

Article

Beat Estermann

Université de Berne (Suisse)

OPENGLAM BENCHMARK SURVEY – MESURE DE LA PROMOTION DES DONNÉES OUVERTES / CONTENUS OUVERTS DANS LE SECTEUR DU PATRIMOINE)

Résumé

En raison de la numérisation généralisée du contenu patrimonial et de l'usage grandissant des médias sociaux, de nombreuses institutions liées à la préservation du patrimoine (musées, bibliothèques, centres d'archivage) ont ouvert et rendu disponible un nombre grandissant de collections et de méthodes de travail permettant d'y accéder depuis l'extérieur, comme en témoignent les données ouvertes / les contenus ouverts, l'utilisation des réseaux sociaux, la création de contenu commun et les données liées. Afin de mesurer le degré d'ouverture des institutions vouées à la préservation du patrimoine, une enquête modèle a été élaborée et administrée aux institutions de divers pays. Cet article expose certaines trouvailles fondamentales, les forces et les faiblesses, ainsi que les défis et les possibilités liés à l'approche choisie pour mesurer les « données ouvertes », le « contenu ouvert » et les pratiques afférentes dans le secteur patrimonial. Afin de rendre compte des particularités de notre approche, nous la comparons à celle utilisée pour l'enquête ENUMERATE, une étude internationale visant à mesurer la numérisation au sein des institutions vouées à la préservation du patrimoine, ainsi qu'à celle utilisée pour l'Open Data Barometer et le Global Open Data Census, deux initiatives internationales visant à surveiller la progression des données ouvertes publiques dans un grand nombre de pays. L'article se conclut sur des propositions visant de nouveaux développements.

MOTS-CLÉS : patrimoine culturel, numérisation, médias sociaux, données ouvertes, production participative (crowdsourcing en anglais), données liées, enquête de référence

Paper

Beat Estermann

University of Bern (Switzerland)

OPENGLAM BENCHMARK SURVEY – MEASURING THE ADVANCEMENT OF OPEN DATA / OPEN CONTENT IN THE HERITAGE SECTOR

Abstract

The widespread digitization of heritage content and the increasing use of social media have triggered many heritage institutions (museums, libraries, archives) to increasingly open up collections and work processes for participation from the outside, as exemplified by open data/open content, social media use, collaborative content creation, and linked data. In order to measure the degree of openness of heritage institutions, a benchmark survey was developed and administered among institutions in several countries. In the present paper we present some of the key findings and discuss the strengths and weaknesses, as well as the challenges and opportunities related to the approach chosen to measure 'open data', 'open content', and related practices within the heritage sector. In order to give an account of the particularities of our approach we compare it to the one used for the ENUMERATE survey, an international study aimed at the measurement of digitization among heritage institutions, and to the one employed for the Open Data Barometer and the Global Open Data Census, two international efforts to survey the advancement of open government data in a large number of countries. We conclude the paper by suggesting lines for further development.

KEYWORDS: cultural heritage, digitization, social media, open data, crowdsourcing, linked data, benchmark survey.

Introduction

The emerging collaborative culture on the Internet provides heritage institutions with new opportunities, but creates also new challenges for them. Some of the new emerging practices can be subsumed under the term ‘OpenGLAM’, the equivalent of ‘open government’, applied to the cultural heritage sector. More specifically, ‘OpenGLAM’ relates to practices such as the digitization of holdings, inter-organizational cooperation involving the exchange of metadata, open data, the engagement of audiences through social media, crowdsourcing, collaborative content creation, and linked data. In order to measure the state of advancement of OpenGLAM in various countries, and to identify the main challenges and obstacles with regard to the promotion of OpenGLAM and free access to knowledge, an international benchmark survey is carried out. So far, the survey has been rolled-out in Brazil, Bulgaria, Finland, New Zealand, Poland, Portugal, Switzerland, The Netherlands, and Ukraine. In the present paper we present the background, the motivations, and the actual implementation of the survey. Drawing on earlier papers, we summarize the main findings that have so far been derived from it. In addition, we propose an OpenGLAM Index summarizing the advancement of heritage institutions with regard to various practices relevant to OpenGLAM. Based on a comparison with other surveys covering similar aspects we discuss its strengths and weaknesses, as well as the challenges and opportunities that come with the chosen approach. We conclude the paper with a few suggestions for further development.

The Open Glam Benchmark Survey

OpenGLAM

The term ‘OpenGLAM’ is used by the Open Knowledge Foundation as a rough equivalent of ‘Open Government’, but applied to the cultural heritage sector. The acronym ‘GLAM’ stands for ‘*galleries, libraries, archives, and museums*’ and is used to refer to heritage institutions in general. According to the introduction to the Open Knowledge Foundation’s ‘OpenGLAM Principles’ (OKFN 2013), the objective of ‘OpenGLAM’ consists in encouraging heritage institutions to seize the opportunities offered by the Internet by engaging ‘*global audiences*’, by making their collections ‘*more discoverable and connected than ever*’, and by allowing users ‘*not only to enjoy the riches of the world’s memory institutions, but also to contribute, participate, and share*’. The principles themselves focus on aspects of openness, in the sense of the Sunlight Foundation’s definition of ‘open data’ (see below), and on ‘*novel ways of engaging audiences on the web*’, i.e. various forms of e-participation.

Motivation / Purpose of the Survey

The OpenGLAM Benchmark Survey has been carried out in a collaborative effort of NGOs, heritage institutions, and research institutions with a common interest in the promotion of OpenGLAM. The declared purpose of the survey is to measure the state of advancement of OpenGLAM in the participating countries, to identify the main challenges and obstacles with regard to the promotion of OpenGLAM and free access to knowledge, and to inform the heritage community about the latest developments in the area of OpenGLAM.

For the NGOs involved (mainly national chapters of Wikimedia, Open Knowledge, and Creative Commons), the survey is also a means to identify potential partners for open data and/or crowdsourcing projects, and they may use the study reports as a communication instrument to promote OpenGLAM within their local heritage community and related sectors as well as for their political work as interest groups in favor of open data and free knowledge.

In addition, the international benchmark study provides international comparisons, which allow each country to see where it stands compared to other countries and provide the international OpenGLAM community with a tool that helps it better understand the particularities of each country, to put insights gained in a country into a broader perspective, and to better adapt strategies and best practices to the specific situation of each country.

The Life-Cycle of the OpenGLAM Benchmark Survey

The beginnings of the survey date back to a pilot survey, carried out by the author with the support of a team of students of the Bern University of Applied Sciences in 2012. The pilot survey targeted some 200 heritage institutions in Switzerland (Estermann 2013). Given the positive reception of the survey, both among the Swiss heritage institutions and the international OpenGLAM community, it was decided to run the survey at an international level.

For the purpose of the international survey, the questionnaire was overhauled in an iterative process: an initial version was produced based on the questionnaire used for the Swiss pilot survey 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 adaptations were made to the final version of the questionnaire after its deployment in the first two countries, Poland and Finland.

The OpenGLAM Benchmark Survey is organized in a federative manner, which means that its roll-out depends on volunteers and partners in each country who take care of the translation of the questionnaire into the national language(s) and the compilation of contact lists. Ideally, the national teams also play a pro-active role when it comes to disseminating the survey results in their respective countries. This also means that data analysis takes place both centrally, with a focus on international comparisons, and in a decentralized manner, with a particular focus on a given country's burning issues. To this effect, the survey data has been made available as 'open data' from the outset, to the extent possible without encroaching on the confidentiality of individual institutions' responses.

And finally, the national teams also take care of the follow-up of respondents which indicated that they would like to be contacted in order to receive further information, training or consulting in one or several areas relevant to OpenGLAM.

Definition of Core Concepts

Before we get into the details of the methodological approach, we would like to clarify a few core concepts used in the context of this article, such as 'OpenGLAM', 'open data', 'open content', 'linked open data', and 'crowdsourcing', as well as the theory of innovation diffusion that serves as our primary theoretical lens:

Heritage institutions

In the context of our survey we followed 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'* (Nauta et al. 2011, p. 5).

Open data / open content

The open data movement experienced its worldwide breakthrough around 2009 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. 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.

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 economic 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).

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'. 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 collaborative activities taking place within online communities, such as the Wikipedia community.

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, as well as in preceding papers (e.g. Estermann 2015, Estermann 2016), 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. In the present paper we mainly refer to the innovation adoption process which 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).

Methodology of Data Collection

Survey instrument

The questionnaire used for the survey contains 34 questions (for a more detailed account of the questionnaire elaboration process, refer to Estermann 2015): 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 weren't included in all the countries.

Sampling approach and response rates

In each country we attempted to invite all the known heritage institutions to participate in the survey. The availability of lists of heritage institutions varied from country to country, so that distribution lists for the different countries contained between ca. 60% and more than 90% of all heritage institutions. 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 when accounting for methodological differences in the way the distribution lists have been assembled (Estermann 2015).

The overall response rate for the nine countries covered so far was 11.3%. There were however significant differences among the various countries: The highest response rate has been observed in Finland (25.8%), followed by Switzerland (19.5%). The lowest response rates have been registered for Brazil (6.3%) and Bulgaria (10.4%).

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 across countries. Most of the findings presented in this article are based on a combined sample of 1030 institutions from 7 European countries (Bulgaria, Finland, Poland, Portugal, Switzerland, The Netherlands, and Ukraine), Brazil and New Zealand. Some findings from an earlier conference paper (Estermann 2015) are based on a smaller sample of 584 institutions from Poland, Finland, Switzerland, and The Netherlands.

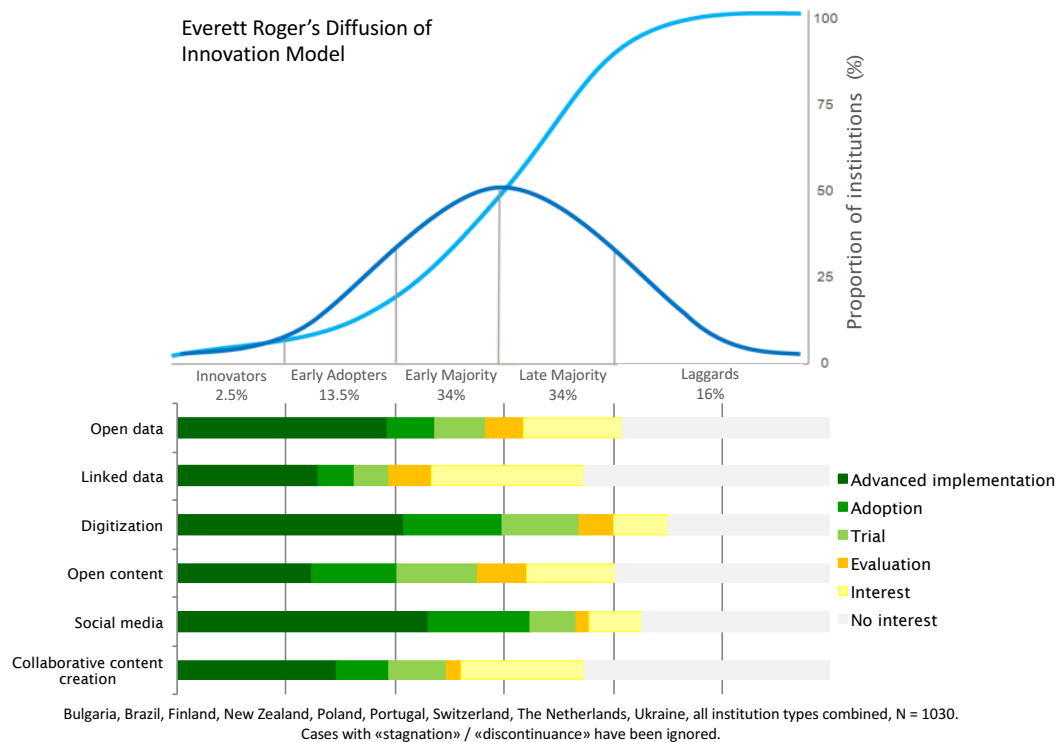
Summary of Findings

In this section we will give a brief summary of the main findings of the survey so far and present country comparisons for some of the aspects that have turned out to be particularly relevant with regard to the progress of OpenGLAM.

Diffusion of Internet-related practices among heritage institutions

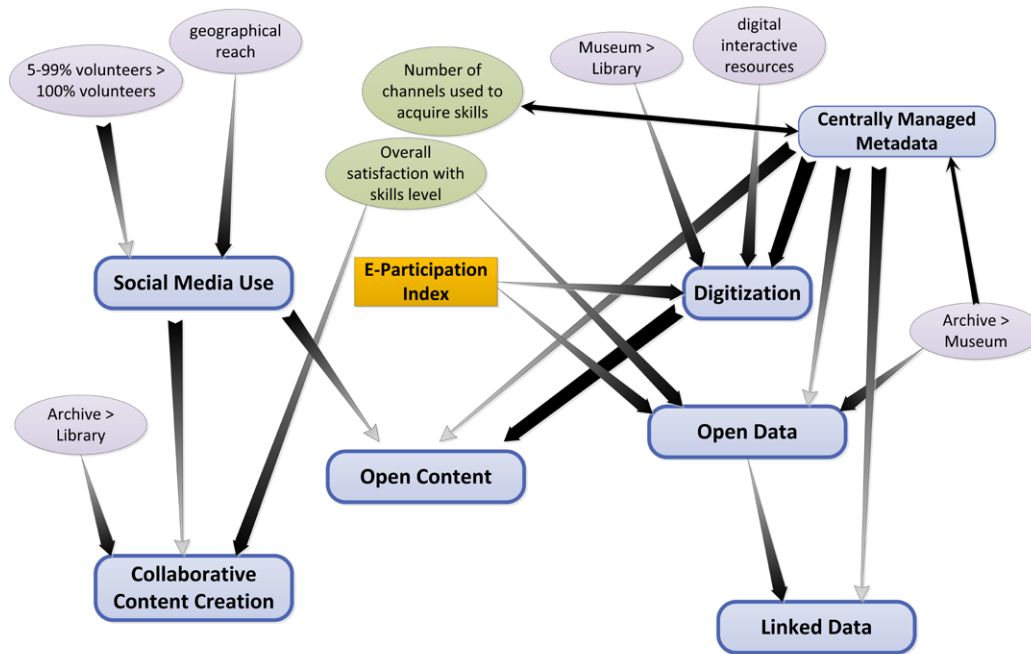
Figure 1 shows the present state of the diffusion of various Internet-related practices within the heritage sector (see Estermann 2015 for a detailed account of the operationalization of the various concepts): The use of social media and digitization are the most widespread practices with adoption rates of around 50%. They are followed by open data (27% adoption) and open content (17% adoption). In both cases, the 'early majority' of institutions is starting to adopt the practice. Collaborative content creation (15% adoption) is lagging slightly behind and may soon reach the 'early majority', while linked data (9% adoption) is for the moment just being embraced by the 'early adopters'.

FIGURE 1 Diffusion of Internet-related practices among heritage institutions



In order to reach a better understanding of the innovation diffusion processes, the factors influencing the adoption of the various practices have been investigated and the links and mutual influences between various Internet-related practices have been examined by means of a series of regression analyses. Figure 2 gives an overview of the influences that have been found to be significant at the 0.01 level (see Estermann 2016 for a detailed account of the factors that have been found to be significant at the 0.05 level).

FIGURE 2 Factors influencing the diffusion of Internet-related practices among heritage institutions



Results based on multinomial regression analysis; represented are correlations that are significant at 0.01.

The color and the shape of the arrows indicate whether a correlation holds for the first, the second, or both phases of the adoption process.

As can be seen in figure 2, the existence of centrally managed metadata (e.g. in the form of catalogues, inventories, or finding aids) plays an important role with regard to the adoption of 'digitization' as well as to the initiation of the adoption process of 'open data', 'open content', and 'linked data'. It is interesting to note that 36% of institutions indicated that they do not have such centrally managed metadata.

Furthermore, several dependencies between the different Internet-related practices have been identified. Thus, the adoption of 'open content' is generally preceded by the adoption of 'social media use' and 'digitization', while the adoption of 'collaborative content creation' tends to be preceded by the adoption of 'social media use'. 'Open data' in turn is a prerequisite of 'linked data', which indicates that, within the heritage sector, the latter primarily amounts to publishing 'linked open data'.

Several context factors at the country level were introduced into the model: the GDP and the E-Participation Index at the macro level, as well as the overall effectiveness of the use of different methods of skills acquisition by the heritage institutions of a given country, as it appears from the survey data.

Most interestingly, the analyses showed a strong correlation of the E-Participation Index with the adoption of 'open data', 'digitization', and indirectly with the adoption of 'open content'. The E-Participation Index is calculated as part of the UN E-Government Survey (UN 2014) and focuses on the use of online services to facilitate the provision of information by governments to citizens ('e-information sharing'), interaction with stakeholders ('e-consultation') and engagement in decision-making processes ('e-decision making'). Based on a qualitative assessment of the availability and relevancy of participatory services available on government websites, it is reflective of the online participation culture within a given country.

Surprisingly, the adoption of the different practices under examination turned out to be unrelated to the GDP, which serves as a proxy not only for the economic situation in a given country, but also for the quality of its IT infrastructure, given the high correlation between the two.

As to the effective use of different methods of skills acquisition among a country's heritage institutions, no positive influence was found regarding the adoption of the various Internet-related practices. The skills-related factors influencing the adoption of some of the practices therefore do not seem to be related to the level of effective information and/or educational offers in a given country. They rather appear to be linked to the skills-acquisition strategy employed by a given institution. This is exemplified by the strong correlation between the number of channels used by an institution to acquire skills and know-how and the presence of centrally managed metadata, as well as by the strong links between the 'overall skills level' of an institution and its adoption of 'open data' and 'collaborative content creation'.

Differences between various types of heritage institutions

While an institution's strategy of acquiring skills and know-how may be subject to change, other characteristics of an institution that were found to play a role with regard to the adoption of some of the Internet-related practices are immutable or at least less prone to change. Thus, the adoption of social media use is more likely among institutions with a wider geographical reach and among institutions with a workforce comprising both paid staff and volunteers compared to purely volunteer based institutions. The institution type was also found to play a role: museums are more likely than libraries to adopt 'digitization', while archives are more likely to adopt 'open data' than museums and more likely to adopt 'collaborative content creation' than libraries. And finally, institutions which hold digital interactive resources are more likely to digitize their non-digital born holdings than their counterparts. Again, in order to focus on the big picture we are just mentioning the correlations that were found to be significant at the 0.01 level; a more detailed account has been provided by Estermann 2016.

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

In order to provide further insights into the dynamics of the innovation adoption processes, institutions were asked about the challenges, risks, and opportunities of 'open content' and 'crowdsourcing'. As noted in an earlier paper, based on the analysis of a sample of 584 institutions, *'the greatest benefit 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'* (Estermann 2015, p. 20). On the opposing side are the challenges and risks associated with 'open content': *'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%) [...] [W]hen 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%)'* (Estermann 2015, pp. 20-21).

As to the pros and cons of crowdsourcing, all the purposes suggested in the questionnaire were mentioned by at least half of the institutions: *'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%)', [...] [while] 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%)'* (Estermann 2015, p. 21).

Dynamics of the adoption processes

In order to get an idea of the dynamics of the adoption of 'open data', 'open content', and crowdsourcing, different indicators have been used, including a comparison of 'importance' versus 'desirability' ratings for the different practices, the analysis of the changes in the institutions' attitudes as they progress in the innovation adoption process, the analysis of the presence (or absence) of important prerequisites or 'show-stoppers', such as the existence of centrally managed metadata within an institution, as well as explicit declarations by the responding institutions regarding their future practice. First of all, it should be noted that all Internet-related practices appear to be self-reinforcing: institutions that have reached higher adoption levels tend to perceive most practices as more important and also as more desirable for them, and only a very small number of institutions indicated that they would abandon a given practice (Estermann 2016). As to the speed of innovation diffusion, we reached the following conclusions based on the analysis of various indicators in the four-country sample (N = 584): If the dissemination of 'open data' and 'open content' continue at the rates suggested by our survey data, we can expect that all the institutions which have centrally managed metadata (i.e. 70% of institutions) will have adopted 'open data' in about 10 years and 'open content' in about 15 years from now; in the case of crowdsourcing, the data suggest that widespread adoption would take a bit longer (see Estermann 2015 for a detailed account of the estimation approach).

Do changes in attitudes precede or follow practice?

In one of our earlier papers (Estermann 2016) we examined the changes in attitudes as institutions pass through the different adoption stages for the various practices. While some differences in attitudes could be observed depending on the adoption stage at which a given institution finds itself, there were no dramatic effects that would change the dynamics of the adoption processes, and the changes in attitudes appeared to be rather slow. This was in particular the case for the institutions' attitudes regarding 'open content', which seemed somewhat disconnected from the declared practice regarding the opening up of collections. In fact, even institutions that reportedly had started to make their collections available as 'open content' persisted in their hesitant attitudes with regard to making content available for re-use by third parties for any purpose, including commercial use. As it turns out – provided that the institutions correctly reported their practice – engaging in the practice of 'open content' does not seem to require a previous change of attitude, and it remains to be seen whether and when these attitudes may actually change in the future so as to reflect actual practice.

Country comparisons

Given the findings presented above, we propose a series of country comparisons: Figure 5 in the annex shows the advancement of the heritage institutions of the different countries with regard to the adoption of the six practices under consideration (note that in order to account for the different sample sizes for the different countries, we have indicated the error bars for a confidence level of 95%): The picture is rather heterogeneous, with the constant that Switzerland is lagging behind on most practices, while the Netherlands are among the forerunners, except with regard to 'collaborative content creation', where Finland is leading (the value for Ukraine on 'collaborative content creation' appears to be over-evaluated due to self-selection bias, as established by propensity score matching, cf. Austin 2011). Figure 6 in the annex shows the values of a series of institutional-level variables for the different countries. Here, some interesting differences appear: While most countries have a similar rate of institutions with centrally managed metadata, this rate is notably lower for Brazil (46% vs. around 70% on average). When it comes to acquiring skills and know-how, Swiss institutions are less likely to use many different ways to do so than the average, while their counterparts in Finland and New Zealand are more likely to do so (48% vs. 78% and 88%, respectively). Regarding the geographical reach of the institutions, Polish institutions are less likely to have a purely 'local/regional' focus than their counterparts in most other countries. Interesting differences also appear regarding the ratio of pure volunteer organizations within the sample (the Netherlands has the highest rate with an astonishing 46%, followed by Switzerland with 22%) and the ratio of very small institutions with no more than 1 FTE paid staff and/or volunteers (where Switzerland shows the highest ratio with 22% of institutions). And finally, figure 7 in the annex shows the country-level indicators that were entered into the regression model presented above (E-Participation Index, GDP per capita, and the effective use of four basic methods of skills acquisition by the country's heritage institution) as well as two further variables shown for illustrative

purposes: the E-Government Development Index as well as the density of heritage institutions based on the estimates we arrived at when gathering the contacts lists. As can be seen from the first graphic, the countries covered by the survey score much more unevenly on the E-Participation Index than on the E-Government Development Index, with Bulgaria, Switzerland, Poland, and Ukraine lagging clearly behind in the area of e-participation, while The Netherlands are clearly ahead. In contrast, Switzerland is clearly ahead on GDP, while the values for Ukraine, Bulgaria, Brazil, Poland and Portugal are 5 to 10 times smaller. Given the low advancement of Swiss heritage institutions regarding the adoption of various Internet-related practices, this illustrates again the irrelevance of the economic situation with regard to the adoption of said practices among heritage institutions of a given country. As can be seen from the third graphic, the effective use of four basic methods of skills acquisition (guidelines/desk research; conferences; training; advice from peers) by the country's heritage institutions has little distinctive power, with most values ranging between 3 and 3.5 on a scale from 0 to 5. And finally, the last graphic shows the highly heterogeneous picture when it comes to the density of heritage institutions in the different countries, with values ranging from 21 institutions (Ukraine) to 199 institutions (Switzerland) per million inhabitants.

OpenGLAM Index

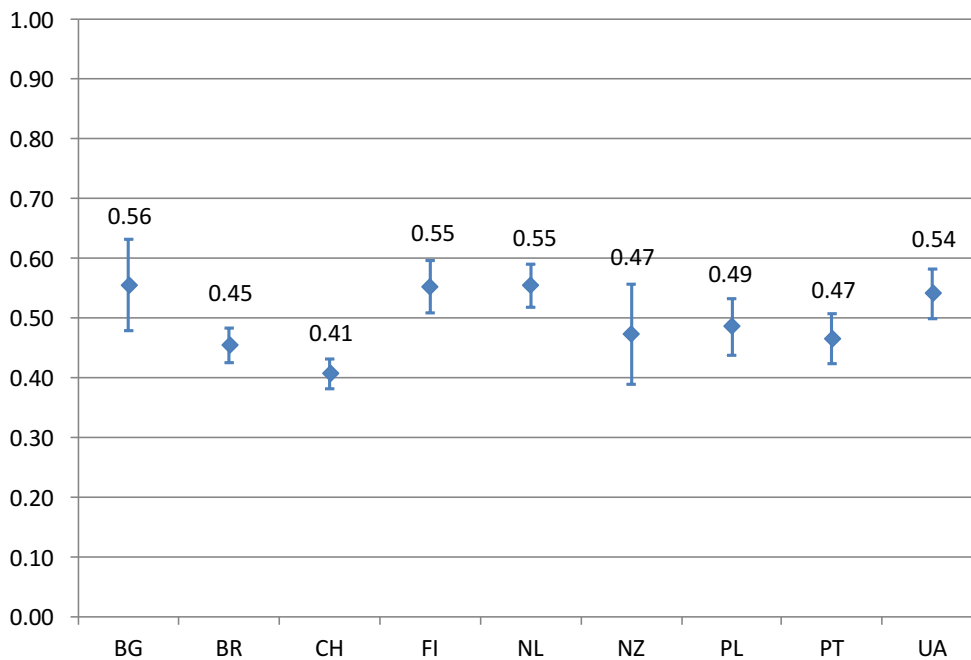
Based on the insights gathered by means of the regression analyses, we calculated an OpenGLAM Index composed of three sub-indexes. Table 1 gives an overview of the composition of the OpenGLAM Index and the relative weight of its sub-indexes and their components. The calculation model employed reflects the fact that the practices relevant to OpenGLAM can be roughly divided into three areas, as it appeared from the regression analyses: engaging audiences on the Internet, 'open data', and 'open content'. In the calculation of the index, each of these areas is given the same weight. For each of them, we are taking into account several variables, representing different levels of the given practice. This may include pre-cursors of the actual practice, such as the existence of 'centrally managed metadata' in the case of 'open data' and 'open content', or the adoption level of digitization in the case of 'open content'. In the case of 'open data' we also included the adoption level of 'linked data' as an indicator of higher 'open data' maturity.

TABLE 1 Composition of the OpenGLAM Index and relative weight of its sub-indexes and their components

Engaging audiences (Sub-index)	33%
adoption level of social media use (40%) adoption level of 'collaborative content creation' (60%)	
Open data (Sub-index)	33%
existence of centrally managed metadata (20%) adoption level of 'open data' (60%) adoption level of 'linked data' (20%)	
Open Content (Sub-index)	33%
existence of centrally managed metadata (20%) adoption level of digitization (20%) adoption level of 'open content' (60%)	

The OpenGLAM Index provides a summary statistic with regard to the advancement in the area of OpenGLAM that is calculated at the level of individual institutions and can be aggregated in different ways for comparative purposes. Figure 3, for example, gives an overview of the OpenGLAM Index for the countries so far covered by the survey.

FIGURE 3 OpenGLAM Index for the different countries



Discussion

In the present paper we have described the making of the OpenGLAM Benchmark Survey as well as the main insights that have so far been derived from it. We have shown how the adoption of the various Internet-related practices by heritage institutions can be understood in terms of the innovation diffusion approach popularized by Everett M. Rogers. We have provided an overview of the advancement of the diffusion of the different practices among heritage institutions, shown how the different practices relate to each other, and given estimates with regard to the further diffusion of OpenGLAM-related practices. In order to complement the picture, we have given an account of the driving and hindering factors from the point of view of heritage institutions. In addition, the survey has allowed us to give an account of the heritage institutions' attitudes with regard to various aspects of OpenGLAM and to track shifts in these attitudes as the institutions progressively adopt various Internet-related practices. Given the important differences regarding the adoption of the various practices by the heritage institutions of different countries, we introduced several country-level variables into our regression models, concluding that a country's online participation culture (as measured by the E-Participation Index) plays an important role with regard to the adoption of OpenGLAM, while its economic situation (as measured by the GDP per capita) is irrelevant. It has to be noted, however, that the analyses involving country-level variables are rather exploratory in nature compared to the rest of the analyses presented in this paper. This is due to the fact that only nine countries have been covered so far by the survey, which doesn't allow for very sophisticated analyses of country-level influences. By providing various country comparisons, we have not only been able to live up to the promise of creating an international benchmarking tool, but also to point out particularities of individual countries and their heritage sector. Thus, we have been able to point to important differences in the structure of the heritage domain across the different countries – for example with regard to the role of volunteer work or with regard to the average size of institutions and their density. These insights may lead to the formulation of concrete challenges for memory politics in individual countries, as is the case for Switzerland, where the question needs to be tackled what it means to have a particularly high density of particularly small institutions in view of their ability to cope with current changes in the heritage sector.

One of the main methodological difficulties of the survey also relates to these country differences: due to the different structure of the heritage domain in the various countries it is a challenge to paint an equitable picture, while the varying responding behavior among heritage institutions in different countries only adds to this challenge. To a certain extent, these challenges can however be addressed by adequate analytical methods. Another challenge related to the implementation of the survey relates to its crowdsourcing aspect: it heavily relies on the personal engagement of volunteers in various countries for its roll out. While the reliance on people's intrinsic motivation is certainly instrumental in attaining a good level of quality regarding the questionnaire translation and the compilation of contacts lists, it has at the same time also its limits, as is demonstrated by the fact that not more countries have been covered so far, in spite of people in many other countries having at some point manifested an initial interest in helping to run the survey.

Strengths and Weaknesses Compared to Other Surveys

In order to better bring to the fore the strengths and weaknesses of the OpenGLAM Benchmark Survey, we will conclude our discussion by comparing our survey to three other surveys which are at least partly similar in nature: The ENUMERATE survey, whose aim is to monitor the progress on digitization of cultural heritage across Europe (Nauta et al. 2011), as well as the World Wide Web Foundation's Open Data Barometer (World Wide Web Foundation 2015) and the Open Knowledge Foundation's Global Open Data Census (OKFN 2015), which both aim to monitor the state of 'openness' of government data.

The OpenGLAM Benchmark Survey was from the outset meant to be complementary to the ENUMERATE survey in that it attempted to cover several Internet-related practices that build upon or are complementary to the digitization of heritage objects. As both surveys rely on heritage institutions filling in an online questionnaire, it was also clear from the outset that the OpenGLAM Benchmark Survey wouldn't be able to cover all the details related to the digitization of cultural heritage covered by the ENUMERATE survey. As a result, there are major thematic differences between the two surveys, but also methodological ones:

From a thematic point of view, ENUMERATE's focus is almost exclusively on questions relating to digitization. As a consequence, it covers this topic in much more detail than the OpenGLAM survey. The OpenGLAM survey, on the other hand, has a stronger focus on open data, open content, social media, and crowdsourcing. There is some thematic overlap between the two surveys regarding digitization, but just enough to allow for some comparisons: While the ENUMERATE survey is the only one to ask about the presence of a digitization strategy, the size of collections, access statistics, digitization expenses and staff effort, the OpenGLAM survey asks about planned digitization activities in the next five years and about reasons not to digitize certain content, aspects the ENUMERATE survey doesn't cover. Furthermore, the ENUMERATE survey is much more sophisticated when it comes to describing the individual institution's collections, and it asks institutions which fraction of their heritage objects should eventually be digitized. In that respect, the OpenGLAM survey is much more rudimentary, and it remains difficult to put the adoption levels of 'open data' and 'open content' in perspective by controlling for a collection's size or importance. The big plus of the OpenGLAM survey consists however in the fact that it allows to study inter-relations between digitization activities and other Internet-related practices, while the ENUMERATE survey is blind with regard to this aspect.

As to the methodological differences, the data for the ENUMERATE survey is (mostly) gathered through an open call, which poses some methodological challenges. In particular, it is impossible to make methodologically sound country comparisons based on the ENUMERATE data. The data for the GLAM survey, in contrast, is collected by means of individual invitations sent to heritage institutions to participate (ideally including all heritage institutions of a given country). By sending reminder emails, the GLAM survey arrives at better and more homogenous response rates. In sum, the GLAM survey presently covers fewer countries than the ENUMERATE survey, but covers the heritage sector in the countries that are included more thoroughly. When it comes to analyzing country-level factors, the ENUMERATE survey is superior in that it allows for more sophisticated analyses given the larger number of countries included (see for example Borowiecki & Navarrete 2015); it poses however the challenge that it is difficult to estimate to what extent the institutions included in the survey are actually representative for the heritage sector in their respective countries.

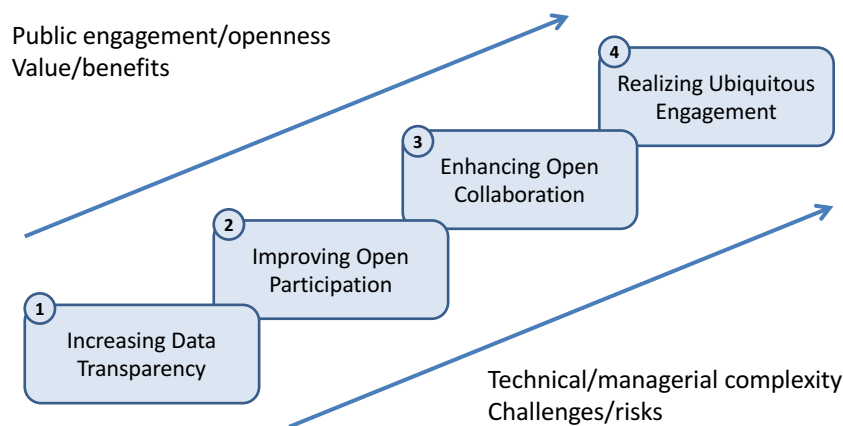
With regard to the Open Data Barometer and to the Global Open Data Census, the OpenGLAM survey is insofar complementary as it covers the heritage sector, which is not covered by the two surveys focusing on open government data. As with the ENUMERATE survey, we can again point to differences from a thematic and from a methodological point of view:

Both the Open Data Barometer and to the Global Open Data Census focus on open government data in a limited number of thematic areas; assuming that a series of standard datasets should exist in every country (e.g. on government spending or government contracts). There is an underlying assumption regarding the importance of certain datasets for certain purposes (e.g. with regard to the transparency of government). It appears however that this country focus fails to cover important datasets at the local and regional levels, as is illustrated by the fact that the Open Knowledge Foundation has lately launched a Local Open Data Census as a complement to the initial Global Open Data Census. The OpenGLAM survey, in contrast, is totally open regarding thematic areas, as there are no assumptions regarding the existence of data on particular topics. In that sense, it follows a more generic approach by assuming that every heritage institution has holdings which ought to be accompanied by metadata and which can be digitized (if not digital born). As mentioned above, the GLAM survey itself does not make any assumptions regarding the relevance of particular holdings or their publication as 'open data' / 'open content'. Up to a certain extent it is possible to take into account the relevance of holdings, provided that the holdings have been officially evaluated (which is the case for example in Switzerland): Taking into account the evaluation of holdings works fairly well in the case of smaller institutions; in the case of large institutions, the challenge lies in the fact that the OpenGLAM survey is not able to capture anything in reasonable detail at a collection level (as opposed to the institution level). While the Global Open Data Census focuses on the sole publication of data, the Open Data Barometer is broader in scope and covers also the countries' readiness to secure positive outcomes from an open government data initiative as well the impacts that arise from open data publication. In this regard, the OpenGLAM survey is narrower in scope, as it does not attempt to evaluate a country's readiness for OpenGLAM or to track any effective impacts. To a certain extent, the impact of OpenGLAM-related practices is however covered by the questions asking institutions about the purpose and the perceived benefits of 'open content', social media use, and 'crowdsourcing', combined with the possibility to analyze the institutions' views on those aspects for different stages of the innovation adoption process. The OpenGLAM survey is certainly broader in scope when it comes to covering 'open government' / 'OpenGLAM' as a whole, which includes not only the publication of data, but also online participation and collaboration (Lee & Kwak 2011, see figure 4). A possible rationale for focusing the enquiry on the sole publication of data is provided by Lee and Kwak themselves:

'A thesis of the OGIM states that government agencies should advance their open government initiatives in a progressive and orderly manner by focusing on one implementation stage at a time, starting from increasing data transparency (Stage 1), and then moving on to improving open participation (Stage 2), enhancing open collaboration (Stage 3), and realizing ubiquitous engagement (Stage 4). We argue that, by following this sequence, agencies can minimize risk and effectively harness the power of social media in order to engage the public' (Lee & Kwak 2011, p. 254).

Lee's and Kwak's postulation that the opening up of data should come before the extensive use of social media is not supported by our empirical data – at least not in the heritage sector, where the use of social media was even found to play a role in triggering the adoption of 'open content'.

FIGURE 4: Open Government Implementation Model' (Lee & Kwak 2011)



The main methodological difference lies in the fact that both the Open Data Barometer and the Global Open Data Census are polling open data experts and open data activists, respectively, while the OpenGLAM survey is polling individual institutions. By polling experts on the situation in a given country, it is possible to capture a series of technical aspects that are relevant to data users in more detail. On the other hand, it is rather difficult to capture the perspective of data providers in a way the OpenGLAM survey does. Furthermore, there are methodological difficulties in covering the sub-national and local levels as the main focus lies on the country level, which is particularly challenging in countries with a federal structure and shared responsibilities across different levels of government. As mentioned above, the Open Knowledge Foundation presently tries to fill some of the gaps by launching a Local Open Data Census. In the case of the GLAM survey there is no hierarchical approach: the primary level of analysis is the institution. From there, different types of aggregation are possible. One of the great strengths of the Open Data Barometer and the Global Open Data Census lies certainly in their broad coverage in terms of the number of countries included: the latter presently covers 122 countries, and the former 86 countries. The crowdsourcing approach seems to work better for their methodological approach than for the one chosen for the GLAM survey.

Conclusions and Outlook

After this thorough discussion of the strengths and weaknesses of the approach pursued in the context of the OpenGLAM Benchmark Survey, we would like to conclude the paper by suggesting a few lines for further development and analysis:

- Including further countries in the survey would allow for more solid analyses regarding influencing factors at the country level. Furthermore, it would allow sensitizing more institutions for the issues surrounding OpenGLAM. It is uncertain, however, whether a substantially larger number of countries can be reached by means of the crowdsourcing approach pursued so far; dedicated funding might be required
- As has been shown in the previous section, the OpenGLAM survey is complementary both to the ENUMERATE survey and to the Open Data Barometer and the Global Open Data Census. This complementarity should be leveraged in future analyses, e.g. by analyzing the issues surrounding the digitization of cultural heritage based on data both from the OpenGLAM survey and the ENUMERATE survey.
- Many questions of the survey have been inspired by previous qualitative studies (see Estermann 2014). Now that quantitative data is available for many aspects, and many new issues have been brought up in the course of the analysis, it might be time to get back to qualitative investigations, e.g. in the area of attitudinal changes among heritage institution's staff, with regard to coping strategies of smaller institutions, with regard to the role of volunteering in the heritage sector and its relation to OpenGLAM, or in order to better understand the (absence of) influence of country level factors, such as the e-participation culture or the economic situation on an institution's practice.

- Efforts should be undertaken in order to measure the impact of OpenGLAM among the heritage institutions themselves, but also at a societal and possibly also at an economic level.
- With regard to the dissemination of survey results and the sensitization of heritage institutions for the issues surrounding OpenGLAM, it is important that country specific analyses be made and disseminated locally (as has already been the case for Poland and partly for The Netherlands). There is also anecdotal evidence from the author's practice that the survey results are well suited to inform consulting in the area of OpenGLAM; it remains to be seen to what extent the example is followed by others.
- Another measure that could be taken in order to sensitize heritage institutions for the issues surrounding OpenGLAM consists in creating an interactive web app that allows institutions around the world to explore where they stand compared to similar institutions in their country or abroad.

Acknowledgements

I would like to thank all the lovely people who have made this research possible by supporting it in one or the other manner: My colleagues of the OpenGLAM Working Group of the Open Knowledge Foundation: Joris Pekel, Maarten Brinkerink, Dorota Kawęcka, Sanna Marttila, Helene Hahn, Thomasin Sleigh, Lieke Ploeger, and Mariana Valente, for helping to get the project started and to implement it in their respective countries. All the other people who provided feedback during the development of the questionnaire: Gemma Echevarria, Maria de Vallibana Serrano, Merete Sanderhoff, Laura Sillanpää, Gerhard Jan Nauta, Max Hammond, Laura Carletti, Alek Tarkowski, and Liliana Melgar. The people who helped translating the questionnaire: Joris Pekel, and Erwin Verbruggen (Dutch); Laura Sillanpää, Katja Bargum, Pia Lonardi, Antti Pakarinen, Tove Ørsted, and Maria Virtanen (Finnish); Léonard Kwuida and Frédéric Noyer (French); Helene Hahn and Sylvia Petrovic-Majer (German); Dorota Kawęcka, Aleksandra Janus, and Anna Buchner (Polish); Vladislav Nedeleev and colleagues (Bulgarian); Maria Valente and colleagues (Brazilian Portuguese); Helena Patrício, Margarida Lopes, and Dulce Fontes (adaptation to European Portuguese), Anna Khrobolova, Bohdan Melnychuk, Ekaterina Estermann, and Vadim Zaytsev (Russian), as well as Anna Khrobolova, Vira Motorko, Bohdan Melnychuk, Oksana Brui, and Galyna Onysko (Ukrainian). And last, but not least the national teams in the various countries who helped rolling out the survey.

The survey data, along with the questionnaire, are available on: [Online].[survey.openglam.ch]

References

- AUSTIN, P. C. (2011). "An introduction to propensity score methods for reducing the effects of confounding in observational studies", *Multivariate behavioral research*, 46(3), 399-424.
- BAUER, F. and M. KALTENBÖCK (2001). *Linked Open Data: The Essentials: A Quick Start Guide for Decision Makers*, SEMANTIC-WEB COMPANY, 2011.
- BEAL, G.M. and J. M. BOHLEN (1957). "Special Report No. 18", *Agriculture Extension Service*, Iowa State College, 1, pp. 56-77, 1957, Reprinted in 1981.
- BOROWIECKI, K. J. and T. NAVARRETE (2015). "Digitization of heritage collections as indicator of innovation", *Discussion Papers on Business and Economics*, Denmark, University of Southern.
- ESTELLÉS-AROLAS, E. and F. GONZÁLEZ-LADRÓN-DE-GUEVARA (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". *Journal of theoretical and applied electronic commerce research*, 9(3), 15-31.
- ESTERMANN, B. (2015). "Diffusion of Open Data and Crowdsourcing among Heritage Institutions. Based on data from Finland, Poland, Switzerland, and The Netherlands" *Paper Presented at the EGPA 2015 Conference*, in Toulouse, France, 26-28 August 2015.
- ESTERMANN, B. (2016). "'OpenGLAM' in Practice – How Heritage Institutions Appropriate the Notion of Openness", *Paper presented at the IRSPM 2016 Conference*, in Hong Kong, 13-15 April 2016.

- GREENHALGH, T., G. ROBERT, F. MACFARLANE, P. BATE and O. KYRIAKIDOU (2004). *Diffusion of innovations in service organizations: systematic review and recommendations*, *Milbank Quarterly*, 82(4), 581-629.
- HOWE, J. (2006). "Crowdsourcing: A Definition", *Crowdsourcing Blog*, 2 June 2006. [Online]. [crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html] Last accessed on January 24, 2013.
- JANKOWSKI, J., Y. COBOS, M. HAUSENBLAS and S. DECKER (2009). "Accessing cultural heritage using the web of data", *Paper presented at the 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST* (2009).
- LEE, G. and Y. H. KWAK (2011). "Open government implementation model: a stage model for achieving increased public engagement", *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, *ACM* (pp. 254-261).
- NAUTA, G.J., S. BAKKER and de M. NIET (2011). *ENUMERATE Core Survey 1 Methodology*, November 2011. [Online]. [www.egmus.eu/fileadmin/ENUMERATE/deliverables/ENUMERATE-D2-04.pdf] Last accessed on 30 July 2015.
- OKFN (2013). *OpenGLAM Principles*, version 1.0, as established by the Open Knowledge Foundation's OpenGLAM Working Group, [Online]. [openglam.org/principles/] Last accessed on 8 March 2016.
- OKFN (2015). *Global Open Data Index – Methodology*. [Online]. [index.okfn.org/methodology/] Last accessed on 12 April 2016.
- ROGERS, E. M. (2003). *Diffusion of innovations/ Everett M. Rogers*.
- STROEKER, N. and R. VOGELS (2014). "Survey Report on Digitization in European Cultural Heritage Institutions 2014", *ENUMERATE Project*. [Online]. [www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2014.pdf] Last accessed on 30 July 2015.
- SUNLIGHT FOUNDATION (2010). *Ten Principles for Opening Up Government Information*, August 11, 2010. [Online]. [sunlightfoundation.com/policy/documents/ten-open-data-principles/]
- UN (2014). *UN E-Government Survey*, United Nations, Department of Economic and Social Affairs. [Online]. [publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2014-Survey/E-Gov_Complete_Survey-2014.pdf] Last accessed on March 17, 2016.
- WORLD WIDE WEB FOUNDATION (2015). *Open Data Barometer Global Report*, Second Edition.

Annex

FIGURE 5: Adoption rates of various Internet-related practices (country comparisons)

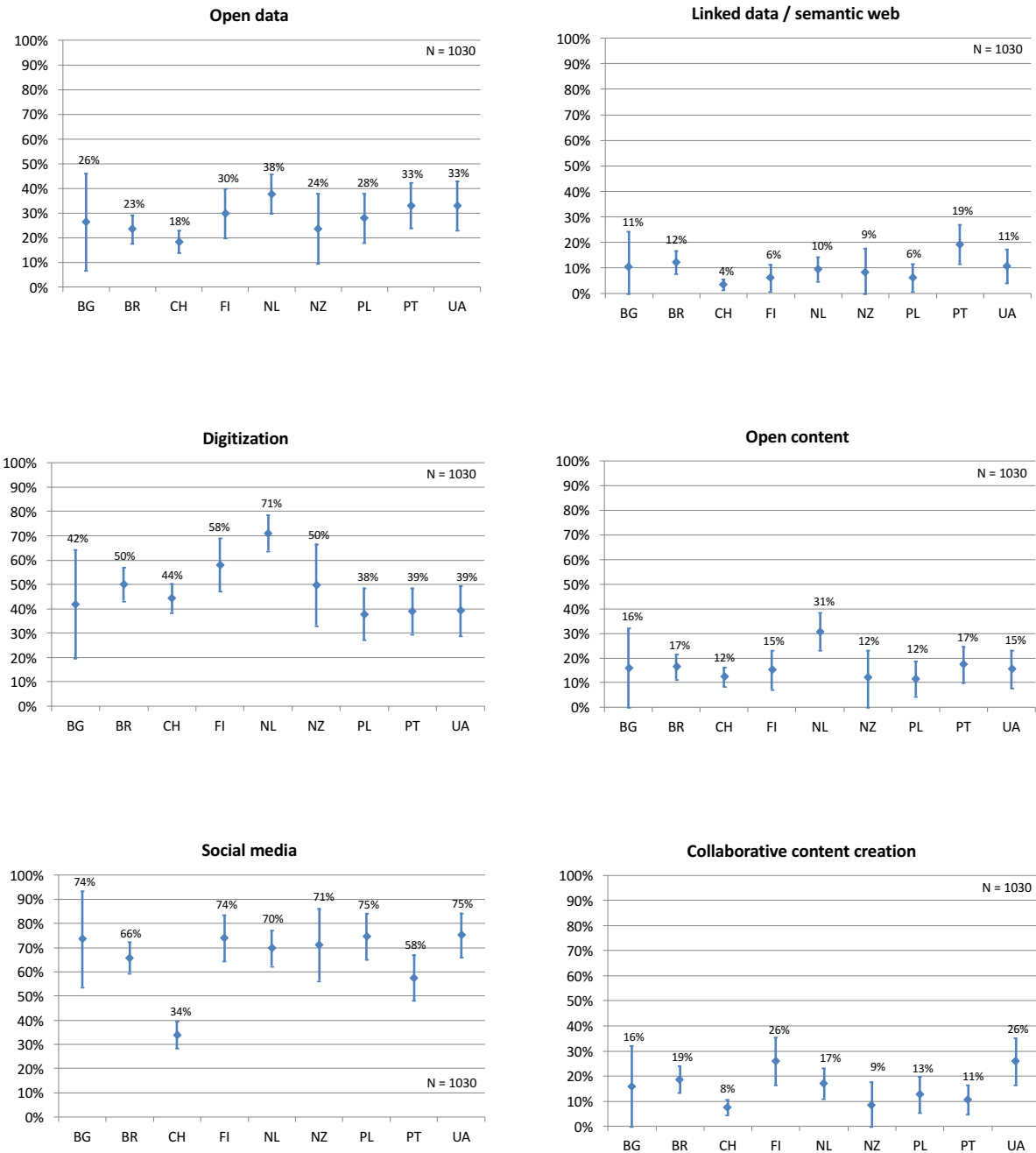


FIGURE 6: Selection of institution-level variables (country comparisons)

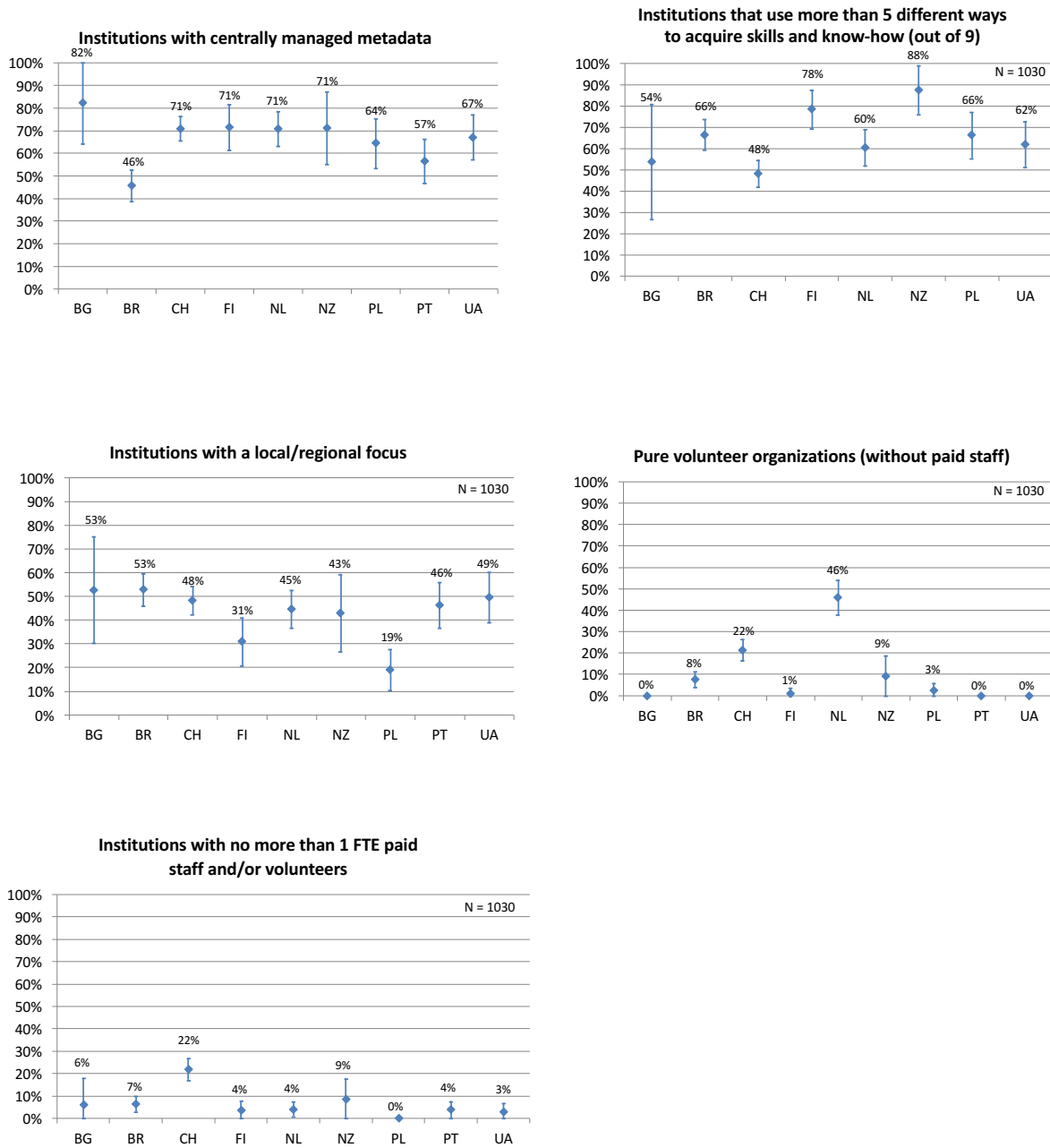


FIGURE 7: Selection of country-level indicators

