland is lagging behind with an adoption rate of only 34%. Differences between institution types are significant, but not as pronounced as with regard to digitization and in different order (libraries and archives showing higher adoption rates than museums).
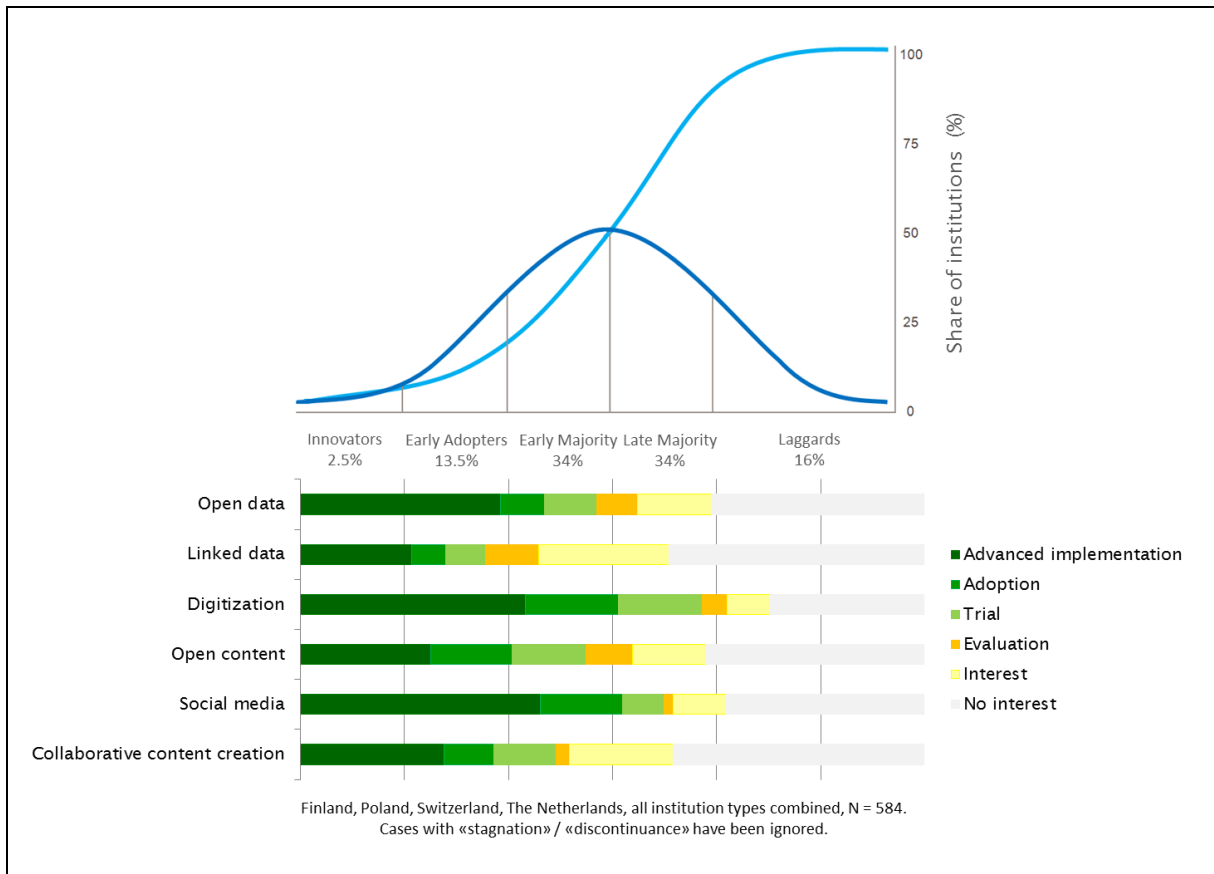
Figure 1: Diffusion of innovative practices among heritage institutions
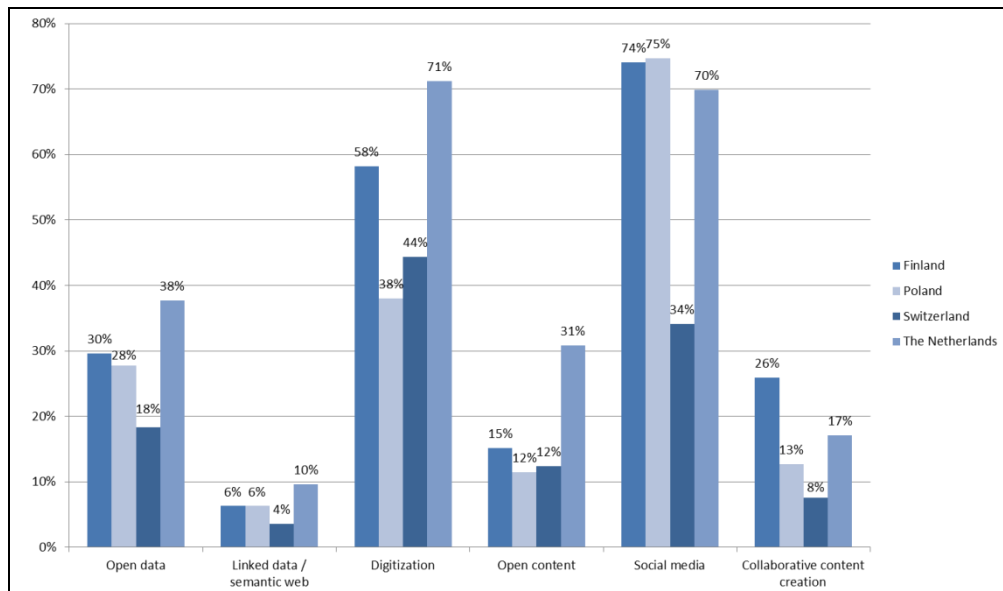

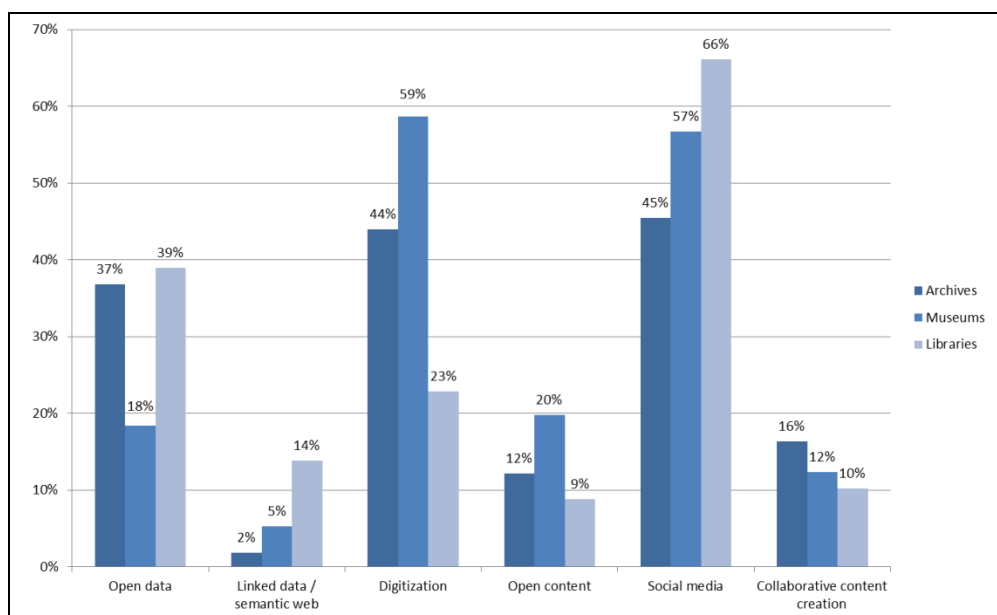Figure 2: Adoption rates per country

Figure 3: Adoption rates per institution type

Open data and open content are about to diffuse within the heritage sector. When it comes to putting open data into practice, the Dutch and the Finnish institutions are ahead of the Swiss institutions (NL: 38%; FI: 30%; PL: 29%; CH: 18%). Regarding the opening up of content, the Dutch institutions are clearly ahead of the institutions in the other three countries (NL: 31%; FI: 15%; CH: 12%; PL: 12%). Museums are lagging behind the other two institution types regarding the adoption of open data, but are ahead when it comes to adopting an open content approach.

Regarding collaborative content creation, Finnish institutions are clearly in the lead (26% adoption, compared to NL: 17%; PL: 13%; CH: 8%), and regarding the adoption of linked data, the libraries are ahead of the other two institution types (14% adoption, compared to museums: 5%; archives: 2%).

6.2   The institutions' characteristics and the level of adoption of the various practices

In order to gather further insights into the factors that influence the adoption of the Internet-related practices under examination, we generated a series of crosstabs and ran z-tests to check whether the differences in the proportions of adopting institutions are significant at the 0.95 level. Here a summary of (selected) findings:

-   The types of objects which are characteristic for an institution affect its likelihood to adopt various practices, especially open data, digitization, and the use of social media (for example, 30% of institutions for which text based resources are characteristic adopt open data, compared to only 17% of their counterparts). And most notably, institutions for which digital interactive resources are characteristic have higher odds to adopt a whole series of Internet-related practices: open data, linked data, digitization, open content, and social media use.
-   The types of users which an institution counts among its main users affect its likelihood to adopt various practices, such as open data, linked data, social media use, and collaborative content creation. Thus, institutions targeting (other) cultural institutions have a 12% likelihood to adopt linked data (compared to 3% for those which don't). Or institutions which count journalists among their main users have a higher likelihood to adopt open data than their counterparts (33% vs. 22%), to name just a couple of such correlations.
-   Not very surprisingly, institutions with a local/regional focus are less likely to use social media than other institutions (local-regional: 39%; regional-national: 60%; national-international or global: 68%). Regarding the adoption of the other practices, the geographical reach of an institution doesn't seem to play a role.
-   The institutions' size (in terms of number of paid staff and of total revenues) is positively correlated with the adoption of open data, linked data, social media use, and collaborative content creation. In contrast, the institutions' size is not related to their adoption of digitization and open content.
-   Institutions whose workforce counts only volunteers are less likely to adopt open data than other institutions. Interestingly, institutions whose workforce is made up of 5 to 99% of volunteers are more likely to adopt digitization, than those with a smaller percentage of volunteers. Similarly, institutions with at least 5% volunteers in their workforce are more likely to adopt open content than institutions with fewer volunteers. Regarding the use of social media, there seems to be a non-linear relationship between the involvement of volunteers and adoption rates: the latter are highest among the institutions whose workforce is made up by 5 to 99% of volunteers; both the institutions with fewer volunteers and the institutions exclusively run by

11

volunteers have lower adoption rates. A similar pattern can be observed for collaborative content creation, although not all the differences are significant.

- The composition of revenue sources mainly plays a role regarding the adoption of digitization and open content. Somewhat surprisingly, institutions where institutional funding makes up for 80% or more of the total annual revenues are less likely to adopt digitization and open content than other institutions. In contrast, higher levels of public funding seem to be positively related to the adoption of open data. At the same time, the adoption of open content is positively related to high levels of sponsorship / donations within the institutions' revenue mix as well as with high levels of revenues from commercial operations. Regarding digitization, it seems that a good balance of institutional funding from public funds and revenues from commercial operations increases the likelihood of adoption.
- The organizational form of an institution (public institution / private nonprofit / commercial organization) doesn't play a role regarding the adoption of the various practices, except for the fact that institutions with a mixed form (e.g. premises provided by a public institution; exploitation taken care of by a private nonprofit) are less likely to adopt social media use than public institutions (41% compared to 57%). This may however be due to the fact that this organizational form is highly correlated with "local / regional" focus, which in turn is associated with low levels of social media use.
- The fact that a heritage institution has metadata (an astonishing 29% say that they don't have catalogues, inventories, or finding aids) is positively associated with their adoption of open data, linked data, digitization, social media use, and collaborative content creation.
- Another good predictor for the adoption of all the six examined Internet-related practices is the fact that an institution's staff is involved in Wikipedia. The cause-effect-relationship is however unclear: for example, are they involved in Wikipedia because it's part of their social media strategy? Or are they embracing social media, because they are involved in Wikipedia?
- If we look at the two practices that are the least developed within the heritage sector, linked data and collaborative content creation, it is the staff's involvement in Wikidata and Wikimedia Commons which serve as the strongest predictors. Thus, institutions whose staff is involved in Wikidata have a 22% likelihood to have adopted linked data (compared to 5% for their counterparts); and institutions whose staff is involved in Wikimedia Commons have a 35% likelihood to have adopted collaborative content creation (compared to 11% for their counterparts).
- We also looked into the relationship between staff skills and the adoption of the various practices. Not very surprisingly, institutions which generally report insufficient staff skills also show lower adoption rates for most of the practices. However, the relationship between the perceived level of staff skills (compared to the institution's goals and requirements) and the adoption of innovative practices is not a linear one (at the exception of linked data). In fact, when it comes to digitization and the adoption of open content, the institutions reporting "rather sufficient" overall staff skills show higher adoption rates than those reporting "sufficient" staff skills; and in the case of digitization, social media use, and collaborative content creation, the adoption rates of the two groups are very similar. It could well be that the heritage institutions' ambitions in the sphere of Internet-related practices are not only related to high adoption rates, but may also ensue a certain level of dissatisfaction regarding present levels of staff skills.
- Institutions whose staff members are using a larger number of different approaches to acquire skills and know-how tend to have higher adoption rates for the various practices.
- It shouldn't come to a surprise that the institutions which consider that their ways of acquiring skills and know-how are "rather effective" have a higher adoption rate of open data than institutions which think that they are "ineffective". But strangely enough, there seems to be no correlation between the effectiveness of skills acquisition and the adoption of digitization, open content, social media use, and collaborative content creation.
- And finally, we also compared the institutions' adoption rates with their attitudes towards making content available on the Internet without receiving payment in exchange. Quite unsurprisingly, their attitudes can serve as good predictors for their adoption rates regarding all the practices under examination, except for digitization. Thus, institutions which would open up their content for non-profit projects permitting commercial use of the content are more likely to adopt open data (36% vs. 22%), linked data (11% vs. 3%), open content (28% vs. 14%), social media use (71% vs. 49%), and collaborative content creation (19% vs. 11%).
- Institutions which would make content available only under the condition that the works be used without modification are less likely to adopt open data and collaborative content creation. However, they show the same likelihood as their counterparts regarding the adoption of open content. This comes to a surprise, as this limitation isn't compatible with an open content approach. As with the question whether they would be ready to make content freely available for commercial users, declared attitudes seem to lag behind declared practices in some of the cases (in fact, 54% of those institutions which report that they are making a significant amount of content available as "open content" say that they wouldn't make content available for commercial users, which seems contradictory).

This list of findings gives an idea of the factors influencing the adoption of the various practices; but it doesn't yield any insights as to the interplay between the various factors. In order to single out the most relevant factors while controlling for all the others, we ran ordinal regressions with the adoption stages as dependent variables and all the potential factors discussed in the preceding section as independent variables, at the exception of the variables related to Wikipedia, Wikimedia Commons, or Wikidata involvement of staff members. The revenues variable and the

variable regarding the share of volunteers in the workforce, which had shown non-linear relationships with some of the dependent variables, were entered as categorical variables. When running the regression for the adoption of open content, we had to drop the annual revenues variable altogether in order to meet the test assumptions (test of parallel lines). As dependent variables we used categorical variables with three categories for the adoption of open data, digitization, open content, and social media use, and categorical variables with six categories for the adoption of linked data and collaborative content creation, as this fitted best with the model assumptions.

A cumulative odds ordinal logistic regression with proportional odds was run to determine the effect of the independent variables on the level of adoption of the various practices. There were proportional odds, as assessed by full likelihood ratio tests comparing each of the fitted models to a model with varying location parameters[2]. As most cells were sparse with zero frequencies in 66.7%, respectively 83.3% of cells, the goodness-of-fit tests were not expected to yield any reliable results. However, the final models statistically significantly predicted the dependent variables over and above the intercept-only models, with very good results for the adoption of open data, linked data, digitization, open content, and social media use (with p-values of .001 or below), and a satisfactory result for collaborative content creation (with a p-value of .047)[3].

Regarding the level of adoption of open data (3 categories), the following predictors were found:
- An increase in the number of metadata types[4] present in the institution (scale: 0-3) was associated with an increase in the odds of having a higher level of open data adoption, with an odds ratio of 2.013[5].
- The odds of Dutch heritage institutions having a higher level of open data adoption were 4.578 times that for their Swiss counterparts[6]; the other differences between countries were not significant.
- An increase of the score for the item saying that they would make content only available if the works can only be used without modification (5-point Likert scale) was associated with a decrease in the odds of having a higher level of open data adoption, with an odds ratio of .773[7].
- An increase in the number of ways used by the institution's staff to acquire new skills and know-how (scale: 0-9) was associated with an increase in the odds of having a higher level of open data adoption, with an odds ratio of 1.196[8].

Regarding the level of adoption of linked data (6 categories), the following predictors were found:
- The odds of Dutch heritage institutions having a higher level of linked data adoption were 4.403 times that for their Swiss counterparts[9], and 3.199 times that for their Polish counterparts[10]. The odds of Finnish heritage institutions having a higher level of linked data adoption were 2.275 times that for their Swiss counterparts[11]. The other differences between countries were not significant.
- An increase in the number of metadata types present in the institution (scale: 0-3) was associated with an increase in the odds of having a higher level of linked data adoption, with an odds ratio of 1.671[12].
- An increase in the number of ways used by the institution's staff to acquire new skills and know-how (scale: 0-9) was associated with an increase in the odds of having a higher level of linked data adoption, with an odds ratio of 1.182[13].
- An increase of the score for the item saying that they would make content available for non-profit projects which permit their users commercial use of the content (5-point Likert scale) was associated with an increase in the odds of having a higher level of linked data adoption, with an odds ratio of 1.371[14].
- The odds for heritage institutions considering the overall skills level of their staff as sufficient to have a higher level of linked data adoption were 1.943 times higher than for those which considered that the overall skills level of their staff was insufficient[15].

---

[2] The test statistics for the full likelihood ratio tests were for the adoption of open data: χ^2 = 42.066, p = .638; for linked data: χ^2 = 187.438, p = .416; for digitization: χ^2 = 49.262, p = .344; for open content: χ^2 = 44.773, p = .397; for social media use: χ^2 = 32.213, p = .938; and for collaborative content creation: χ^2 = 157.077, p = .926.

[3] The model fitting information was for the adoption of open data: $\chi^2(46)$ = 115.409, p < .001; for linked data: $\chi^2(46)$ = 91.742, p < .001; for digitization: $\chi^2(46)$ = 80.589, p = .001; for open content: $\chi^2(43)$ = 82.131, p < .001; for social media use: $\chi^2(46)$ = 117.342, p < .001; and for collaborative content creation: $\chi^2(46)$ = 63.141, p = .047.

[4] The survey differentiated between three metadata types: (1) catalogues, inventories, finding aids; (2) glossaries, vocabularies, ontologies; (3) name authority files.

[5] 95% CI: 1.526 to 2.655, $\chi^2(1)$ = 24.565, p < .001

[6] 95% CI : 1.830 to 11.448, $\chi^2(1)$ = 10.580, p = .001

[7] 95% CI: .607 to .983, $\chi^2(1)$ = 4.418, p = .036

[8] 95% CI: 1.061 to 1.348, $\chi^2(1)$ = 8.614, p = .003

[9] 95% CI: 1.886 to 10.275, $\chi^2(1)$ = 11.749, p = .001

[10] 95% CI: 1.097 to 9.325, $\chi^2(1)$ = 4.537, p = .033

[11] 95% CI: 1.019 to 5.080, $\chi^2(1)$ = 4.026, p = .045

[12] 95% CI: 1.290 to 2.163, $\chi^2(1)$ = 15.143, p < .001

[13] 95% CI: 1.053 to 1.327, $\chi^2(1)$ = 8.031, p = .005

[14] 95% CI: 1.058 to 1.776), $\chi^2(1)$ = 5.703, p = .017

Regarding the level of adoption of digitization (3 categories), the following predictors were found:
- The odds of museums having a higher level of adoption of digitization were 13.836 times that for libraries[16], and 4.943 times that for archives[17]. Similarly, the odds for institutions representing a combination of archive, museum, or library to have a higher level of adoption of digitization were 11.576 times that for libraries[18], and 4.135 times that for archives[19]. The other differences between institution types were not significant.
- The odds of Dutch heritage institutions having a higher level of adoption of digitization were 5.022 times that for their Finnish counterparts[20], and 4.980 times that for their Swiss counterparts[21]. The other differences between countries were not significant.
- The odds of institutions for which digital interactive resources are characteristic were 2.329 times that for the other institutions[22].

Regarding the level of adoption of open content (3 categories), the following predictors were found:
- An increase of the perceived effectiveness of skills acquisition (expressed on a continuous scale from 1 to 5) was associated with an increase in the odds of having a higher level of open content adoption, with an odds ratio of 1.448[23].
- An increase of the score for the item saying that they would make content available for commercial users (5-point Likert scale) was associated with an increase in the odds of having a higher level of open content adoption, with an odds ratio of 1.351[24].
- An increase of the level of public funding (expressed as a percentage of institutional funding from public funds in the overall revenue mix) was associated with a decrease in the odds of having a higher level of open content adoption, with an odds ratio of .985[25].

Regarding the level of adoption of social media use (3 categories), the following predictors were found:
- The odds of institutions which count private individuals (the general public) among their main users to have a higher level of adoption of social media were 25.242 times that of the other institutions[26].
- An increase of the number of paid employees (expressed in terms of full-time equivalents) was associated with an increase in the odds of having a higher level of adoption of social media use, with an odds ratio of 1.077[27].
- For institutions with a share of 5% to 99.9% of volunteers in their workforce the odds to have a higher level of adoption of social media use were 14.708 times that of institutions relying exclusively on volunteers[28], and 3.378 times that of institutions with a lower percentage of volunteers in their workforce[29]. At the same time, the odds for institutions with a share of less than 5% of volunteers in their workforce to have a higher level of adoption of social media use were 4.353 times that of institutions relying exclusively on volunteers[30].
- The odds of Dutch heritage institutions to have a higher level of adoption of social media use were 6.785 times that for their Swiss counterparts[31], while the odds of Finnish heritage institutions to have a higher level of adoption of social media use were 5.563 times that for their Swiss counterparts[32]. The other differences between countries were not significant.
- The odds of institutions for which natural resources are characteristic to have a higher level of adoption of social media use were 5.775 times that of the other institutions[33]; the odds of institutions for which time

---

[15] 95% CI: 1.109 to 3.406, $\chi^2(1) = 5.386$, $p = .020$

[16] 95% CI: 2.573 to 74.407), $\chi^2(1) = 9.369$, $p = .002$

[17] 95% CI: 1.305 to 18.723), $\chi^2(1) = 5.530$, $p = .019$

[18] 95% CI: 2.143 to 62.531), $\chi^2(1) = 8.079$, $p = .004$

[19] 95% CI: 1.137 to 15.038), $\chi^2(1) = 4.645$, $p = .031$

[20] 95% CI: 1.312 to 19.224), $\chi^2(1) = 5.552$, $p = .018$

[21] 95% CI: 1.627 to 15.242), $\chi^2(1) = 7.914$, $p = .005$

[22] 95% CI: 1.026 to 5.285), $\chi^2(1) = 4.090$, $p = .043$

[23] 95% CI: 1.012 to 2.074, $\chi^2(1) = 4.090$, $p = .043$

[24] 95% CI: 1.051 to 1.737, $\chi^2(1) = 5.531$, $p = .019$

[25] 95% CI: .972 to .999, $\chi^2(1) = 4.430$, $p = .035$

[26] 95% CI: 3.725 to 171.029, $\chi^2(1) = 10.938$, $p = .001$

[27] 95% CI: 1.019 to 1.138, $\chi^2(1) = 6.794$, $p = .009$

[28] 95% CI: 4.250 to 50.901, $\chi^2(1) = 18.012$, $p < .001$

[29] 95% CI: 1.134 to 10.065, $\chi^2(1) = 4.777$, $p < .029$

[30] 95% CI: 1.000 to 18.945, $\chi^2(1) = 3.843$, $p < .050$

[31] 95% CI: 1.992 to 23.115, $\chi^2(1) = 9.374$, $p = .002$

[32] 95% CI: 1.812 to 17.086, $\chi^2(1) = 8.987$, $p = .003$

[33] 95% CI: 1.365 to 24.427, $\chi^2(1) = 5.679$, $p = .017$

based resources are characteristic were 3.398 times that of the other institutions[34]; And the odds of institutions for which text based resources are characteristic were 3.054 times higher than for the other institutions[35].

- An increase of the level on the local vs. global orientation scale (range: 1 to 4) was associated with an increase in the odds of having a higher level of adoption of social media use, with an odds ratio of 1.864[36].
- An increase of the level of public funding (expressed as a percentage of institutional funding from public funds in the overall revenue mix) was associated with a decrease in the odds of having a higher level of adoption of social media use, with an odds ratio of .979[37]. Similarly, an increase of the level of project funding (in percent of the overall revenue mix) was associated with a decrease in the odds of having a higher level of adoption of social media use, with an odds ratio of .969[38]. Interestingly, when entered into the model, all revenue sources (institutional funding from public funds, from private funds, revenues from sponsorship/donations, from commercial operations, as well as project funding) showed this pattern, even if the effects were not always significant. This might suggest that a good balance between different revenue sources increases the likelihood that a heritage institution has a higher level of adoption of social media use. We tried to test this hypothesis by entering a variable giving the number of different revenue sources with a share of at least 20% of overall revenue into the model; this did however not yield any significant result.

And finally, regarding the level of adoption of collaborative content creation (6 categories), the following predictors were found:
- The odds of Finnish heritage institutions to have a higher level of adoption of collaborative content creation were 3.240 times that for their Swiss counterparts[39], while the odds of Dutch heritage institutions to have a higher level of adoption of collaborative content creation were 2.881 times that for their Swiss counterparts[40]. The other differences between countries were not significant.
- And quite curiously, for heritage institutions for which two-dimensional visual resources are *not* characteristic, the odds of having a higher level of adoption of collaborative content creation were 1.969 times that for institutions for which two-dimensional visual resources *are* characteristic[41].

Summing up, the regression analyses regarding the influence of an institution's characteristics on its likelihood to adopt Internet-related practices yielded the following results:

- The factor "country" has a great influence on all the dependent variables at the exception of the adoption level of open content. We need to be cautious however regarding the interpretation of the statistics regarding the amplitude of this effect, given the fact that some countries showed a much greater self-selection bias regarding the theme of the survey than others. But the observed effects clearly go beyond the self-selection bias, as significant differences also exists between countries with similar levels of selection bias.
- Interestingly, when controlling for all the other factors, the type of heritage institution plays a role only with regard to adoption levels of digitization.
- The type of objects that are characteristic for an institution mainly play a role with regard to the adoption of social media use. Thus, the presence of time based resources, of natural resources, and text based resources seems to favor the adoption of social media use. Regarding the adoption of the other Internet-related practices, the type of objects held by an institution plays no or only a marginal role when controlling for the other factors.
- Interestingly, the type of target users of the institution plays only a role with regard to the adoption of social media use; quite unlike what has been suggested above when relating the results of the z-tests on the various adoption rates.
- As suggested by the results of the z-tests, the geographical reach of an institution plays a role only regarding the adoption of social media use.
- The size of an institution (in terms of full-time equivalents of paid staff) is a very strong predictor regarding the adoption of social media use, but appears to be irrelevant regarding the adoption of the other Internet-related practices. Also we didn't find any effect related to the size of an institution in terms of its budget.
- While the absolute number of volunteers (in terms of full-time equivalents) isn't related to the adoption of any of the Internet-related practices, the presence of volunteers alongside paid staff seems to greatly favor the adoption of social media use.

[34] 95% CI: 1.363 to 8.469, $\chi^2(1) = 6.891$, $p = .009$

[35] 95% CI: 1.091 to 8.550, $\chi^2(1) = 4.518$, $p = .034$

[36] 95% CI: 1.156 to 3.007, $\chi^2(1) = 6.524$, $p = .011$

[37] 95% CI: .962 to .997, $\chi^2(1) = 5.214$, $p = .022$

[38] 95% CI: .942 to .996, $\chi^2(1) = 5.109$, $p = .024$

[39] 95% CI: 1.503 to 6.987, $\chi^2(1) = 8.993$, $p = .003$

[40] 95% CI: 1.294 to 6.414, $\chi^2(1) = 6.716$, $p = .010$

[41] 95% CI: 1.020 to 3.800, $\chi^2(1) = 4.075$, $p = .044$

- While the composition of revenue sources seems to be related to the adoption of open content and social media use, the findings remain somewhat puzzling: On one hand, higher levels of institutional public funding seem to negatively affect an institution's propensity to adopt open content and social media use. On the other hand, none of the other funding sources seems to have a positive effect on the propensity to adopt these practices. Nor does a balanced mix of different funding sources.
- As suggested by the inspection of the crosstabs, the organizational form of an institution (public / private nonprofit / commercial) doesn't have an influence on its adoption of the various Internet-related practices.
- The existence of data is of course a prerequisite for the adoption of an open data approach or the implementation of linked data. It is therefore not surprising that the presence of different metadata types is positively related to an institution's likelihood to adopt open data and linked data.

Apart from the "hard facts", we also tested the predictive value of skills-related factors and "open content"-related attitudes:
- Regarding skills in general, we got some mixed results: regarding the adoption of open data and linked data, it appears that the number of different ways the staff of an institution uses to acquire skills and know-how has quite a strong predictive value. With regard to the other practices, this factor doesn't seem to play a role though. It was however found that the perceived effectiveness of skills acquisition played a role with regard to the adoption of open content, and a higher overall skills level was found to be positively associated with the adoption of linked data. Due to multicollinearity issues, we did not include the skills in individual areas of competence into the model. This further differentiation of the role of skills regarding the adoption of the various practices should be the object of further investigation.
- Regarding attitudes we found rather unsurprisingly that an institution's readiness to make content available for commercial users is positively related to its adoption of open content. Somewhat less intuitive is the insight that institutions with a general "sharing" attitude, as expressed through their readiness to make content available for non-profit projects which permit their users commercial use of the content, are more likely to engage in linked data endeavors. And the same goes for the finding that institutions which would make content available only under the restriction that it won't be modified are less likely to adopt an open data approach. This is somewhat surprising as opening up metadata isn't necessarily related to opening up content (although the opposite is generally true).

All in all, there is no set of characteristics of an institution which would make it generally prone to be among the most advanced among its peers regarding the adoption of all the Internet-related practices under examination. It appears that only the "country" factor has an influence on the adoption of almost all the practices.

6.3 Attitudes regarding open data, open content, and collaborative content creation

We asked the responding institutions about their attitudes regarding the innovative practices under examination, namely the perceived importance of the various practices as well as the perceived risks and opportunities. Unsurprisingly, perceived importance of a particular practice is strongly correlated with the adoption of the given practice. Similarly, the perceived desirability (opportunities vs. risks) of a given practice is correlated with the adoption of the practice, although the association is less strong in some cases (as expressed by the Pearson χ^2 statistics). There is, however, one exception: adoption of collaborative content creation doesn't correlate with the perceived desirability of this practice; there are in fact many institutions which consider the opportunities of collaborative content creation at least as important as the risks, but haven't started to implement it. These findings are consistent with the observation that the perceived importance statistics (figure 4) follows exactly the same pattern as the adoption rate statistics (figure 1), while the perceived desirability statistics (figure 5) show a different picture.

While open data is clearly rated more important than linked data, it appears that the risks of open data are rated higher (both differences are significant at the .05 level). A similar pattern can be observed for open content compared to collaborative content creation, with the risk evaluation for collaborative content creation being about the same as for open content. It seems that the idea of opening up data or content in a way that they can be freely used, modified, and shared by anyone for any purpose (including commercial use) is still met with some hesitations in the heritage sector. This is probably best exemplified by the results regarding the question under what conditions institutions would be ready to make their content available on the Internet without receiving payment in exchange (figure 6): while a "sharing" attitude clearly prevails regarding education and research or non-profit projects (with acceptation rates of 91% and 81%, respectively), heritage institutions are less inclined to make content available for private use (67%) or for non-profit projects, such as Wikipedia, which permit their users commercial use of the content (48%). One of the big barriers appears to be the prospect of "commercial use" of the content by third parties without compensation for the institution, as illustrated by the particularly low acceptation rate for commercial use (18%). There is quite a big dissonance between this reluctance of heritage institutions regarding the opening up of content for commercial re-use on one hand, and the increasing calls for open data and open content policies in the heritage sector on the other hand. Similarly, there is a gap of 30 percent points between the institutions' readiness to make content available for Wikipedia and their readiness to release content also for commercial re-use; in fact, releasing content under a public domain mark or under an open copyright license, which by definition allows for commercial use of the licensed material, is a prerequisite of its re-use in the context of Wikipedia. As we have seen in the regression analysis reported above, an institution's attitude regarding commercial re-use is an important predictor of the adoption of open content. Another "blocker" regarding the adoption of an open content policy is many institutions'

(64%) reluctance to let others modify content. Interestingly, the fact that an institution wasn't insisting on this restriction was found to be a strong predictor for the adoption of open data.
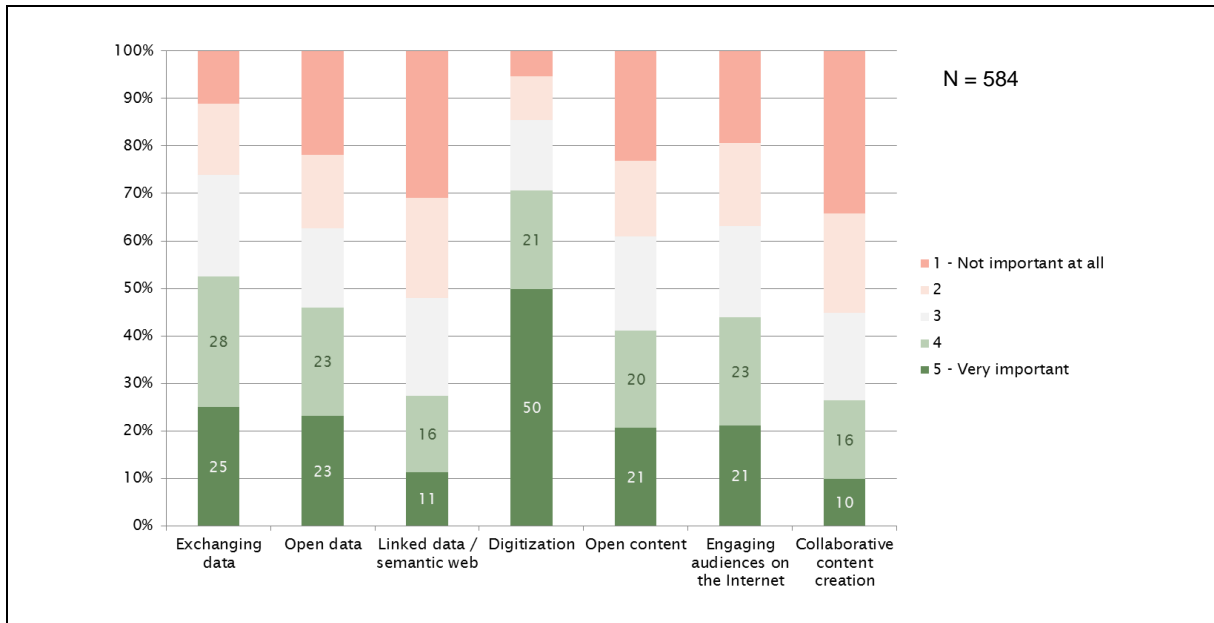

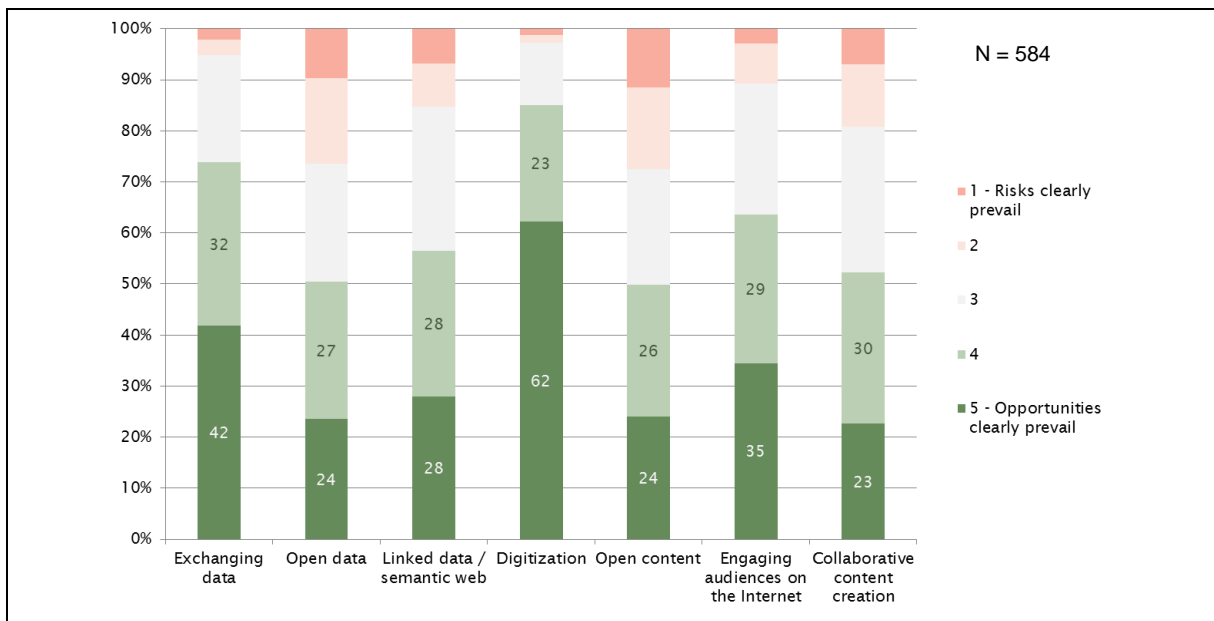Figure 4: Perceived importance of various practices

Figure 5: Desirability (opportunities vs. risks) of various practices
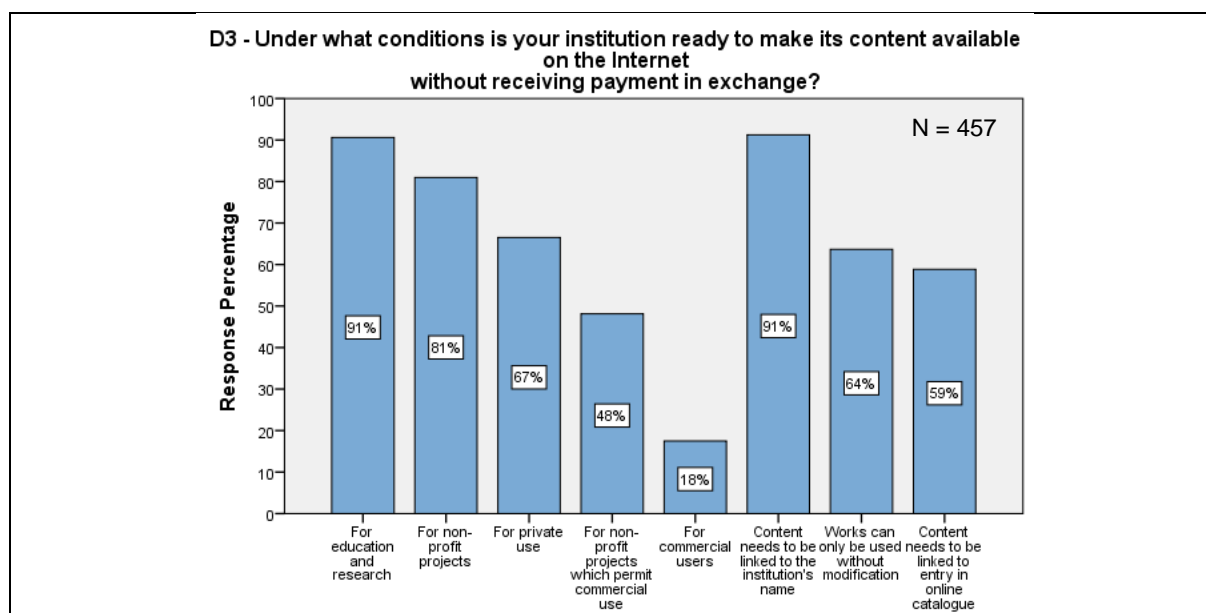
Figure 6: Attitudes regarding the opening up of content

6.4 Expected dynamics regarding the adoption of open data, open content and crowdsourcing

In order to get an idea of the dynamics regarding the future adoption of open data, open content, and crowdsourcing we have put the importance and the desirability statistics in relation to each other, which allowed us to identify on one hand the percentage of institutions which consider a particular practice both as important and desirable (i.e. the expected "motors" of the future adoption process), and on the other hand the percentage of institutions which think that a given practice is both unimportant for them and associated with higher risks than opportunities (i.e. the institutions which can be expected to be still very reluctant to adopt a given practice). Regarding open data, the group of "motors" is made up of 29% of institutions, while 13% fall into the reluctant category, and another 19% failed to provide information on the perceived importance and/or the desirability. Regarding open content, 27% of institutions are among the "motors", while 14% fall into the reluctant category (with 20% missing). And regarding collaborative content creation, 18% are among the "motors", while 11% fall into the reluctant category (with 25% missing). By relating the share of the "motors" to the share of the reluctant institutions, we calculated a ratio that can serve as an indicator for the dynamics of the future adoption of the given practices. As it turns out, among the three practices, this ratio is highest for open data, with a motors/hesitant ratio of 2.14, compared to 1.94 for open content, and 1.61 for collaborative content creation.

A further indicator regarding the existing potential for the adoption of open data and open content is the presence (or absence) of important prerequisites or "show-stoppers". In the case of open data, this is primarily the existence of metadata within the institution, for as we have seen in the regression analysis, availability of metadata is an important predictor for the adoption of open data and linked data. 29.9% of responding institutions indicate that they don't have metadata (e.g. catalogues inventories or finding aids). This means that for roughly 30% of institutions, adopting an open data policy would presently not be very meaningful, unless they first create metadata for their heritage objects. Given the present adoption rate for open data of 26%, this leaves us with 44% of institutions which may still be expected to adopt an open data policy.

In view of the adoption of open content, important prerequisites comprise the existence of digitized material and the institution's entitlement to release the content as public domain material or under a free copyright license. As to digitization, the advancement depends a lot on the type of heritage objects (as can be seen in table 3): while 20% of two-dimensional visual resources and three-dimensional man-made movable objects have already been digitized by the average institution (median), this percentage is much lower for archival resources (2%), text based resources (5%), or natural resources (5%). At the same time, the statistics in the lower part of table 3 shows that we can expect large advances in digitization over the coming 5 years: For all object types the percentage of objects that will have been digitized by the average institution, is expected to increase by at least 100%, and in some cases the percentage of digitized objects is expected to be at least 4 to 5 times higher than at present (archival resources, time-based resources, geography-based resources, and text-based resources). As table 4 shows, for the average institution the advancement of digitization is clearly not the limiting factor when it comes to releasing materials as open content today. Within the next five years, this picture is however expected to change, as much higher percentages of digitized material will have been made available as open content.

Statistics

		Text based resources: already digitized	Two-dimensional visual resources: already digitized	Archival resources: already digitized	Three-dimensional man-made movable objects: already digitized	Natural resources: already digitized	Geography based resources: already digitized	Time based resources: already digitized	Average percentage of objects already digitized
N	Valid	366	357	282	314	47	120	221	543
	Missing	218	227	302	270	537	464	363	41
Mean		17.2350	33.1997	14.9610	34.6153	25.1277	32.5258	32.3385	30.2547
Median		5.0000	20.0000	2.0000	20.0000	5.0000	10.0000	10.0000	25.0000
Percentiles	25	.0000	2.0000	.0000	.0750	.0000	.0000	.0000	5.0000
	50	5.0000	20.0000	2.0000	20.0000	5.0000	10.0000	10.0000	25.0000
	75	20.0000	60.0000	10.0000	70.0000	50.0000	67.5000	60.0000	49.0000

Statistics

		Text based resources: digitized in 5 years	Two-dimensional visual resources: digitized in 5 years	Archival resources: digitized in 5 years	Three-dimensional man-made movable objects: digitized in 5 years	Natural resources: digitized in 5 years	Geography based resources: digitized in 5 years	Time based resources: digitized in 5 years	Average percentage of objects expected to be digitized in 5 years
N	Valid	319	317	252	279	39	100	197	489
	Missing	265	267	332	305	545	484	387	95
Mean		32.1520	51.7565	30.2012	53.9631	33.1538	44.9410	50.0706	44.3729
Median		20.0000	50.0000	15.0000	50.0000	10.0000	45.0000	50.0000	40.0000
Percentiles	25	2.0000	15.0000	2.0000	15.0000	2.0000	4.2500	10.0000	20.0000
	50	20.0000	50.0000	15.0000	50.0000	10.0000	45.0000	50.0000	40.0000
	75	50.0000	90.0000	50.0000	95.0000	60.0000	88.7500	90.0000	66.6667

Table 3: Progress of digitization (today and in 5 years), according to different types of heritage objects

Statistics

		Text based resources: available as open content (% of digitized)	Two-dimensional visual resources: available as open content (% of digitized)	Archival resources: available as open content (% of digitized)	Three-dimensional man-made movable objects: available as open content (% of digitized)	Natural resources: available as open content (% of digitized)	Geography based resources: available as open content (% of digitized)	Time based resources: available as open content (% of digitized)	Digital interactive resources: available as open content (% of digitized)	Average percentage of objects presently available as "open content" (in % of digitized objects)
N	Valid	222	257	160	203	22	64	143	124	430
	Missing	362	327	424	381	562	520	441	460	154
Mean		39.1164	29.2038	37.1921	20.9547	29.6591	38.4321	22.9816	13.4040	23.7521
Median		6.3492	.0000	.0000	.0000	.0000	5.0000	.0000	.0000	4.1479
Percentiles	25	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	50	6.3492	.0000	.0000	.0000	.0000	5.0000	.0000	.0000	4.1479
	75	100.0000	66.6667	100.0000	20.0000	70.0000	100.0000	20.0000	10.0000	44.3254

Statistics

		Text based resources: available as open content in 5 years (% of digitized)	Two-dimensional visual resources: available as open content in 5 years (% of digitized)	Archival resources: available as open content in 5 years (% of digitized)	Three-dimensional man-made movable objects: available as open content in 5 years (% of digitized)	Natural resources: available as open content in 5 years (% of digitized)	Geography based resources: available as open content in 5 years (% of digitized)	Time based resources: available as open content in 5 years (% of digitized)	Digital interactive resources: available as open content in 5 years (% of digitized)	Average percentage of objects available as "open content" in 5 years (in % of digitized objects)
N	Valid	211	234	166	188	28	64	144	109	390
	Missing	373	350	418	396	556	520	440	475	194
Mean		47.2222	42.6961	44.7670	39.2118	45.6349	60.3398	38.1857	28.1385	36.0464
Median		40.0000	30.6250	25.0000	22.2222	45.0000	81.6667	20.0000	5.0000	25.8681
Percentiles	25	.0000	.0000	.0000	.0000	.0000	11.6667	.0000	.0000	1.0000
	50	40.0000	30.6250	25.0000	22.2222	45.0000	81.6667	20.0000	5.0000	25.8681
	75	100.0000	100.0000	100.0000	88.5526	100.0000	100.0000	95.8333	50.0000	60.0000

Table 4: Open content (today and in 5 years), according to different types of heritage objects, in % of digitized objects

Regarding the copyright status of heritage material, it appears that for all the object types, more than half of the objects are either in the public domain or the copyright is controlled by the institution (figure 7). Thus, it appears that across for the average institution, copyright is not the limiting factor so far when it comes to opening up content. This doesn't however exclude that for individual institutions this may be the case. Furthermore, this data doesn't indicate whether the institutions already know for all the objects in their collection what copyright status they have or whether they still would have to put in some time for research to assess the copyright status before releasing material under a free copyright license. Again, our data suggests that in five years from now, copyright may more generally start becoming a limitation regarding the release of open content. Thus, 11% of institutions report that in 5 years from now they will have released at least half of their text based resources as open content. For archival resources, natural resources, and time based resources, the share of institutions planning to release at least half of their material as open content is roughly the same (12-13%), while it is notably higher for three-dimensional man-made movable objects (17%), for two-dimensional visual resources (21%), for digital interactive resources (22%), and for geography based resources (24%).
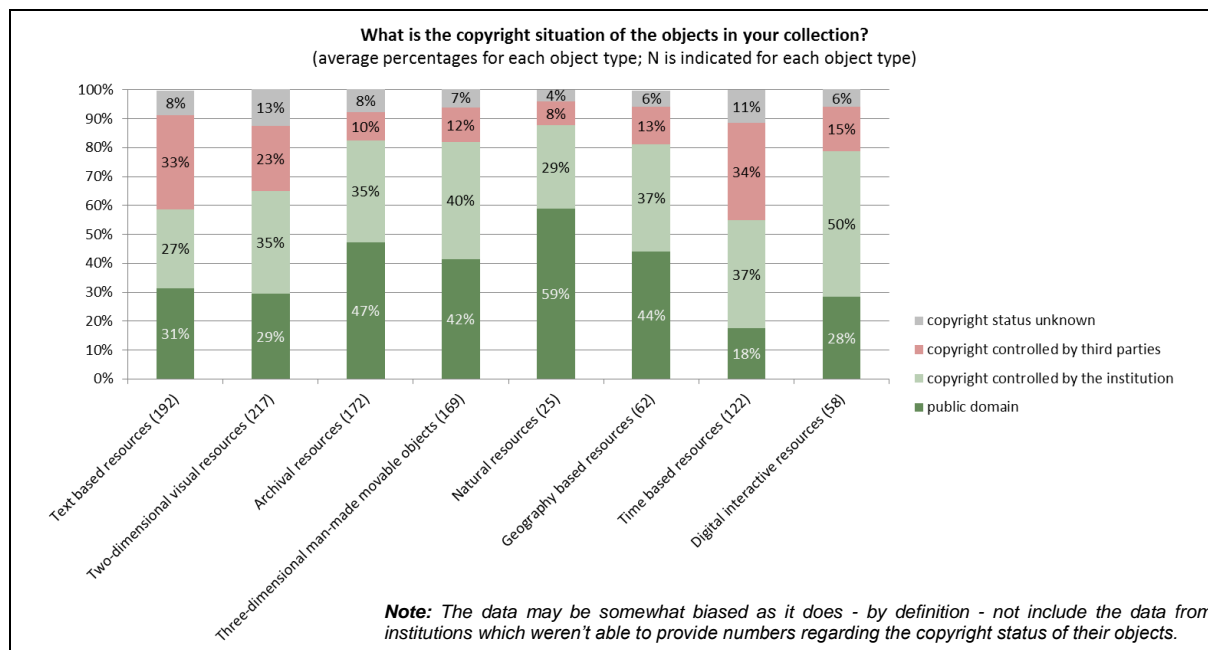
Figure 7: Copyright situation of the objects in the collection.

And finally, we can estimate the future adoption rate of the various practices based on the institutions' indications regarding their future practices. Thus, 48% of the responding institutions expect to have released more than 10% of their metadata as open data in 5 years from now (compared to 26% today), while 43% of institutions expect to have released more than 5% of their content as open content in 5 years from now (compared to 26% today). As to collaborative content creation, 27% of institutions expect to use at least one type of collaborative content creation within the next year (compared to 21% today). Given these results, it appears that the diffusion of open data and open content will progress at a similar pace (which is in line with our findings regarding the motors/hesitant ratio). Regarding the diffusion of collaborative content creation, the numbers are not really comparable; in this context, it should also be kept in mind that the practice of collaborative content creation is reversible (i.e. an institution may abandon the practice), while the indicators we use for open data and open content are irreversible, due to the fact that data or content that has been released under a free license generally cannot be taken back in order to again be locked up.

6.5 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 content policies and crowdsourcing approaches. In order not to overload the questionnaire and to avoid confusion among the respondents due to repetitive items we did not ask specifically about risks and opportunities of open data, but focused exclusively on open content.

6.5.1 Challenges, risks and opportunities of open content

The greatest benefits of open content from the point of view of the responding institutions is the fact that it improves the visibility or perceived relevance of the institution (mentioned by 88% of responding institutions), that it improves the discoverability of its holdings (85%), and that content becomes more easily available to existing users (80%). Other important benefits and opportunities include the fact that open content attracts new users (73%), that it facilitates networking among heritage institutions (72%), and that it improves interactions with users (71%). As a result, 70% of the responding institutions reckon that opening up their content allows the institution to better fulfill its core mission. Potential benefits that were mentioned less often are the enhancement of transparency and accountability (49%), one of the key arguments for the promotion of open government data, as well as the reduction of legal complexity (32%).

When it comes to implementing an open content strategy, the main challenges the responding institutions are facing are the extra time effort and expense related to the digitization of holdings (90% consider this as a challenge) and the time effort and expense related to proper documentation of the content (83%). Further important challenges include technical challenges (63%), challenges related to staff skills (59%), as well as time effort and expense related to rights clearance (54%). Other potential challenges, such as difficulties to track the use of the content, e.g. for usage statistics (37%), unknown copyright holders (orphan works) (36%), and third party copyright holders unwilling to release content under a "free" license (33%) are of concern only to about a third of institutions. When asked about the risks of opening up content, roughly three quarters mention re-use without proper attribution to the institution

(77%), re-use without proper attribution to the author/creator (76%), and mis-use / mis-representation of content (72%), whereas a bit more than half of them are worried about copyright infringements by third parties (56%). These risks, which were rated highest, all seem to be related to a sense of loss of control. Further risks comprise legal risks related to copyright infringements committed by the own institution (47%), increased time effort in order to respond to enquiries (47%), risks due to infringements of data protection regulations / divulgence of classified information (41%), desecration of places, rites, objects (24%), as well as destruction / robbery / disturbance of endangered or valuable objects or living organisms (23%). Interestingly, loss of revenues (17%) and loss of brand value (15%) were rated very low.

6.5.2 Purpose, risks and challenges of crowdsourcing

In order to get an idea what crowdsourcing may mostly be used for by heritage institutions, we asked them which crowdsourcing approaches they are using at present or planning to use within the next year. The typology used in the questionnaire corresponds to the classification scheme proposed by Oomen and Aroyo (2011) based on the digital content life cycle model of the National Library of New Zealand. According to our data[42], the approaches used most often aim at the supplementation of collections (53%) and at contextualization (46%), followed by classification and supplementation of metadata (39%), as well as correction and transcription tasks (37%), whereas crowdfunding (26%), and co-curation (26%) were mentioned less often.

The institutions which indicated that they are presently using at least one crowdsourcing approach or are planning to do so within the next year were asked about the purpose of their use of crowdsourcing and the main risks and challenges. All the suggested purposes were mentioned by more than half of the respondents: gaining access to external expertise (71%), experimenting with new ways of relating to the users/visitors (70%), increasing trust and loyalty of the users/visitors with regard to the institution (63%), giving the users/visitors a sense of public ownership and responsibility (59%), and having certain tasks carried out in spite of resource constraints (56%).

The risks and challenges mentioned most often were "extensive preparation and follow-up required" (70%) and "difficulties to estimate the time scope" (67%), followed by "limited planning security" (57%) and "the continuity of data maintenance is not guaranteed" (57%). In contrast, "little influence on results" (39%) seems to be of minor concern, and "anxiety among employees (loss of job, changes to roles and tasks, etc.) (14%) hardly preoccupies any institutions.

---

[42] The data reported here is based on the responses by the institutions which indicated that they are presently using at least one crowdsourcing approach or are planning to use one within the next year (N = 171).

7 Discussion

We would like to focus our discussion mainly on three points: First, we shall discuss the insights gathered regarding the influence of various characteristics of heritage institutions on their adoption of open data, open content, and crowdsourcing in light of some of the hypotheses we formulated at the time of questionnaire development. Second, we shall shortly discuss the insights gathered regarding the dynamics of the adoption process of the three practices. And third, we shall discuss our findings regarding the driving and hindering factors of open data / open content and crowdsourcing in light of the existing literature.

7.1 Influence of various characteristics of heritage institutions on their adoption of open data, open content, and crowdsourcing

When developing the questionnaire we formulated a series of hypotheses, which were inspired by our earlier research (Estermann 2014), by insights from the literature on innovation diffusion (mainly Rogers 2003), as well as by our practical experience while reaching out to heritage institutions on behalf of the Wikipedia/Wikimedia community and the Open Knowledge Foundation. In the following, we will shortly discuss these hypotheses in the light of the findings presented in this paper:

H1.1: There are differences regarding the adoption of innovative practices based on contextual factors related to country, region, institution type, geographical reach (the wider the reach, the more innovation-prone), and revenue sources (the higher the share of project funding, the more innovation prone)

We have indeed found that the "country" factor has a great influence on the adoption of all the examined practices at the exception of open content (differences between regions of a single country haven't been covered yet). The observed effects go beyond the country differences regarding the self-selection bias. It should however be kept in mind that adoption rates for The Netherlands and for Poland are most likely overestimated. In order to allow for reliable benchmarking between the countries, the differences regarding the self-selection bias should be adjusted for. The institution type does play a role, but when controlling for other factors it has an influence only with regard to digitization activities. Interestingly, this effect doesn't seem to be linked to the types of heritage objects an institution holds. So far, no convincing explanation has been presented for this finding. It would be interesting to verify whether this effect is also present in the data gathered by the ENUMERATE surveys. Interestingly, geographical reach seems to play a role only with regard to the use of social media. The finding from the innovation diffusion literature that earlier adopters are more cosmopolite than are later adopters (Rogers 2003, p. 290) doesn't seem to hold among heritage institutions; or it might be that geographical reach of an institution is a bad proxy for the cosmopolite character of its staff members. While the composition of revenue sources seems to play a certain role regarding the adoption of open content and social media, the hypothesis that higher shares of project funding would be correlated with increased innovativeness hasn't been confirmed.

H1.2: When controlling for the other variables, there are no differences related to language, types of heritage objects, main users, and form of organization.

As with the differences between geographical regions of a single country, we haven't yet analyzed the differences between various language regions, nor does the data allow for the analysis of linguistic groups at an international level. The type of heritage objects plays a role with regard to the adoption of social media use, which is rather unexpected. Furthermore, some minor effects have been identified regarding the adoption of digitization and collaborative content creation. With regard to the other practices, the type of heritage objects that are characteristic for an institution doesn't seem to play a role. The main users don't play a role as expected, except for the use of social media, which is highly correlated with having private individuals (the general public) as main target users. And finally, the form of organization does not play a role as expected.

H1.3: The size of the institutions plays a role in that larger institutions are more innovative than smaller institutions on average. The group of first mover institutions, however, is composed of similar shares of large, medium-sized, and smaller institutions.

One of the insights derived from innovation diffusion research is that earlier adopters tend to be larger in size than do later adopters (Rogers 2003, p. 288). If analyzed separately, the institutions' size (in terms of number of paid staff and of total revenues) was indeed found to be positively correlated with the adoption of open data, linked data, social media use, and collaborative content creation, but not with their adoption of digitization and open content. However, our data suggests that this effect is not directly related to the size of an institution; the only size effect that was found after controlling for other factors was related to social media use. Given the much larger number of small institutions compared to larger institutions, they also represent a relatively large share of the early adopters. Thus, institutions with 5 or fewer full-time equivalents of paid staff (which make up 59% of the entire sample) represent 40% of institutions at the advanced implementation stage of open data, 24% of institutions at the advanced implementation stage of linked data, 64% of institutions at the advanced implementation stage of open content, and 52% of institutions at the advanced implementation stage of collaborative content creation.

H1.4:   The higher the perceived overall competency level, the more innovation-prone is an institution.

According to the innovation diffusion literature, early adopters tend to have higher levels of education, and to have more intelligence than later adopters (Rogers 2003, p. 288-289). In our analyses we haven't found however any evidence for a strong link between perceived overall competency and early adoption of innovations. The only area where we found a (rather weak) effect was related to the adoption of linked data. This may be due to the fact that we didn't measure competency objectively, but through self-reporting without controlling for different levels of ambition. Thus, an institution which permanently innovates may also perceive a permanent need to improve its staff skills.

H1.5:   Institutions with more effective know-how acquisition are more innovation-prone.

According to innovation diffusion research, early adopters have greater knowledge of innovations than later adopters (Rogers 2003, p. 291). A link between more effective know-how acquisition and early adoption of innovative practices could however only be established regarding the adoption of open content. In the other areas, this didn't seem to play an important role.

H1.6:   Institutions using more channels of know-how acquisition are more innovation-prone.

According to innovation diffusion research, early adopters seek information about innovations more actively than later adopters (Rogers 2003, p. 291). Institutions using more channels for know-how acquisition were indeed found to have higher adoption levels regarding open data and linked data. But again, the same relationship hasn't been found with regard to the other practices.

In addition to the hypotheses regarding innovation adoption in general, we have also formulated hypotheses regarding the adoption of particular Internet-related practices, some of which have been covered by the analyses presented in this paper. Here a summary of our findings: The hypotheses that institutions holding various types of metadata are more likely to adopt open data and that institutions holding name authority files and/or glossaries/vocabularies/onotologies are more likely to adopt linked data are supported by the data. So is the hypothesis that positive attitudes regarding open data principles are positively related to the adoption of open content. Surprisingly, no link was found between the number of volunteers an institution has and its propensity to adopt collaborative content creation. There are several further hypotheses that haven't been covered yet, but might be worthwhile analyzing in the future, e.g. pertaining to the correlation between levels of specific skills and the adoption of particular practices, to the correlation between the adoption of different Internet-related practices (e.g. the link between the progress of digitization and the adoption of open content), or to the relation between the copyright status of heritage objects in an institution's collection and its adoption of open content.

7.2   The dynamics of the adoption of open data, open content, and crowdsourcing

Various indicators suggest that open data and open content will diffuse at a similar rate throughout the heritage sector, while the widespread adoption of crowdsourcing may take a bit longer. These findings are in line with our earlier findings based on the Swiss pilot survey. The expected adoption rate of open data is quite high: the number of institutions which will have released more than 10% of their metadata as open data is expected to rise from 26% today to 48% within the next five years. There is a rather hard limit to the diffusion of open data: about 30% of responding institutions don't have metadata for their heritage objects. This number seems rather high and roughly corresponds to the 32% of institutions which either belong to the "reluctant" category regarding open data or do not have an opinion about its importance, risks, and opportunities. While it is true that the absence of metadata concerns mainly small and very small institutions, we don't have any reliable information regarding the size and the importance of the holdings concerned. If the dissemination of open data continues at the rate suggested by our survey results, we can expect that all the institutions which have metadata (i.e. 70% of institutions) will have adopted open data in about 10 years from now.

Despite some serious challenges, the adoption rate of open content is not much lower than the one for open data: the number of institutions which will have released more than 5% of their content as open content is expected to go up from 26% to 43% over the next five years. At present, there are two main impediments regarding the opening up of content: On one hand the time effort and expense related to the digitization of the holdings, to the documentation of the content, and to rights clearance. And on the other hand negative attitudes among many institutions regarding the opening up of content, inspired by a feeling of loss of control and the wish to prevent commercial use of content by third parties without due compensation. Further challenges include technical issues and insufficient staff skills. There are "hard" limits regarding the adoption of open content posed by the advancement of digitization and by copyright. While both limits only play a minor role regarding the opening up of content today, they are expected to increase in importance over the coming years, as they become for more and more institutions the limiting factor regarding the release of content. By the same measures of calculation as we have used for open data, we can expect that 70% of institutions will have adopted open content in about 15 years from now.

21% of responding institutions are using crowdsourcing today, and 6% said that they will start using it within the coming year. At first sight these data suggest a high adoption rate. We have however no explicit data that would give us an idea how many institutions will stop using crowdsourcing. Other indicators, like the ratio between "motors" and

institutions that are rather reluctant regarding crowdsourcing, suggest that the adoption rate will be lower than for open data and open content.

7.3 Driving and hindering factors of open content and crowdsourcing from the point of view of the institutions

Regarding the opportunities, challenges and risks of open content, the earlier findings of the Swiss pilot survey (Estermann 2014) have been widely confirmed[43]. The findings were also mostly in line with those of the earlier studies (Balthussen et al. 2013, Kelly 2013, Eschenfelder and Caswell 2010), although none of them allows for a direct comparison of results. Interestingly, the extra time effort and expenses, which were perceived as the greatest challenge in our surveys, was mentioned only by Kelly (2013) 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 roughly half of our respondents, was mentioned only by Kelly (2013). 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. Our survey shows however that there are serious reservations among heritage institutions regarding the commercial use of their content. It is therefore possible that many institutions are abhorred by the (theoretical) prospect of third parties earning money thanks to the release of an institution's content, without shouldering their "fair" share of the costs. This would correspond to Balthussen's finding that heritage institutions were afraid that they may fail to generate extra income in the future as third parties develop new business models based on their datasets (Balthussen et al. 2013).

With regard to crowdsourcing, the ratings of the risks were less uniform than the ones in the pilot survey. They may also have been more informed, as the question was only asked to institutions which actually are using crowdsourcing approaches or are planning to use them in the following year, while in the pilot survey the question was asked to all respondents. In both surveys, the item regarding the extensive preparation and follow-up ranked highest at about the same rate (70% vs. 72% of institutions), followed by the difficulties to estimate the time scope (67% vs. 70%). The low level of planning security (57% vs. 60%) also received very similar scores, while the fact that the continuity of data maintenance was not guaranteed, scored lower in the international survey (57% vs. 66%), although given the relatively small sample sizes, this difference is not significant at the 0.05 level. The fact that institutions had little influence on the results scored 39% in the international survey, while the item "unforeseeable results" had scored with 61% in the pilot survey; this difference may however be due to the different formulation of the item. Finally, the item "Anxiety among employees (loss of job, changes to roles and tasks, etc.)" ranked very low with 14% (compared to 23% in the pilot survey). The item partly covers the issues recently brought up by Perry and Beale (2015), who point to ethical challenges related to fair pay, fair working conditions, social equality, and basic human rights as heritage practitioners (or more specifically archeologists) increasingly rely on crowdsourcing. We did however not specifically ask about "ethical" issues in connection to the increasing reliance on volunteers. This aspect should be considered for inclusion in future studies.

As to the most popular crowdsourcing approaches, the findings were somewhat different from the findings of the pilot survey, with "supplementation of collections" and "contextualization" ranking first, while the pilot survey had found that "classification / completion of metadata" and "correction and transcription tasks" were the greatest opportunities related to crowdsourcing from the perspective of the responding heritage institutions. For the first time, the international survey also included a question about the purpose of crowdsourcing. Here the main focus was on gaining access to external expertise and on experimenting with new ways of relating to users/visitors, but the other items, such as increasing trust and loyalty of the users/visitors, giving them a sense of public ownership and responsibility, as well as having certain tasks carried out in spite of resource constraints were also mentioned by more than half of the respondents. There is little research into heritage institutions' motivations for crowdsourcing. One notable exception is a case study carried out by Alam and Campbell (2013) in order 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. At first sight, the results of our study are in line with these findings in that all the motivations and purposes mentioned were also indicated by many of our respondents. However, we haven't been able to verify the dynamic aspect by comparing the motivations of the institutions which are going to use crowdsourcing to the ones who have been using crowdsourcing before.

---

[43] The Swiss pilot survey did not differentiate between open data and open content. The way the questions regarding "open data" were formulated implied that "open content" was meant as well.

8 Conclusions and Outlook

In the present paper, we have been able to largely confirm the earlier findings from the Swiss pilot survey and to refine them by examining the different factors influencing the diffusion of the various Internet-related practices and by getting a better grasp of the dynamics of the diffusion of open data and crowdsourcing among heritage institutions in Europe.

The analysis of the data has shown that, if the survey is to be used for benchmarking purposes between different countries, the self-selection bias present in some of the samples should be corrected for. Furthermore, based on the insights presented in this paper, we would suggest the following paths for further inquiry:

- Carry out an analysis of the context factors in the various countries which influence the adoption of open data and crowdsourcing among heritage institutions.
- Further investigate the links and mutual influences between the various Internet-related practices, in order to refine the model regarding the factors that influence the adoption of open data and crowdsourcing by heritage institutions.
- 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.

Further aspects that could also be investigated based on our data:
- Analyze the dissemination of digitization within the heritage sector, by comparing our data to the data collected through the ENUMERATE survey.
- Get a deeper understanding of the use of social media by heritage institutions and examine potential relations between the use of social media and the adoption of crowdsourcing approaches.

Acknowledgements

References

- Alam, S. and Campbell, J. (2013). Dynamic Changes in Organizational Motivations to Crowdsourcing for GLAMs. Thirty Fourth International Conference on Information Systems, Milan 2013.
- Allen, N. (2009). Art museum images in scholarly publishing. Rice University Press.
- Bakker, S., de Niet, M, and Nauta, G. J. (2012). Overview of National and International Initiatives. Deliverable 2.1 of the ENUMERATE Project. Retrieved on January 28, 2014 from: http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE_Overview-Initiatives.pdf
- Baltussen, L.B., Brinkerink, M., Zeinstra, M. J. Oomen, and N. Timmermans (2013) Open Culture Data: Opening GLAM Data Bottom-up. Paper presented at Museums and the Web, 2013, Portland OR, USA.
- Bauer, F. and Kaltenböck, M. (2001). Linked Open Data: The Essentials: A Quick Start Guide for Decision Makers. SEMANTIC-WEB COMPANY, 2011.
- Beal, G.M. and Bohlen, J. M. (1957) Special Report No. 18, Agriculture Extension Service, Iowa State College, 1, pp. 56–77, 1957. Reprinted in 1981.
- Bundesamt für politische Bildung (2011). Open Data. Retrieved on 23 April 2013 from http://www.bpb.de/gesellschaft/medien/opendata/64055/was-sind-offene-daten
- Carletti, L., Giannachi, G., Price, D., McAuley, D. and Benford, S. (2013) Digital Humanities and Crowdsourcing: An Exploration. Paper presented at Museums and the Web 2013, Portland OR, USA.
- Christensen, J. R. (2010). Four steps in the history of museum technologies and visitors' digital participation. MedieKultur. Journal of media and communication research, 50, pp. 7-29.
- Crews, K.D. (2012). Museum Policies and Art Images: Conflicting Objectives and Copyright Overreaching (July 1, 2012). Fordham Intellectual Property, Media & Entertainment Law Journal, Vol. 22, p. 795, 2012. Available at SSRN: http://ssrn.com/abstract=2120210
- Dooley, J.M., and Luce, K. (2010). Taking our pulse: The OCLC Research survey of special collections and archives. Dublin, Ohio: OCLC.
- Eschenfelder, K. R., and Caswell, M. (2010). Digital cultural collections in an age of reuse and remixes. Proceedings of the American Society for Information Science and Technology, 47(1), pp. 1-10.
- Estellés-Arolas, E. and González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information Science, 38(2), pp. 189-200.
- Estermann, B. (2013). Swiss Heritage Institutions in the Internet Era. Results of a pilot survey on open data and crowdsourcing. Bern University of Applied Sciences, E-Government Institute.
- Estermann, B. (2014). Diffusion of Open Data and Crowdsourcing among Heritage Institutions: Results of a Pilot Survey in Switzerland. JTAER(2014)15-31.
- European Commission (2001a). European Content in Global Networks: Coordination Mechanisms for Digitisation Programmes (Lund Principles), 4 April 2001.
- European Commission (2001b). Action Plan on Coordination of Digitisation Programmes and Policies (Lund Action Plan), 23 July 2001.
- European Commission, Information Society DG., and Salzburg Research (2002). 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.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Quarterly, 82(4), 581-629.
- Holley, R. (2010). Crowdsourcing: how and why should libraries do it?. D-Lib Magazine, 16(3), 4.
- Howe, J. (2006), Crowdsourcing: A Definition. Crowdsourcing Blog, June 2, 2006. Retrieved on January 24, 2013 from http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html
- Jankowski, J., Cobos, Y., Hausenblas, M., and Decker, S. (2009). Accessing cultural heritage using the Web of Data. The 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST, 2009.
- Kelly, Kristin (2013). 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.

- Manžuch, Z. (2009). Monitoring digitisation: lessons from previous experiences. Journal of Documentation, 65(5), 768-796.
- Nauta, G.J., Bakker, S., and de Niet, M. (2011). ENUMERATE Core Survey 1 Methodology, November 2011. Available: http://www.egmus.eu/fileadmin/ENUMERATE/deliverables/ENUMERATE-D2-04.pdf (accessed on 30 July 2015).
- Nicholls, Ann / Pereira, Manuela / Sani, Margehrita (eds.) (2012) Report 1 - The Virtual Museum. The Learning Museum Network Project.
- Oomen, J. and Aroyo, L. (2011). Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. C&T' 11, QUT, Brisbane, Australia, 20 June - 2 July 2011.
- Perry, S. and Beale, N. (2015). The Social Web and Archaeology's Restructuring: Impact, Exploitation,
- Disciplinary Change. Open Archeology 2015; 1:153-165.
- Phillips, L. B. (2013). The Temple and the *Bazaar: Wikipedia as a Platform for Open Authority in Museums. Curator: The Museum Journal, 56(2), 219-235.*
- Rogers, E. M. (2003). Diffusion of innovations/Everett M. Rogers.
- Smith-Yoshimura, K., Godby, C.J., Koffler, H., Varnum, K. and Yakel, E. (2011). Social Metadata for Libraries, Archives and Museums. Part 2: Survey Analysis. Dublin, Ohio: OCLC Research.
- Smith-Yoshimura, K. / Schein, C. (2011). Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews. Dublin, Ohio: OCLC Research.
- Stroeker, N. and Vogels, R. (2012). Survey Report on Digitisation in European Cultural Heritage Institutions 2012. ENUMERATE Thematic Network, May 2012.
- Stroeker, N., and Vogels, R. (2014). Survey Report on Digitisation in European Cultural Heritage Institutions 2014. ENUMERATE Project. Available: http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2014.pdf (accessed on 30 July 2015)
- Stroeker, N., Vogels, R., Nauta, G.J. and de Niet, M. (2013) Report on the ENUMERATE Thematic Surveys on Digital Collections in European Cultural Heritage Institutions 2013.
- Sunlight Foundation (2010). Ten Principles for Opening Up Government Information. August 11, 2010. Available: http://sunlightfoundation.com/policy/documents/ten-open-data-principles/
- Terras, M. (2010) Digital curiosities: resource creation via amateur digitization. Literary and Linguistic Computing, 25(4), pp. 425-438.
- Wyatt, L. (2011). Wikipedia & Museums: Community Curation. Uncommon Culture, 2(1), 33-41.
- 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, 9(2), 2014.
- Zuiderwijk, and M. Janssen, A Coordination Theory Perspective to Improve the Use of Open Data in Policy-Making. In: M. A. Wimmer, M. Janssen, and H. J. Scholl (eds.): EGOV 2013, LNCS 8074, pp. 38–49, 2013.